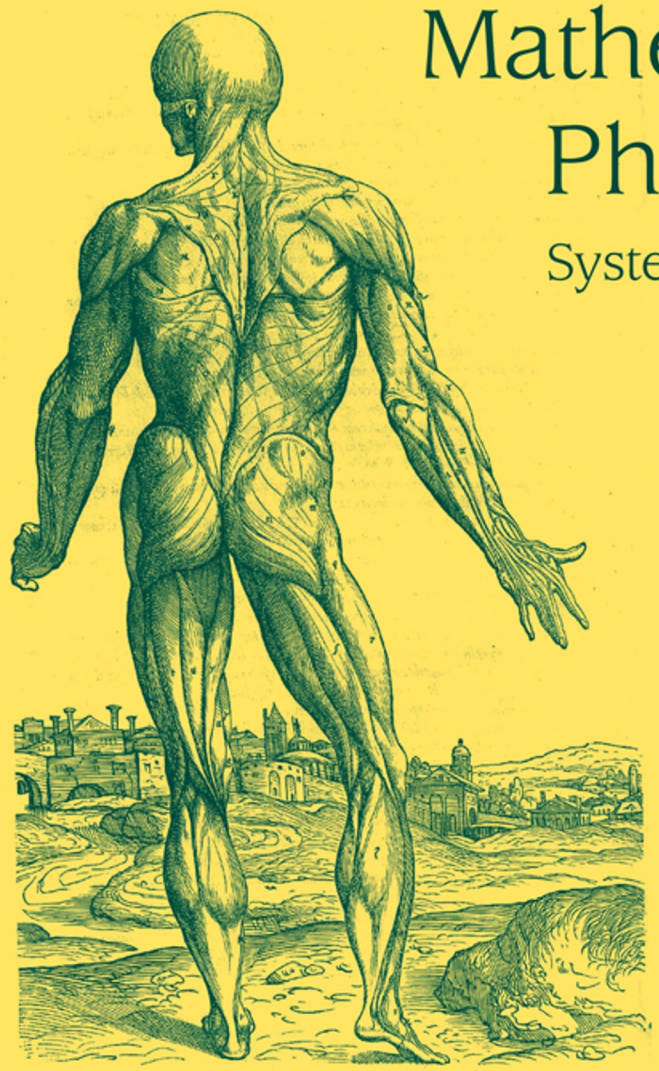
INTERDISCIPLINARY APPLIED MATHEMATICS

MATHEMATICAL BIOLOGY



Mathematical Physiology

Systems Physiology

James Keener
James Sneyd

Second Edition

Springer

James Keener James Sneyd

Mathematical Physiology

I: Cellular Physiology

Second Edition

Springer

James Keener
Department of Mathematics
University of Utah
Salt Lake City, 84112
USA
keener@math.utah.edu

James Sneyd
Department of Mathematics
University of Auckland
Private Bag 92019
Auckland, New Zealand
sneyd@math.auckland.ac.nz

ISBN 978-0-387-75846-6 e-ISBN 978-0-387-75847-3

DOI 10.1007/978-0-387-75847-3

Library of Congress Control Number: 2008931057

Printed on acid-free paper.

springer.com

CHAPTER 2

Cellular Homeostasis

2.1 The Cell Membrane

The cell membrane provides a boundary separating the internal workings of the cell from its external environment. More importantly, it is selectively permeable, permitting the free passage of some materials and restricting the passage of others, thus regulating the passage of materials into and out of the cell. It consists of a double layer (a *bilayer*) of phospholipid molecules about 7.5 nm (75 angstroms) thick (Fig. 2.1). The term *lipid* is used to specify a category of water-insoluble, energy rich macromolecules, typical of fats, waxes, and oils. Irregularly dispersed throughout the phospholipid bilayer are aggregates of globular proteins, which are free to move within the layer, giving the membrane a fluid-like appearance. The membrane also contains water-filled pores with diameters of about 0.8 nm, as well as protein-lined pores, called *channels*, which allow passage of specific molecules. Both the intracellular and extracellular environments consist of, among many other things, a dilute aqueous solution of dissolved salts, primarily NaCl and KCl, which dissociate into Na^+, K^+, and Cl^- ions. The cell membrane acts as a barrier to the free flow of these ions and maintains concentration differences of these ions. In addition, the cell membrane acts as a barrier to the flow of water.

