Physical Biology of the Cell
Hints to the Problems

Rob Phillips, Jane Kondev, Julie A. Theriot and Hernan G. Garcia

January 18, 2013

Contents

Preface

This document contains a series of hints to guide the reader with some of the more difficult problems in the book. The goal of the hints is to give the reader a better sense of what we are asking for in the problem since often the problem statements can be complemented by extra information. Second, the goal of the hints is to sketch a logical attack on the problem of interest.

If our readers find a specific problem obscure or deserving of a hint, please let us know and we will consider adding a hint for the problem of interest.

Chapter 2

Biological Structures: Rulers at Many Different Scales

2.4. A feeling for the numbers: molecular volumes and masses.
(a) If you want to use actual PDB files of amino acids to calculate their volumes
you'll have to look around on the Internet. A simple Google search for "amino
acid pdb" should give you plenty of relevant pages. One useful one, for example,
is `http://wbiomed.curtin.edu.au/teach/biochem/tutorials/pdb/index.html`.
These coordinates can also be found on the book website.

2.6. Atomic-Level representations of biological molecules.
(a) The following are the PDB accession number of the molecules used in this
problem or the link to the relevant websites. All of them are also available on
the book website.

- ATP: `ATP.pdb` from `http://xray.bmc.uu.se/hicup`.

- Phosphatidylcholine: `stearyl-oleyl-phophacholine.pdb` from
  `http://faculty.gvsu.edu/carlsont/mm/lipids/pg.html`.

- B-DNA: `bdna.pdb` from `http://chemistry.gsu.edu/glactone/PDB/pdb.html`.

- G-actin: `1J6Z.pdb` from the PDB.

- Lambda repressor/DNA complex: `1LMB.pdb` from the PDB has the DNA-binding region of lambda repressor complexed with DNA.

- Lac repressor/DNA complex: `1lbg.PDB` and `1tfl.pdb` from the PDB.

- Hemoglobin: `1hga.pdb` from the PDB.

- Myoglobin: `1mbo.pdb` from the PDB.

- HIV gp120: `1GC1.PDB` from the PDB.

- Green fluorescent protein (GFP): `1GFL.pdb` from the PDB.

- RNA polymerase: `1L9U.pdb` from the PDB.

2.7. Coin flips and partitioning of fluorescent proteins.

(a) You can calculate both $\langle n_1^2 \rangle$ and $\langle n_1 \rangle^2$ by using the observation that

$$\langle n_1^b \rangle = \sum_{n_1=0}^{N} n_1^b \frac{N!}{n_1!(N-n_1)!} p^{n_1} q^{N-n_1}. \tag{2.1}$$

If we recall that

$$(p+q)^N = \sum_{n_1=0}^{N} \frac{N!}{n_1!(N-n_1)!} p^{n_1} q^{N-n_1}, \tag{2.2}$$

then we can find $\langle n_1 \rangle$, for example, as

$$\langle n_1 \rangle = p\frac{\partial}{\partial p} \sum_{n_1=0}^{N} \frac{N!}{n_1!(N-n_1)!} p^{n_1} q^{N-n_1}. \tag{2.3}$$

This same strategy can be used to find $\langle n_1^2 \rangle$.

(b) The key assumption for this part of the problem is $I = \alpha N$, $I_1 = \alpha N_1$, etc. By substituting these values into the expression for $\langle (I_1 - I_2)^2 \rangle$ we can beat this into a form where we need nothing more than to just compute quantities of the form $\langle n_1^b \rangle$ as we did in part (a) of the problem.

Chapter 4

Who: Bless the Little Beasties

4.1. Structure of hemoglobin and myoglobin.
Check out the hints for problem 2.6 for information about the relevant PDB files. They can also be found on the book website.

4.4. Mutation correlation and physical proximity on the gene.
The key concept in the problem is that the frequency of recombination between two alleles is proportional to the distance between them. The first thing to do is to evaluate the ratios in table 4.2, thus obtaining "distances" whose units Sturtevant later dubbed "centimorgans"—one centimorgan or cM corresponds to 1% recombination frequency. Now the problem becomes one of finding the linear ordering of the factors B, C, O, P, R, and M consistent with the distances derived from the data.

Chapter 5

Mechanical and Chemical Equilibrium in the Living Cell

5.1. Energy cost of macromolecular synthesis.

As an example we will determine the biosynthetic cost of the amino acid valine. We start by accessing "valine biosynthesis" in "ecocyc.org". Clicking on "More Detail" once shows the structures of the intermediate species.

Notice that 2 pyruvate molecules are the precursor for this amino acid. Let's first calculate its energy cost. In order to do that we can look at figs. 5.2Constructing the Cell: Managing the Mass and Energy Budget of the Cellfigure.5.2 and 5.5Constructing the Cell: Managing the Mass and Energy Budget of the Cellfigure.5.5. In "ecocyc.org" this information can be found by searching for "glycolysis I".

Let's go to "glycolysis I" and click on "More Detail". Remember that in order to go from glucose to glucose-6-phosphate one ATP molecules had to be turned into ADP. This corresponds to step one in fig. 5.2Constructing the Cell: Managing the Mass and Energy Budget of the Cellfigure.5.2. Now, from glucose-6-phosphate we go to fructore-6-phosphate reversibly (step 2 in fig. 5.2Constructing the Cell: Managing the Mass and Energy Budget of the Cellfigure.5.2). Notice the double arrow between these two molecules in "glycolysis I". In order to continue into fructose-1,6-biphosphate we pay one more ATP molecules (step 3 in fig. 5.2Constructing the Cell: Managing the Mass and Energy Budget of the Cellfigure.5.2). In step 5 we go to two molecules of glyceraldehyde-3-phosphate reversibly.

From this point on we obtain one NADH molecule, two ATP molecules and one pyruvate molecules per glyceraldehyde-3-phosphate molecule. In conclusion, we started from one glucose molecule and ended up with two pyruvate molecules.

The total ATP cost is

$$\text{ATP cost} \quad = \quad \underbrace{+1}_{\text{glucose}\rightarrow\text{glucose-6-phosphate}} \tag{5.1}$$

$$\underbrace{+1}_{\text{fructose-6-phosphate}\rightarrow\text{fructose-1,6-phosphate}}$$

$$+2 \times (\quad \underbrace{-1}_{\text{fructose-6-phosphate}\rightarrow\text{1,3-bisphosphoglycerate}}$$

$$\underbrace{-1}_{\text{phosphoenolpyruvate}\rightarrow\text{pyruvate}} \quad) = -2,$$

which is a net ATP gain of one per pyruvate molecule. We also gained one NADH molecule per pyruvate in the conversion from glyceraldehyde-3-phosphate into 1,3-bisphosphoglycerate.

Finally, we turn to the synthesis on valine itself. We see that we start off with two pyruvate molecules and that in the process of getting to valine we had to pay one NADPH molecule. We also had to transform glutamate into 2-ketoglutarate. The cell will have to pay an extra energy cost to turn this 2-ketoglutarate back into glutamate.

Let's go to "glutamate biosynthesis III". Notice that the precursor for glutamate is in fact 2-ketoglutarate. In order to obtain it one NADPH molecule and an ammonia molecule have to be paid. Inorganic ions such as phosphate and ammonium are present in bacterial growth medium, which always contains a mixture of salts as well as a carbon source such as glucose. As noted in table 5.1Constructing the Cell: Managing the Mass and Energy Budget of the Celltable.5.1, however, we ignore the cost of phosphate and ammonium.

The net cost of making valine out of two pyruvate molecules is two NADPH molecules. If we add the cost of making two pyruvate molecules it results in

- One glucose molecule consumed.

- Two ATP molecules gained.

- Two NADH molecules gained and two NDAPH molecules consumed.

Which means that the net cost of making a valine molecule is -2 ATP molecules or $-40 \; k_B T$ (energy is gained).

Chapter 6

Entropy Rules!

6.3. Statistical mechanics of an optical trap

In this problem, we are asked to find the mean excursion $\langle x^2 \rangle$ of the bead in the trap. For simplicity, we are pretending that the bead can move in only one dimension and moves on the energy landscape $U(x) = \frac{1}{2}kx^2$. To find averages in statistical mechanics, we follow the general prescription $\langle f(x) \rangle = \sum_x f(x)p(x)$, where we are being intentionally vague about what x might be. For the case of interest here, the key point is to evaluate

$$\langle x^2 \rangle = \frac{1}{Z} \int_\infty^\infty x^2 e^{-\beta \frac{1}{2}kx^2} dx, \tag{6.1}$$

where

$$Z = \int_\infty^\infty e^{-\beta \frac{1}{2}kx^2} dx. \tag{6.2}$$

Hence, the reader is really asked to perform two tasks in this problem. First, to justify the claims above more carefully and second, to perform the relevant calculations to actually find how the stiffness depends upon $\langle x^2 \rangle$. Note that in our definition of Z, we have been sloppy in the sense that there are prefactors we have ignored since they are not relevant to the actual solution of the problem as a result of the fact that they drop out of the problem.

6.4. Free vs. bound ligand

We start by considering the different species involved in a reaction of the form

$$L + R \rightleftharpoons LR. \tag{6.3}$$

The total concentrations of each of these species can be written in terms of the free concentrations as

$$R_{tot} = R + LR \tag{6.4}$$

$$L_{tot} = L + LR \tag{6.5}$$

$$K_d = \frac{L \cdot R}{LR}. \tag{6.6}$$

$$\tag{6.7}$$

As we have seen repeatedly, the probability of finding a ligand bound to a receptor is given by

$$p_{bound} = \frac{LR}{R + LR} = \frac{1}{R/LR + 1} = \frac{1}{K_d/L + 1}. \tag{6.8}$$

From eqn. 6.8 we see that we need L in terms of L_{tot}, R_{tot}, and K_d. We do this by using eqns. 6.4 and 6.5 to rewrite eqn. 6.6

$$
\begin{aligned}
LR &= L_{tot} - L \\
R &= R_{tot} - LR = R_{tot} - L_{tot} + L \\
K_d &= \frac{L \cdot (R_{tot} - L_{tot} + L)}{L_{tot} - L},
\end{aligned}
\tag{6.9}
$$

which we can solve for L which can then be substituted backed into the expression for p_{bound}.

6.9. Lattice model of the chemical potential   In this problem, we are interested in making a toy model of a solution so that we can evaluate the free energy difference in the system between the state where there are N_s solutes and the state where there are $N_s + 1$ solutes. If we imagine a lattice model for the solution with a total Ω sites on the lattice, then the free energy of our N_s solutes will be

$$G(N_s) = N_s \epsilon_s - k_B T \ln \frac{\Omega!}{N_s!(\Omega - N_s)!}, \tag{6.10}$$

where the first term tells us the enthalpy cost associated with putting a solute molecule in solution and the second term is the entropy associated with all the different ways those solute molecules can be distributed in the solution. Now to finish the problem requires two steps: 1) use the Stirling approximation to simplify the expression for the entropy, 2) take the difference in the free energy for the states in which there are N_s and $N_s + 1$ solutes. An alternative scheme is to use the expression given above and to compute the difference directly and use the standard algebraic rules for logarithms to simplify the resulting expression.

Chapter 7

Two-State Systems: From Ion Channels to Cooperative Binding

7.5. Carbon monoxide and hemoglobin.

The general expression appropriate for carrying out the numerical part of the problem is given by

$$p_{O_2} = \frac{\left(\frac{[O_2]}{K_d^{O_2}}\right)^{n_{O_2}}}{1 + \left(\frac{[O_2]}{K_d^{O_2}}\right)^{n_{O_2}} + \left(\frac{[CO]}{K_d^{CO}}\right)^{n_{CO}}}, \tag{7.1}$$

with a similar expression for the binding of CO. It is worth exploring how expressions like this arise by using simple lattice models like those explored in the chapter.

Chapter 8

Random Walks and the Structure of the Macromolecules of the Cell

8.7. Chromosome tethering.

We need to solve the partial differential equation

$$\frac{\partial G}{\partial N} = \frac{b^2}{6}[\frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial G}{\partial r}) + \frac{1}{r^2}\frac{\partial^2 G}{\partial \theta^2} + \frac{\partial^2 G}{\partial z^2}], \tag{8.1}$$

where G is a function of the arguments r, θ, z and N. The strategy the reader should adopt is separation of variables by adopting a trial solution of the form

$$G(r, \theta, z, N) = R(r)\Theta(\theta)Z(z)n(N). \tag{8.2}$$

In light of this trial solution, we have

$$\frac{1}{N}\frac{dn}{dN} = \frac{b^2}{6}[\frac{1}{R}\frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial R}{\partial r}) + \frac{1}{r^2}\frac{1}{\Theta}\frac{\partial^2\Theta}{\partial\theta^2} + \frac{1}{Z}\frac{\partial^2 Z}{\partial z^2}] \tag{8.3}$$