Molecules can be transported across the cell membrane by passive transport or active processes. An active process is one that requires the expenditure of energy, while a passive process results solely from the inherent, random movement of molecules. *Osmosis*, i.e., the diffusion of water down its concentration gradient, is the most important process by which water moves through the cell membrane. Simple diffusion

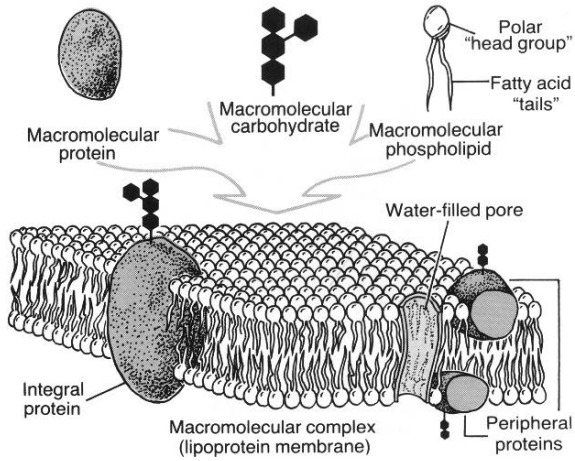


Figure 2.1 Schematic diagram of the cell membrane. (Davis et al., 1985, Fig. 3-1, p. 41.)

accounts for the passage of small molecules through pores and of lipid-soluble molecules through the lipid bilayer. For example, water, urea (a nitrogenous waste product of metabolism), and hydrated Cl^- ions diffuse through membrane pores. Oxygen and carbon dioxide diffuse through the membrane readily because they are soluble in lipids. Sodium and K^+ ions pass through ion-specific channels, driven by diffusion and electrical forces. Some other mechanism must account for the transport of larger sugar molecules such as galactose, glucose, and sucrose, as they are too large to pass through membrane pores.

Concentration differences are set up and maintained by active mechanisms that use energy to pump ions against their concentration gradient. One of the most important of these pumps is the Na^+–K^+ ATPase, which uses the energy stored in ATP molecules to pump Na^+ out of the cell and K^+ in. Another pump, the Ca^{2+} ATPase, pumps Ca^{2+} out of the cell or into the endoplasmic reticulum. There are also a variety of exchange pumps that use the energy inherent in the concentration gradient of one ion type to pump another ion type against its concentration gradient. For example, the Na^+–Ca^{2+} exchanger removes Ca^{2+} from the cell at the expense of Na^+ entry, and similarly for the Na^+–H^+ exchanger. Typical values for intracellular and extracellular ionic concentrations are given in Table 2.1.

Differences in ionic concentrations create a potential difference across the cell membrane that drives ionic currents. Water is also absorbed into the cell because of concentration differences of these ions and also because of other large molecules contained in the cell, whose presence provides an osmotic pressure for the absorption of water. It is the balance of these forces that regulates both the cell volume and the membrane potential.

Table 2.1 Typical values for intracellular and extracellular ionic concentrations, Nernst potentials and resting potentials, from three different cell types. Concentrations are given in units of mM, and potentials are in units of mV. Extracellular concentrations for the squid giant axon are for seawater, while those for frog muscle and red blood cells are for plasma. (Adapted from Mountcastle, 1974, Table 1-1.)