The concept with separation of variables is that the two sides are functions of different variables but are equal to each other. The only way to guarantee that is true is that both sides are equal to the same constant. For example, for the case shown above, we have both equal to a separation constant $-k^2$, resulting in

$$\frac{dn}{dN} = -k^2 N, \tag{8.4}$$

which implies in turn

$$n(N) = e^{-k^2 N}. \tag{8.5}$$

We can once again banish different terms to different sides of the equation resulting in

$$-\frac{b^2}{6}\left[\frac{1}{R}\frac{1}{r}\frac{\partial}{\partial r}(r\frac{\partial R}{\partial r})+\frac{1}{r^2}\frac{1}{\Theta}\frac{\partial^2\Theta}{\partial\theta^2}-k^2=\frac{b^2}{6}\frac{1}{Z}\frac{\partial^2 Z}{\partial z^2}.\right. \tag{8.6}$$

Once again, given that we have a function of r and θ on the left side and a function of z on the right side, both sides must be equal to a constant which this time we christen $-m^2$. This process is repeated until we have solved four separate differential equations for the four unknown functions. Next, the reader needs to impose boundary conditions which amounts to making sure that the solution vanishes on the surface of the cylinder. This leads us to the ability to write the solution as

$$\begin{aligned}
G(r,\theta,z,N) &= \sum_{p=0}^{\infty}\sum_{j=1}^{\infty}\sum_{q=1}^{\infty}e^{-(\frac{\alpha_{pj}^2 b^2}{6a^2}+\frac{q^2\pi^2 b^2}{6L^2})N}J_p(\frac{\alpha_{pj}}{a}r) \\
&\quad\times\sin\frac{q\pi z}{L}(A_{pjq}\cos p\theta + B_{pjq}\sin p\theta).
\end{aligned} \tag{8.7}$$

To determine the constants A_{pjq} and B_{pjq}, we now need to appeal to our "initial condition" which tells us the origin of the polymer chain (i.e. the tether point). In particular, we have

$$G(r,\theta,z,0)=\sum_{p=0}^{\infty}\sum_{j=1}^{\infty}\sum_{q=1}^{\infty}J_p(\frac{\alpha_{pj}}{a}r)\sin\frac{q\pi z}{L}(A_{pjq}\cos p\theta + B_{pjq}\sin p\theta). \tag{8.8}$$

To determine the coefficients, we resort to the use of the orthonormality of the basis functions by writing

$$\begin{aligned}
\int_0^{2\pi}\int_0^L\int_0^a &G(r,\theta,z,0)\sin\frac{q'\pi z}{L}J_{p'}(\frac{\alpha_{p'j'}}{a}r)\cos p'\theta\; rdrdzd\theta = \tag{8.9} \\
&=\sum_{p=0}^{\infty}\sum_{j=1}^{\infty}\sum_{q=1}^{\infty}A_{pjq}\times\int_0^a J_{p'}(\frac{\alpha_{p'j'}}{a}r)J_p(\frac{\alpha_{pj}}{a}r)rdr \\
&\quad\times\int_0^L\sin\frac{q\pi z}{L}\sin\frac{q'\pi z}{L}dz\int_0^{2\pi}\cos p\theta\,\cos p'\theta d\theta.
\end{aligned}$$

The integrals on the right are all straightforward since each of them reflects a convenient orthogonality relation of the form

$$\int_0^a J_{p'}(\frac{\alpha_{p'j'}}{a}r)J_p(\frac{\alpha_{pj}}{a}r)rdr=\delta_{pp'}\delta_{jj'}\frac{a^2}{2}J_{p+1}(\alpha_{pj})^2, \tag{8.10}$$

and

$$\int_0^L\sin\frac{q\pi z}{L}\sin\frac{q'\pi z}{L}dz=\delta_{qq'}\frac{L}{2} \tag{8.11}$$

and

$$\int_0^{2\pi}\cos p\theta\,\cos p'\theta d\theta=\delta_{pp'}\pi. \tag{8.12}$$

As was done in the chapter, to get the probability we need

$$p(r, \theta, z, N) = \frac{G(r, \theta, z, N)}{\int_0^a \int_0^L \int_0^{2\pi} G(r, \theta, z, N) \ r dr d\theta dz}. \tag{8.13}$$

Once the reader does this algebra, the remaining step for making contact with the data is to figure out how to collapse the probability distribution so it is only a function of z or the perpendicular distance from the axis of the cell.

8.8. Three-dimensional random walk and polymer cyclization.
To get the cyclization probability start by expressing the number of walks that return to the origin using the Krönecker delta function $\delta\mathbf{l}$ as

$$\Omega_\circ = \sum_{\{\mathbf{l}_i\}} \delta(\mathbf{l}_1 + \mathbf{l_2} + \mathbf{l}_3 + \ldots + \mathbf{l}_N). \tag{8.14}$$

The sum goes over all possible choices, which for a three-dimensional walk are: $\mathbf{l}_i \in \{\pm\hat{\mathbf{x}}, \pm\hat{\mathbf{y}}, \pm\hat{\mathbf{z}}\}$ for each of the N steps, $i = 1, 2, \ldots, N$. The function $\delta\mathbf{l}$ returns 1 if $\mathbf{l}$ is zero, otherwise the value of the function is 0.

The next step is to use the mathematical identity

$$\delta(\mathbf{l}) = \frac{1}{2\pi} \int_0^{2\pi} e^{il_x k_x} dk_x \cdot \frac{1}{2\pi} \int_0^{2\pi} e^{il_y k_y} dk_y \cdot \frac{1}{2\pi} \int_0^{2\pi} e^{il_z k_z} dk_z \ , \tag{8.15}$$

where the vector $\mathbf{l} = (l_x, l_y, l_z)$ has integer coordinates. This will lead to an expression for the number of random walks that return to the origin written as an integral from which the scaling with N can be derived.

8.12. Scaling of Protein Size
The PDB files with the crystal structures of myoglobin (1MBO), hemoglobin (1HGA), bovine pancreatic trypsin inhibitor (1PIT), lysozyme (132L), cytochrome C (3CYT), G-actin (1ATN), and an α-*beta* tubulin dimer (1TUB) can be found on the book website.

Chapter 9

Electrostatics for Salty Solutions

9.5. A simple model for viral capsid assembly.
To compute the entropy for the n-mers use a lattice model of an ideal solution which consists of $\Omega_n = V/v_n$ boxes among which the N_n n-mers can be distributed. Combine this entropy with the n-mer energy to obtain the free energy for part (a).

9.11. Toy model of salt dependent ligand-receptor binding.
(a) In order to obtain the free energies of the bound and unbound states use the chemical potential for the ligands and the ions, assuming that they make up an ideal gas. Since the free energy cost of removing any one of these molecular species from solution is given by its chemical potential, the concentration dependence of the chemical potential then reveals the dependence of the equilibrium constant on salt concentration.
(b) In this case make use of the free energy of a charged ball in a salty solution to account for the free energy of the unbound state of the ligand and receptor pair.

Chapter 10

Beam Theory: Architecture for Cells and Skeletons

10.1. Persistence length and Fourier analysis.

The Fourier expansion of the angle $\theta(s)$, that takes into account of the boundary conditions, can be written as

$$\theta(s) = \sum_{n=1}^{\infty} \tilde{\theta}_{2n-1} \sin\left(\frac{(2n-1)\pi s}{2L}\right) \; . \tag{10.1}$$

Putting this into the equation for the beam energy leads to an expression for the beam energy in terms of the amplitudes $\tilde{\theta}_{2n-1}$. The average values of the amplitudes can in turn be obtained from equipartition. The idea here is that the energy of the beam acquires the form of a sum over independent modes (each one having an energy quadratic in the amplitude), and in thermal equilibrium each mode has an average energy of $k_B T$.

Chapter 11

Biological Membranes: From the Golgi Apparatus to Membrane Fusion

11.7. Bending modulus and the pippete aspiration experiment.
It is much more convenient to carry out our analysis in q-space. Our Fourier transform convention is

$$h(\mathbf{x}) = \frac{A}{(2\pi)^2} \int \tilde{h}(\mathbf{q}) e^{-i\mathbf{q}\cdot\mathbf{x}} \mathrm{d}^2\mathbf{q}. \tag{11.1}$$

The goal is to write the energy in Fourier transformed variables as

$$E_{tot} = \frac{1}{2} \left(\frac{A}{2\pi}\right)^2 \int |\tilde{h}(\mathbf{q})|^2 \left(\kappa_b q^4 + \sigma q^2\right) \mathrm{d}^2\mathbf{q} \tag{11.2}$$

This follows by using the real space energy and Fourier transforming. To do so, you will need to use the tricks for Fourier transforming terms with derivatives which can be obtained by integrating by parts. The next step is to use the equipartition theorem to show

$$\langle h(\mathbf{q})^2 \rangle = \frac{k_B T}{A(\kappa_b q^4 + \sigma q^2)} \tag{11.3}$$

Using the above results we can write the average difference betwen real and projected area as:

$$\langle A_{act} - A \rangle = \frac{1}{2} \frac{A^2}{(2\pi)^2} \int q^2 \langle |\tilde{h}(\mathbf{q}|^2 \rangle \mathrm{d}^2\mathbf{q} \tag{11.4}$$

By substituting the result for $\langle h(\mathbf{q})^2 \rangle$, we can then find a relation between the tension and the area change.

11.8. Variational approach to deformation induced by MscL.
The first point is to obtain the partial differential equation describing the defor-
mation field. From the chapter, we know that the free energy associated with
the hydrophobic mismatch is of the form

$$G_{hydrophobic} = \frac{K_b}{2} \int (\nabla^2 u)^2 d^2\,r + \frac{K_t}{2} \int \left(\frac{u}{a}\right)^2 d^2\,r.$$

In order to determine the Euler-Lagrange equations, we need to evaluate

$$\frac{\delta G[u(r)]}{\delta u(r)} = lim_{\epsilon \to 0} \frac{G_{hydrophobic}(u(r) + \epsilon\eta(r)) - G_{hydrophobic}(u(r))}{\epsilon} = 0.$$

$$(11.5)$$

What this means precisely is that the reader should literally take the free energy
expression given in eqn. 11.5 and evaluate the difference called for in eqn. 11.5.
There are many separate terms to consider and the objective is to beat them
all into the form
$$\frac{\delta G[u(r)]}{\delta u(r)} = \int_R^\infty \eta \times \text{stuff } dr = 0. \qquad (11.6)$$

Effectively, the differential equation of interest will be the part we have labeled
"stuff" above.

The differential equation can be shown to be related to a Bessel differential
equation. The most intuitive idea is to recognize that problems with cylindrical
symmetry like that of our ion channel often lead to equations that can be beaten
into the form of Bessel's differential equation which is of the form

$$r^2 \frac{d^2 u}{dr^2} + r \frac{du}{dr} + k^2 r^2 u = 0, \qquad (11.7)$$

or to the modified Bessel equation which can be written as

$$r^2 \frac{d^2 u}{dr^2} + r \frac{du}{dr} - k^2 r^2 u = 0. \qquad (11.8)$$

Note that these two equations are *not* the most general form of the Bessel or
modified Bessel equation, but suffice for our purposes. The details of how to
solve such an equation itself are beyond the scope of this book, but we will
show the relation between the differential equation we need to solve given in
eqn. 11.172 and the Bessel equation. In particular, we can write the modified
Bessel equation as
$$\nabla^2 u = k^2 u. \qquad (11.9)$$

If we now apply ∇^2 to both sides of this equation, we are left with

$$\nabla^4 u = k^2 \nabla^2 u, \qquad (11.10)$$

but because of the previous equation $\nabla^2 u = k^2 u$, this can be rewritten as

$$\nabla^4 u = k^4 u, \qquad (11.11)$$

which is precisely our eqn. 11.172 as long as we make the identification

$$k = (-\frac{K_t}{K_b a^2})^{\frac{1}{4}} \tag{11.12}$$

As appeared in the chapter itself, we have four allowed values of k corresponding to the four fourth roots of -1 and given by

$$k_1 = \sqrt[4]{\frac{K_t}{K_b a^2}} e^{i\pi/4} = \sqrt[4]{\frac{K_t}{K_b a^2}} \left(\frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2}\right) \tag{11.13}$$

$$k_2 = \sqrt[4]{\frac{K_t}{K_b a^2}} e^{i3\pi/4} = \sqrt[4]{\frac{K_t}{K_b a^2}} \left(-\frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2}\right) \tag{11.14}$$

$$k_3 = \sqrt[4]{\frac{K_t}{K_b a^2}} e^{i5\pi/4} = \sqrt[4]{\frac{K_t}{K_b a^2}} \left(-\frac{\sqrt{2}}{2} - i\frac{\sqrt{2}}{2}\right) \tag{11.15}$$

$$k_4 = \sqrt[4]{\frac{K_t}{K_b a^2}} e^{i7\pi/4} = \sqrt[4]{\frac{K_t}{K_b a^2}} \left(\frac{\sqrt{2}}{2} - i\frac{\sqrt{2}}{2}.\right) \tag{11.16}$$