	Squid Giant Axon	Frog Sartorius Muscle	Human Red Blood Cell
Intracellular concentrations			
Na^+	50	13	19
K^+	397	138	136
Cl^-	40	3	78
Mg^{2+}	80	14	5.5
Extracellular concentrations			
Na^+	437	110	155
K^+	20	2.5	5
Cl^-	556	90	112
Mg^{2+}	53	1	2.2
Nernst potentials			
V_{Na}	+56	+55	+55
V_K	−77	−101	−86
V_{Cl}	−68	−86	−9
Resting potentials	−65	−99	−6 to −10

2.2 Diffusion

To keep track of a chemical concentration or any other measurable entity, we must track where it comes from and where it goes; that is, we must write a *conservation law*. If U is some chemical species in some region, then the appropriate conservation law takes the following form (in words):

rate of change of U = rate of production of U + accumulation of U due to transport.

If Ω is a region of space, this conservation law can be written symbolically as

$$\frac{d}{dt}\int_\Omega u\,dV = \int_\Omega f\,dV - \int_{\partial\Omega} \mathbf{J}\cdot\mathbf{n}\,dA, \tag{2.1}$$

where u is the concentration of the chemical species U, $\partial\Omega$ is the boundary of the region Ω, $\mathbf{n}$ is the outward unit normal to the boundary of Ω, f represents the local production density of U per unit volume, and $\mathbf{J}$ is the flux density of U. According to the divergence

theorem, if **J** is sufficiently smooth, then

$$\int_{\partial\Omega} \mathbf{J}\cdot\mathbf{n}\,dA = \int_{\Omega} \nabla\cdot\mathbf{J}\,dV, \tag{2.2}$$

so that if the volume in which u is being measured is fixed but arbitrary, the integrals can be dropped, with the result that

$$\frac{\partial u}{\partial t} = f - \nabla\cdot\mathbf{J}. \tag{2.3}$$

This, being a conservation law, is inviolable. However, there are many ways in which the production term f and the flux **J** can vary. Indeed, much of our study in this book is involved in determining appropriate models of production and flux.

2.2.1 Fick's Law

Suppose that u is a function of a single spatial variable, x, and consider the two situations shown in Fig. 2.2, one where u has a steep gradient, the other with a shallow gradient. It is intuitively reasonable that the flux of u should be greater in magnitude in the first case than in the second, and this is indeed what is found experimentally, provided u is not too large. Thus

$$J = -D\frac{du}{dx}. \tag{2.4}$$

Note the sign of J. By definition, a flux of u from left to right is identified as a positive flux, and thus the flux is opposite in sign to the gradient.

In higher dimensions

$$\mathbf{J} = -D\nabla u. \tag{2.5}$$

Equation (2.5) is called a *constitutive relationship*, and for chemical species it is called *Fick's law*. The scalar D is the *diffusion coefficient* and is characteristic of the solute and the fluid in which it is dissolved. If u represents the heat content of the volume, (2.5) is called *Newton's law of cooling*. Fick's law is not really a law, but is a reasonable

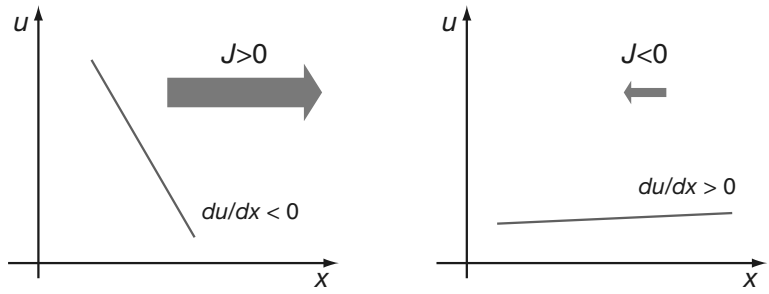


Figure 2.2 Fick's Law. The flux is proportional to the gradient, and opposite in sign.

approximation to reality if the concentration of the chemical species is not too high. When Fick's law applies, the conservation equation becomes the reaction–diffusion equation

$$\frac{\partial u}{\partial t} = \nabla \cdot (D \nabla u) + f, \tag{2.6}$$

or, if D is a constant,

$$\frac{\partial u}{\partial t} = D \nabla^2 u + f. \tag{2.7}$$

The diffusion equation can also be derived from a random walk (Section 2.9.1).

There is a vast literature on reaction–diffusion equations. To mention but a very few, Aronson and Weinberger (1975), Britton (1986) and Grindrod (1991) are biologically oriented, as is Murray (2002), while Smoller (1994) and Fife (1979) are more theoretical presentations.

2.2.2 Diffusion Coefficients

A quantitative understanding of diffusion was given by Einstein (1906) in his theory of Brownian motion. He showed that if a spherical solute molecule is large compared to the solvent molecule, then

$$D = \frac{kT}{6\pi \mu a}, \tag{2.8}$$

where $k = \frac{R}{N_A}$ is Boltzmann's constant, N_A is Avogadro's number, T is the absolute temperature of the solution, μ is the coefficient of viscosity for the solute, and a is the radius of the solute molecule. For nonspherical molecules, Einstein's formula generalizes to

$$D = \frac{kT}{f}, \tag{2.9}$$

where f is the Stokes frictional coefficient of the particle and $f = 6\pi \mu a$ for a sphere of radius a. The molecular weight of a spherical molecule is

$$M = \frac{4}{3}\pi a^3 \rho, \tag{2.10}$$

where ρ is the molecular density, so that, in terms of molecular weight,

$$D = \frac{kT}{3\mu} \left(\frac{\rho}{6\pi^2 M} \right)^{1/3}. \tag{2.11}$$

The density of most large protein molecules is nearly constant (about 1.3−1.4 g/cm^3), so that $DM^{1/3}$ is nearly the same for spherical molecules at a fixed temperature. The diffusion of small molecules, such as the respiratory gases, is different, being proportional to $M^{-1/2}$.

Table 2.2 Molecular weight and diffusion coefficients of some biochemical substances in dilute aqueous solution.

Substance	Molecular Weight	$D(\text{cm}^2/\text{s})$
hydrogen	1	4.5×10^{-5}
oxygen	32	2.1×10^{-5}
carbon dioxide	48	1.92×10^{-5}
glucose	192	6.60×10^{-6}
insulin	5,734	2.10×10^{-6}
Cytochrome c	13,370	1.14×10^{-6}
Myoglobin	16,900	5.1×10^{-7}
Serum albumin	66,500	6.03×10^{-7}
hemoglobin	64,500	6.9×10^{-7}
Catalase	247,500	4.1×10^{-7}
Urease	482,700	3.46×10^{-7}
Fibrinogen	330,000	1.97×10^{-7}
Myosin	524,800	1.05×10^{-7}
Tobacco mosaic virus	40,590,000	5.3×10^{-8}

2.2.3 Diffusion Through a Membrane: Ohm's Law

We can use Fick's law to derive the chemical analogue of Ohm's law for a membrane of thickness L. Suppose that a membrane separates two large reservoirs of a dilute chemical, with concentration c_l on the left (at $x = 0$), and concentration c_r on the right (at $x = L$). According to the diffusion equation, in the membrane (assuming that the only gradients are transverse to the membrane)

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2}, \tag{2.12}$$

subject to boundary conditions $c(0,t) = c_l, c(L,t) = c_r$.

The full time-dependent solution can be found using separation of variables, but for our purposes here, the steady-state solution is sufficient. At steady state, $\frac{\partial c}{\partial t} = 0$, so that $\frac{\partial J}{\partial x} = -D\frac{\partial^2 c}{\partial x^2} = 0$, from which it follows that $J = -D\frac{\partial c}{\partial x} =$ constant, or that $c(x) = ax + b$, for some constants a and b. Applying the boundary conditions, we obtain

$$c(x) = c_l + (c_r - c_l)\frac{x}{L}. \tag{2.13}$$

From Fick's law it follows that the flux of chemical is constant, independent of x, and is given by

$$J = \frac{D}{L}(c_l - c_r). \tag{2.14}$$

The ratio L/D is the effective "resistance" of the membrane, and so D/L is called the *conductance*, or *permeability*, per unit area.