The solutions to our differential equation can be written down directly as modified Bessel functions of the form $K_0(kr)$ and $I_0(kr)$, though from the outset we can reject the solutions $I_0(kr)$ since at large r they diverge, inconsistent with our boundary conditions that $u(\infty) = 0$ and $u'(\infty) = 0$. Further, $K_0(k_2 r)$ and $K_0(k_3 r)$ diverge for large r. As a result, we write our solution to the problem as

$$u(r) = A_1 K_0(k_1 r) + A_4 K_0(k_1^* r) \tag{11.17}$$

To determine the coefficients A_1 and A_4, we need to apply our boundary conditions $u(R) = U_0$ and $u'(R) = 0$. The displacement boundary condition at the protein edge results in

$$u(R) = A_1 K_0(k_1 R) + A_4 K_0(k_1^* R) = U_0. \tag{11.18}$$

Similarly, we can evaluate the slope at the protein boundary resulting in

$$u'(R) = A_1 \frac{dK_0(k_1 r)}{dr} + A_4 \frac{dK_0(k_1^* r)}{dr} = 0. \tag{11.19}$$

Using the fact that $K_0'(x) = -K_1(x)$, we can rewrite this boundary condition in the more convenient form

$$u'(R) = -A_1 k_1 K_1(k_1 R) - A_4 k_1^* K_1(k_1^* R) = 0. \tag{11.20}$$

We can now solve these two equations for A_1 and A_4 resulting in

$$A_1 = U_0 \frac{k_1 K_1(k_1 R)}{k_1^* K_0(k_1 R) K_1(k_1^* R) - k_1 K_0(k_1^* R) K_1(k_1 R)}, \tag{11.21}$$

and

$$A_4 = -U_0 \frac{k_1^* K_1(k_1^* R)}{k_1^* K_0(k_1 R) K_1(k_1^* R) - k_1 K_0(k_1^* R) K_1(k_1 R)}. \tag{11.22}$$

Using these values for A_1 and A_4, we can now plug them back into eqn. 11.17 to yield the deformation profile around the ion channel.

One way to obtain the energy is to use the divergence theorem which permits us to convert the energy over the entire membrane into an integral just around the boundary of the protein itself. Effectively, the divergence theorem in this form amounts to a sophisticated version of integration by parts. The theorem can be stated as

$$\int_{\partial\Omega} \mathbf{F} \cdot \mathbf{n} \, ds = \int_{\Omega} \nabla \cdot \mathbf{F} \, dA. \tag{11.23}$$

Our aim is to figure out how the free energy of deformation can be written such that it is clear that the right side is the divergence of some vector. In particular, our goal is to rewrite the term $(\nabla^2 u)^2$ which can be built around the observation that

$$\nabla \cdot (\nabla^2 u \nabla u) = \nabla^3 u \cdot \nabla u + (\nabla^2 u)^2. \tag{11.24}$$

In addition, we have

$$\nabla \cdot (u \nabla^3 u) = \nabla^3 u \cdot \nabla u + u \nabla^4 u, \tag{11.25}$$

which means that if we subtract these two equations, we can isolate $(\nabla^2 u)^2$ as

$$(\nabla^2 u)^2 = -\nabla \cdot (u \nabla^3 u - \nabla^2 u \nabla u) + u \nabla^4 u. \tag{11.26}$$

We can now plug this way of rewriting $(\nabla^2 u)^2$ back into the expression for the free energy of deformation. The result is

$$G_{hydrophobic} = \frac{K_b}{2} \int [-\nabla \cdot (u\nabla^3 u - \nabla^2 u \nabla u) + u\nabla^4 u] d^2 \, r + \frac{K_t}{2} \int \left(\frac{u}{a}\right)^2 d^2 \, r.$$

This very interesting result brings us much closer to having an expression that will permit us to invoke the divergence theorem. In particular, if we rewrite this as

$$G_{hydrophobic} = \frac{K_b}{2} \int [-\nabla \cdot (u\nabla^3 u - \nabla^2 u \nabla u)] d^2 r + \int (\frac{K_t}{2a^2} u + \frac{K_b}{2} \nabla^4 u) u d^2 r, \tag{11.27}$$

we note that the term in parentheses in the second integral is nothing more than the equilibrium differential equation we solved in the first place and is thus zero. Hence, we can now rewrite the energy as

$$G_{hydrophobic} = \frac{K_b}{2} \int [-\nabla \cdot (u\nabla^3 u - \nabla^2 u \nabla u)] d^2 r = -\frac{K_b}{2} \int \mathbf{r} \cdot (u\nabla^3 u - \nabla^2 u \nabla u) r d\theta. \tag{11.28}$$

Chapter 13

Random Walks: A Statistical View of Biological Dynamics

13.2. Diffusion from a point source.

The diffusion equation reads:

$$\frac{\partial}{\partial t}c(x,t) = D\frac{\partial^2}{\partial x^2}c(x,t). \tag{13.1}$$

Let's take the Fourier transform of both sides (which just means multiplying by $e^{ikx}/\sqrt{2\pi}$ and integrating) and take stock of the situation:

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{ikx}\frac{\partial}{\partial t}c(x,t)\,\mathrm{d}x = \frac{D}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{ikx}\frac{\partial^2}{\partial x^2}c(x,t)\,\mathrm{d}x \tag{13.2}$$

$$\frac{\partial}{\partial t}\left(\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{ikx}c(x,t)\,\mathrm{d}x\right) = \frac{D}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{ikx}\frac{\partial^2}{\partial x^2}c(x,t)\,\mathrm{d}x \tag{13.3}$$

$$\frac{\partial}{\partial t}\tilde{c}(k,t) = \frac{D}{\sqrt{2\pi}}\int_{-\infty}^{+\infty} e^{ikx}\frac{\partial^2}{\partial x^2}c(x,t)\,\mathrm{d}x. \tag{13.4}$$

All that we've done so far is to exchange the order of integration and differentiation on the LHS and then used the definition of $\tilde{c}$. Now comes the important part. The key is to integrate by parts on the RHS to get the partial derivatives w.r.t x off of $c(x,t)$ and move them onto the exponential instead. How is this done? Recall that for two functions F and G that both depend on x the following identity holds (the product rule):

$$\frac{\partial}{\partial x}(FG) = \frac{\partial F}{\partial x}G + F\frac{\partial G}{\partial x}. \tag{13.5}$$