2.2.4 Diffusion into a Capillary

Suppose that a long capillary, open at one end, with uniform cross-sectional area A and filled with water, is inserted into a solution of known chemical concentration C_0, and the chemical species is free to diffuse into the capillary through the open end. Since the concentration of the chemical species depends only on the distance along the tube and time, it is governed by the diffusion equation

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2}, \qquad 0 < x < \infty, \qquad t > 0, \tag{2.15}$$

where for convenience we assume that the capillary is infinitely long. Because the solute bath in which the capillary sits is large, it is reasonable to assume that the chemical concentration at the tip is fixed at $C(0,t) = C_0$, and since the tube is initially filled with pure water we set $C(x,0) = 0$.

The solution of this problem is given by

$$C(x,t) = 2C_0\left(1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}\exp\left(-\frac{s^2}{2}\right)ds\right), \qquad z = \frac{x}{\sqrt{2Dt}}. \tag{2.16}$$

From this, one can easily calculate that the total number of molecules that enter the capillary in a fixed time T is

$$N = A\int_0^\infty C(x,T)\,dx = 2C_0A\sqrt{\frac{TD}{\pi}}. \tag{2.17}$$

From this equation it is possible to determine the diffusion coefficient by solving (2.17) for D, yielding

$$D = \frac{\pi N^2}{4C_0^2A^2T}. \tag{2.18}$$

A second useful piece of information is found from (2.16) by observing that $C(x,t)/C_0$ is constant on any curve for which z is constant. Thus, the curve $t = x^2/D$ is a level curve for the concentration, and gives a measure of how fast the substance is moving into the capillary. The time $t = x^2/D$ is called the *diffusion time* for the process. To give some idea of the effectiveness of diffusion in various cellular contexts, in Table 2.3 is shown typical diffusion times for a variety of cellular structures. Clearly, diffusion is effective for transport when distances are short, but totally inadequate for longer distances, such as along a nerve axon. Obviously, biological systems must employ other transport mechanisms in these situations in order to survive.

2.2.5 Buffered Diffusion

It is often the case that reactants in an enzymatic reaction (as in Chapter 1) are free to diffuse, so that one must keep track of the effects of both diffusion and reaction. Such problems, called *reaction–diffusion systems*, are of fundamental significance in physiology and are also important and difficult mathematically.

Table 2.3 Estimates of diffusion times for cellular structures of typical dimensions, computed from the relation $t = x^2/D$ using $D = 10^{-5}\text{cm}^2/\text{s}$ (typical for molecules the size of oxygen or carbon dioxide).

x	t	Example
10 nm	100 ns	Thickness of cell membrane
1 μm	1 ms	Size of mitochondrion
10 μm	100 ms	Radius of small mammalian cell
100 μm	10 s	Diameter of a large muscle fiber
250 μm	60 s	Radius of squid giant axon
1 mm	16.7 min	Half-thickness of frog sartorius muscle
2 mm	1.1 h	Half-thickness of lens in the eye
5 mm	6.9 h	Radius of mature ovarian follicle
2 cm	2.6 d	Thickness of ventricular myocardium
1 m	31.7 yrs	Length of a (long!) nerve axon

An important situation, in which reaction and diffusion interact to modify the behavior, occurs when a diffusing species is buffered by a larger diffusing molecule. This occurs, for example, with oxygen in muscle (which we discuss below), or Ca^{2+}, or H^+. The earliest studies of the buffered diffusion equation were those of Irving et al. (1990) and Wagner and Keizer (1994), while Neher and his colleagues (Zhou and Neher, 1993; Naraghi and Neher, 1997; Naraghi et al. 1998) have done a great deal of work on Ca^{2+} buffering. More theoretical analyses have been performed by Sneyd et al. (1998), Smith et al. (2001), and Tsai and Sneyd (2005, 2007a,b).