Now rearrange things and integrate:

$$\int_{-\infty}^{+\infty} F \frac{\partial G}{\partial x} \, dx \;=\; \int_{-\infty}^{+\infty} \frac{\partial}{\partial x}(FG) \, dx - \int_{-\infty}^{+\infty} \frac{\partial F}{\partial x} G \, dx \qquad (13.6)$$

$$=\; (FG)\Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} \frac{\partial F}{\partial x} G \, dx. \qquad (13.7)$$

Let's suppose that either or both of F and G vanish at $\pm\infty$, then we have:

$$\int_{-\infty}^{+\infty} F \frac{\partial G}{\partial x} \, dx = - \int_{-\infty}^{+\infty} \frac{\partial F}{\partial x} G \, dx. \qquad (13.8)$$

This is just what we wanted; we've succeeded in moving the derivative off of G on the LHS and onto F on the RHS, at the cost of a minus sign. Let's go back to Eq. 13.4 and integrate by parts:

$$\frac{\partial}{\partial t}\tilde{c}(k,t) \;=\; \frac{D}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{ikx} \frac{\partial^2}{\partial x^2} c(x,t) \, dx \qquad (13.9)$$

$$=\; \left(\frac{D}{\sqrt{2\pi}} e^{ikx} \frac{\partial c}{\partial x} \right)\Big|_{-\infty}^{+\infty}$$

$$\quad - \frac{D}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \left(\frac{\partial}{\partial x} e^{ikx} \right) \frac{\partial}{\partial x} c(x,t) \, dx \qquad (13.10)$$

$$=\; -\frac{D}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (ik) e^{ikx} \frac{\partial}{\partial x} c(x,t) \, dx \qquad (13.11)$$

$$=\; \left(-\frac{Dik}{\sqrt{2\pi}} e^{ikx} c(x,t) \right)\Big|_{-\infty}^{+\infty} + \frac{D}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (ik) \qquad (13.12)$$

$$\quad \times \left(\frac{\partial}{\partial x} e^{ikx} \right) c(x,t) \, dx$$

$$=\; +\frac{D}{\sqrt{2\pi}} (ik)^2 \int_{-\infty}^{+\infty} e^{ikx} c(x,t) \, dx \qquad (13.13)$$

$$=\; -k^2 D \tilde{c}(k,t). \qquad (13.14)$$

We integrated by parts twice and had to assume that the flux $\partial c/\partial x$ and concentration $c(x,t)$ both vanish at infinity which are both very reasonable assumptions. At first glance it may seem like we haven't gained that much. But hold on! Instead of having derivatives w.r.t. both x and t, we now have derivatives w.r.t. time t only! We have transformed our PDE into an ODE; $\tilde{c}$ is a function of k but for our purposes it's just a parameter; let's rewrite our equation to make this all explicit:

$$\frac{d}{dt}\tilde{c}(t;k) = -k^2 D \, \tilde{c}(t;k). \qquad (13.15)$$

Of course this is an equation we know very well how to solve, and we can just write down the answer:

$$\tilde{c}(k,t) = \tilde{c}(k,0) e^{-k^2 Dt}. \qquad (13.16)$$

To get $\tilde{c}(k,0)$ we need to Fourier transform our initial condition. Once we do that, all that remains effectively is to perform the inverse Fourier transform using the

$$
\begin{aligned}
c(x,t) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-ikx} \tilde{c}(x,t)\, \mathrm{d}k && (13.17) \\
&= \frac{c_0}{2\pi} \int_{-\infty}^{+\infty} e^{-ikx} e^{-k^2 Dt}\, \mathrm{d}k. && (13.18)
\end{aligned}
$$

13.5. Two-Dimensional FRAP Analysis

In order to solve this problem we will use separation of variables. We will assume that $c(r,t) = T(t)\rho(r)$, which turns our diffusion equation of interest into

$$
\frac{1}{D\,T}\frac{dT}{dt} = \frac{1}{\rho}\frac{1}{r}\frac{d}{dr}\left(r\frac{d\rho}{dr}\right) = \frac{1}{\rho}\frac{1}{r}\left(\frac{d\rho}{dr} + r\frac{d^2\rho}{dr^2}\right). \tag{13.19}
$$

Since both sides of the equation depend on different variables, they must be equal to a constant, which we set equal to K^2. This means that the solution for the temporal part is

$$
T(t) = e^{-DK^2 t} \tag{13.20}
$$

and the resulting equation for the spatial part

$$
r^2\frac{d^2\rho}{dr^2} + r\frac{d\rho}{dr} + K^2 r^2 \rho = 0. \tag{13.21}
$$

This last equation can be written in a Bessel form by making the substitution $z = kr$. Once the solutions are in hand, the boundary conditions need to be imposed. In particular, we have the following expression for the boundary condition

$$
J_0'(kR) = 0. \tag{13.22}
$$

The roots of J_0' are just the roots of J_1 because of the identity $J_0'(z) = -J_1$. This will determine discrete values for K_i. The full solution is

$$
c(r,t) = a_0 + \sum_{i=1}^{\infty} a_i e^{-DK_i^2 t} J_0(K_i r). \tag{13.23}
$$

We can determine the coefficients a_i using the initial condition $c(r,0)$.

Chapter 15

Rate Equations and Dynamics in the Cell: Cytoskeleton Under Construction

15.6. Microtubule dynamics and dynamic instability.
To write down the master equations we have to consider all the possible transitions to and from the state of interest. For example, in the case of writing $p_+(n,t)$, which is the probability of finding a growing filament of length n at time t we have the following options:

- A filament with length $n-1$ which grows into a filament with length n. This process happens with a speed v_+ and its contribution to the master equation is $+v_+ p_+(n-1,t)$.

- The filament with length n might grow into a filament with length $n+1$, contributing with a term $-v_+ p_+(n,t)$.

- A filament that is growing can undergo a "catastrophe", which turns it into a shrinking one and adds the term $-f_{+-} p_+(n,t)$.