Consider a "one-dimensional" cell is which there are hydrogen ions (for example) and buffer, B. We assume that the buffering reaction follows

$$\text{H}^+ + \text{B} \underset{k_-}{\overset{k_+}{\rightleftarrows}} \text{HB}. \tag{2.19}$$

Conservation implies

$$\frac{\partial u}{\partial t} = D_h \frac{\partial^2 u}{\partial x^2} + k_- w - k_+ uv + f(t,x,u), \tag{2.20}$$

$$\frac{\partial v}{\partial t} = D_b \frac{\partial^2 v}{\partial x^2} + k_- w - k_+ uv, \tag{2.21}$$

$$\frac{\partial w}{\partial t} = D_b \frac{\partial^2 w}{\partial x^2} - k_- w + k_+ uv, \tag{2.22}$$

where $u = [\text{H}^+]$, $v = [\text{B}]$, and $w = [\text{HB}]$. Since the buffer is a large molecule, we assume that the diffusion of B is the same as that of HB. We impose no-flux boundary conditions at the ends of the cell and assume that v and w are initially uniform (for example, if w is initially zero, and the buffer is uniformly distributed). The reaction term $f(t,x,u)$ denotes all the other reactions of u apart from the buffering.

Adding (2.21) and (2.22) we obtain

$$\frac{\partial(v+w)}{\partial t} = D_b \frac{\partial^2 (v+w)}{\partial x^2}. \tag{2.23}$$

Since $v+w$ is initially uniform, it remains uniform for all time, so that $v+w = w_0$, where w_0 is the total amount of buffer.

If the buffering reaction is fast compared to the other reactions (i.e., those described by $f(t,x,u)$), then we can assume u and v to be in quasi-equilibrium, so that

$$k_-(w_0 - v) - k_+ uv = 0, \tag{2.24}$$

which implies that

$$v = \frac{K_{\text{eq}} w_0}{K_{\text{eq}} + u}, \qquad \text{where } K_{\text{eq}} = \frac{k_-}{k_+}. \tag{2.25}$$

Subtracting (2.21) from (2.20) yields

$$\frac{\partial(u-v)}{\partial t} = D_h \frac{\partial^2 u}{\partial x^2} - D_b \frac{\partial^2 v}{\partial x^2} + f(t,x,u). \tag{2.26}$$

However, since we know v as a function of u, we can eliminate v to find a nonlinear reaction–diffusion equation for u alone,

$$\frac{\partial}{\partial t}\left(u - \frac{K_{\text{eq}} w_0}{K_{\text{eq}} + u}\right) = D_h \frac{\partial^2 u}{\partial x^2} - D_b \frac{\partial^2}{\partial x^2}\left(\frac{K_{\text{eq}} w_0}{K_{\text{eq}} + u}\right) + f(t,x,u). \tag{2.27}$$

We expand some of the derivatives and find

$$\left(1 + \frac{K_{\text{eq}} w_0}{(K_{\text{eq}} + u)^2}\right) u_t = D_h \frac{\partial^2 u}{\partial x^2} + D_b \frac{\partial}{\partial x}\left(\frac{K_{\text{eq}} w_0}{(K_{\text{eq}} + u)^2} u_x\right) + f(t,x,u). \tag{2.28}$$

Letting

$$\theta(u) = \frac{K_{\text{eq}} w_0}{(K_{\text{eq}} + u)^2} \tag{2.29}$$

then gives

$$u_t = \frac{D_h + \theta(u) D_b}{1 + \theta(u)} u_{xx} + \frac{D_b \theta'(u)}{1 + \theta(u)} (u_x)^2 + \frac{f(t,x,u)}{1 + \theta(u)}. \tag{2.30}$$

Thus, buffering gives rise to a nonlinear transport equation with a diffusion coefficient that is a nonlinear function of u.

In some cases it is reasonable to assume that $u \ll K_{\text{eq}}$. In this limit we find that u has an effective diffusion coefficient

$$D_{\text{eff}} = \frac{D_h + D_b \frac{w_0}{K_{\text{eq}}}}{1 + \frac{w_0}{K_{\text{eq}}}}, \tag{2.31}$$

a convex linear combination of the two diffusion coefficients, D_h and D_b. In addition, the reaction terms are scaled by the constant factor $\frac{1}{1+w_0/K_{\text{eq}}}$. If, additionally, $D_b = 0$, we

recover the usual diffusion equation, but one for which both the diffusion coefficient and the reaction terms are scaled by the same constant factor.