- Finally, a filament that was shrinking can undergo a "rescue" and turn into a growing one, giving the term $+f_{-+} p_-(n,t)$.

the master equation for a growing filament is therefore:

$$\frac{\partial p_+(n,t)}{\partial t} = v_+(p_+(n-1,t) - p_+(n,t)) - f_{+-}p_+(n,t) + f_{-+}p_-(n,t) \quad (15.1)$$

We can now do a Taylor expansion of the term with $p_+(n-1,t) \simeq p_+(n,t) - \frac{\partial p_+(n,t)}{\partial n}$,

To solve the differential equations that you will obtain from the Taylor expansion strategy, assume steady state (i.e. $\frac{\partial p_+(n,t)}{\partial t} = \frac{\partial p_-(n,t)}{\partial t} = 0$. This implies in turn that

$$v_+ \frac{dp_+(n,t)}{dn} = f_{-+}p_- - f_{+-}p_+ \tag{15.2}$$

$$v_- \frac{dp_-(n,t)}{dn} = f_{-+}p_- - f_{+-}p_+.$$

This is a linear system of first order differential equations. We can write the RHS of both equations in matrix form using

$$\begin{pmatrix} -f_{+-}/v_+ & f_{-+}/v_+ \\ -f_{+-}/v_- & f_{-+}/v_- \end{pmatrix} \tag{15.3}$$

We want to look for its eigenvalues and eigenvectors which will permit us to write down the solutions themselves.

15.8. Antenna model for microtubule length control.
(a) Every monomer of the filament can be characterized by its Kip3 occupancy, which we denote n_i; here $i = 1, 2, \ldots, N$ where $N = L/a$ is the number of monomers in the filament, and a is the length of a single monomer. In steady state the number of Kip3 molecules arriving at the ith monomer per unit time must equal the number of monomers departing the same monomer. In the model we assume that all the monomers in front of the ith one funnel Kip3 molecules to it (i.e. Kip3 molecules do not detach before they fall off the end of the filament) and therefore the balance between incoming and outgoing Kip3 becomes

$$ik_{bind} = vn_i . \tag{15.4}$$

Given these arguments, one can now work out the steady-state occupancy and how it depends upon position on the filament. As with many problems in the book, there are multiple ways to solve the problem. An alternative approach to that described above is to write down a differential equation of the form

$$\frac{dn}{dx} = \frac{k_{bind}}{av} \tag{15.5}$$

which implies that

$$n(x) = \frac{k_{bind}}{av}x = i\frac{k_{bind}}{v}. \tag{15.6}$$

This differential equation arises by writing a mass balance for each monomer and then making a continuum approximation.

(b) The flux of Kip3 into the terminal monomer in steady state is $Nk_{bind} = k_{bind}L/a$, therefore the depolymerization rate for the filament is $k_{off} = aNk_{bind} = k_{bind}L$. The way to see this is to use the result of part (a) which tells us how the number of Kip3 molecules depends upon position.

Chapter 16

Dynamics of Molecular Motors

16.4. Kinetics of two-state motors

Start by taking the continuum limit of eqns. 16.33 and 16.34 so that differences in the variable n become derivatives of the probabilities with respect to this parameter. Then substitute the trial solution into the two differential equations to obtain a two-by-two system of linear equations for C_0 and C_1, where

$$\begin{pmatrix} p_0(n,t) \\ p_1(n,t) \end{pmatrix} = e^{i(Kn-\omega t)} \begin{pmatrix} C_0 \\ C_1 \end{pmatrix} . \qquad (16.1)$$

Now, if we write the equations for C_0 and C_1 in matrix form as

$$\begin{pmatrix} i\omega - (k_A^- + k_B^+) & k_A^+(1 - iK - \frac{1}{2}K^2) + k_B^- \\ k_A^-(1 + iK - -\frac{1}{2}K^2) + k_B^+ & i\omega - (k_A^+ + k_B^-) \end{pmatrix} \begin{pmatrix} C_0 \\ C_1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \qquad (16.2)$$

then the matrix on the left-hand side must have a zero determinant for a solution to exist.

Chapter 17

Biological Electricity and the Hodgkin-Huxley Model

17.6. Time-dependent cable equation

(a) The time-dependent cable equation 17.30 is a partial differential equation. We can solve it numerically using Matlab. Alternatively, Mathematica or any similar program can be used.

The function `pdepe` solves equations of the general form

$$c\left(x,t,u,\frac{\partial u}{\partial x}\right) = x^{-m}\frac{\partial}{\partial x}\left(x^m f\left(x,t,u,\frac{\partial u}{\partial x}\right)\right) + s\left(x,t,u,\frac{\partial u}{\partial x}\right). \quad (17.1)$$

When we compare this to eqn. 17.30 we can make the following identifications

$$u(x,t) \;=\; V(x,t) \quad\quad\quad (17.2)$$

$$c\left(x,t,V(x,t),\frac{\partial V(x,t)}{\partial x}\right) \;=\; -\tau$$

$$m \;=\; 0$$

$$f\left(x,t,V(x,t),\frac{\partial V(x,t)}{\partial x}\right) \;=\; -\lambda^2\frac{\partial V(x,t)}{\partial x}$$

$$s\left(x,t,V(x,t),\frac{\partial V(x,t)}{\partial x}\right) \;=\; \frac{g_{Na}(V(x,t))}{g_K}(V(x,t) - V(x,t)^{Na}_{Nernst}) + (V(x,t) - V(x,t)^{K}_{Nernst}).$$

This function is called in the following way: `sol = pdepe(m,pdefun,icfun, bcfun,xmesh,tspan)`, where `pdefun` returns the values of `c`, `f`, and `s` (`[c,f,s] = pdefun(x,t,u,DuDx)`). The initial conditions are given by

$$u(x,t_0) = u_0(x), \quad\quad\quad (17.3)$$

using `u = icfun(x)`. Finally, the boundary conditions correspond to the equation

$$p(x,t,u) + q(x,t)\,f\left(x,t,u,\frac{\partial u}{\partial x}\right) = 0. \quad\quad\quad (17.4)$$

We input these conditions using `[pl,ql,pr,qr] = bcfun(xl,ul,xr,ur,t)`.
Here, `l` and `r` correspond to the values of `x`, `u`, `p`, and `q` at the left and right
boundaries, respectively.

17.7. Propagation of spikes

Like in the previous problem we can use the function `pdepe` form Matlab. The
biggest difference is that this time we'll use it to solve a system of differen-
tial equations rather than only one differential equation. These equations will
be the cable equation as well as the equations describing the dynamics of the
populations of closed, open and inactive Na^+ channels shown in eqn. 17.44:

$$
\begin{aligned}
\frac{\partial p_C}{\partial t} &= -k_{open}(V)p_C \\
\frac{\partial p_O}{\partial t} &= k_{open}(V)p_C - k_{inactive}(V)p_O \\
\frac{\partial p_I}{\partial t} &= k_{inactive}(V)p_O \\
-\tau\frac{\partial V(x,t)}{\partial t} &= -\lambda^2\frac{\partial^2 V(x,t)}{\partial x^2}\frac{g_{Na}(V)}{g_K}(V(x,t) - V_{Nernst}^{Na}) \\
&\quad +(V(x,t) - V_{Nernst}^{K}).
\end{aligned}
\tag{17.5}
$$

We define different column vectors that will be used in `pdepe`:

$$
\begin{aligned}
u(1) &= p_C(V) \\
u(2) &= p_O(V) \\
u(3) &= p_I(V) \\
u(4) &= V(x,t)
\end{aligned}
\tag{17.6}
$$

Each of these will have corresponding `c`, `f` and `s` column vectors.

You will have to implement eqns. 17.45 and 17.46 in you code. For the
boundary conditions, we still set the derivative of the voltage to be zero. We fix
the channels at the boundaries to the closed state. This will certainly introduce
artifacts if the spikes reach the boundaries in our simulation. Finally, we use
the exact same initial spike as as in problem 17.6 for our initial condition, while
we set all the channels to the closed state initially.

Chapter 18

Light and Life

18.7. Eyes and the diffraction limit

(a) Remember that the vector $\vec{r}$ pointing to the plane of observation is given by

$$\vec{r} = x\hat{x} + y\hat{y} + Z\hat{z}, \tag{18.1}$$

where x and y are the positions along the plane and Z is the distance from the point source to the plane. Notice that we are assuming that $Z \gg x, y$, which means that we can Taylor expand r accordingly.

(b) We want now to write the field at position r' on the plane coming from the patch dS of our source. Using the definitions from Figure 18.39(C) we can write r' as

$$r' = \sqrt{(R\cos\phi - r)^2 + (R\sin\phi)^2 + Z^2}. \tag{18.2}$$

Note that, like in part (a), we can Taylor expand this expression to first order in r. This will allow to obtain a simple functional for the disturbance from dS at the plane given by

$$\psi = \frac{A}{r}e^{-\mathrm{i}(\omega t - \vec{k}\cdot\vec{r})}. \tag{18.3}$$

(c) By integrating the disturbances over the source (over R and ϕ) you will reach an expression of the form

$$\int_0^{2\pi} e^{ia\cos(\phi)}\, d\phi = 2\pi J_0(a), \tag{18.4}$$

where $J_n(x)$ is the n-th Bessel function of the first kind. Another useful identify to keep in mind is the one relating the integral of Bessel functions such as

$$\int J_0(x)x\, dx = xJ_1(x). \tag{18.5}$$

(d) The first zero of the intensity, related to the resolution limit, will be given by the first zero of the first Bessel function of the first kind, $J_1(x)$. The zeros of this

function can be found in tables of integrals or in programs such as Mathematica. In the following table we show the first few zeros for reference.

Zero number for $J_1(x)$	x
1	3.83171
2	7.01559
3	10.1735

Chapter 19

Organization of Biological Networks

19.10. Copy number and the Poisson promoter

(a) f is the fraction of the cell cycle during which there is one copy of the gene. As a result of the change in the number of copies of the gene of interest, the mean number of mRNAs varies depending upon when in the cell cycle we look. The hint is to note that when there are two copies of the gene of interest, then $p(m)$ has a contribution of the form $\frac{(2\lambda)^m e^{-2\lambda}}{m!}$. What fraction of the cell cycle is this the relevant $p(m)$?

(b) We can use the result of part (a) to find the average number of mRNAs as

$$\langle m \rangle = \sum_{m=0}^{\infty} mp(m) = f \sum_{m=0}^{\infty} m\frac{\lambda^m e^{-\lambda}}{m!} + (1-f) \sum_{m=0}^{\infty} m\frac{(2\lambda)^m e^{-2\lambda}}{m!}. \quad (19.1)$$

We can use our usual trick of differentiation with respect to a parameter by remembering that

$$\sum_{m=0}^{\infty} m\frac{\lambda^m e^{-\lambda}}{m!} = e^{-\lambda}\lambda\frac{d}{d\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!}. \quad (19.2)$$

Using this insight, the remainder of the problem comes down to algebra.

(c) Similarly, we can evaluate

$$\langle m^2 \rangle = \sum_{m=0}^{\infty} m^2 p(m) \quad (19.3)$$

using the same tricks as in part (b).

Chapter 21

Sequences, Specificity, and Evolution

21.4. Nucleosome urn and chemical potentials

(a) The chemical potential of nucleosomes and base pairs on genomic DNA is related to the free energy $G(n_{nuc}, n_{bp})$, where n_{nuc} is the number of nucleosomes and n_{bp} the number of free (i.e., not covered by nucleosomes) base pairs. In particular, use the definitions

$$
\begin{aligned}
\mu_{nuc} &= G(n_{nuc}, n_{bp}) - G(n_{nuc} - 1, n_{bp}) \\
\mu_{bp} &= G(n_{nuc}, n_{bp}) - G(n_{nuc}, n_{bp} - 1) \ .
\end{aligned}
\tag{21.1}
$$

With these definitions, the calculations of the chemical potentials comes down to figuring out the entropy for the number of ways of arranging the base pairs and nucleosomes.

(b) With the chemical potentials in hand we turn to computing the probability $p(i)$ that the ith site of a genomic region of length N is occupied by a nucleosome, given that the origin is nucleosome free. This probability, $p(i) = 1 - \bar{p}(i)$, where $\bar{p}(i)$ is the probability that the ith site is nucleosome free (i.e., it is a "base pair" in the sense of the urn model). If we introduce

$$
S(i) = \sum_{n_{nuc}=0}^{n_{nuc}^{max}} \frac{[(i - n_{nuc}d) + n_{nuc}]!}{(i - n_{nuc}d)! n_{nuc}!} e^{\beta(i - n_{nuc}d)\mu_{bp}} e^{\beta n_{nuc}\mu_{nuc}},
\tag{21.2}
$$

as in the text, as the thermodynamic weight that takes into account all the possible ways of covering i sites on the genome with base pairs and nucleosomes, then

$$
\bar{p}(i) = \frac{S(i - 1)S(N - i)}{S(N)} \ .
\tag{21.3}
$$

The numerator in the above expression is the weight of all the states in which the ith site is not covered by a nucleosome while the denominator is the weight

of all the states regardless of whether the ith site is nucleosome free or not. The numerator is obtained by considering the ith site nucleosome free and then noting that this leaves the first $i-1$ sites before it and the $N-i$ sites after it free for occupying with nucleosomes or base pairs. Since occupying these two stretches of genomic DNA with free base pairs and nucleosomes are independent of each other, the total weight is the product of weights, S(i-1) S(N-i).

We can use Mathematica to make plots of $p(i)$ for different values of the average linker length $\langle L \rangle$.