2.3 Facilitated Diffusion

A second important example in which both diffusion and reaction play a role is known as *facilitated diffusion*. Facilitated diffusion occurs when the flux of a chemical is amplified by a reaction that takes place in the diffusing medium. An example of facilitated diffusion occurs with the flux of oxygen in muscle fibers. In muscle fibers, oxygen is bound to myoglobin and is transported as oxymyoglobin, and this transport is greatly enhanced above the flow of oxygen in the absence of myoglobin (Wyman, 1966; Murray, 1971; Murray and Wyman, 1971; Rubinow and Dembo, 1977).

This well-documented observation needs further explanation, because at first glance it seems counterintuitive. Myoglobin molecules are much larger (molecular weight M = 16,890) than oxygen molecules (molecular weight M = 32) and therefore have a much smaller diffusion coefficient ($D = 4.4 \times 10^{-7}$ and $D = 1.2 \times 10^{-5}\text{cm}^2/\text{s}$ for myoglobin and oxygen, respectively). The diffusion of oxymyoglobin would therefore seem to be much slower than the diffusion of free oxygen. Further, from the calculation in the last section, the diffusion of free oxygen is much slower when it is buffered by myoglobin since the effective diffusion coefficient of oxygen is lowered substantially by diffusion.

To anticipate slightly, the answer is that, at steady state, the total transport of oxygen is the sum of the free oxygen transport and additional oxygen that is transported by the diffusing buffer. If there is a lot of buffer, with a lot of oxygen bound, this additional transport due to the buffer can be substantial.

A simple model of this phenomenon is as follows. Suppose we have a slab reactor containing diffusing myoglobin. On the left (at $x = 0$) the oxygen concentration is held fixed at s_0, and on the right (at $x = L$) it is held fixed at s_L, which is assumed to be less than s_0.

If f is the rate of uptake of oxygen into oxymyoglobin, then equations governing the concentrations of $s = [\text{O}_2], e = [\text{Mb}], c = [\text{MbO}_2]$ are

$$\frac{\partial s}{\partial t} = D_s \frac{\partial^2 s}{\partial x^2} - f, \tag{2.32}$$

$$\frac{\partial e}{\partial t} = D_e \frac{\partial^2 e}{\partial x^2} - f, \tag{2.33}$$

$$\frac{\partial c}{\partial t} = D_c \frac{\partial^2 c}{\partial x^2} + f. \tag{2.34}$$

It is reasonable to take $D_e = D_c$, since myoglobin and oxymyoglobin are nearly identical in molecular weight and structure. Since myoglobin and oxymyoglobin remain inside the slab, it is also reasonable to specify the boundary conditions $\partial e/\partial x = \partial c/\partial x = 0$ at $x = 0$ and $x = L$. Because it reproduces the oxygen saturation curve (discussed in

Chapter 13), we assume that the reaction of oxygen with myoglobin is governed by the elementary reaction

$$O_2 + Mb \underset{k_-}{\overset{k_+}{\rightleftarrows}} MbO_2,$$

so that (from the law of mass action) $f = -k_- c + k_+ se$. The total amount of myoglobin is conserved by the reaction, so that at steady state $e + c = e_0$ and (2.33) is superfluous.

At steady state,

$$0 = s_t + c_t = D_s s_{xx} + D_c c_{xx}, \tag{2.35}$$

and thus there is a second conserved quantity, namely

$$D_s \frac{ds}{dx} + D_c \frac{dc}{dx} = -J, \tag{2.36}$$

which follows by integrating (2.35) once with respect to x. The constant J (which is yet to be determined) is the sum of the flux of free oxygen and the flux of oxygen in the complex oxymyoglobin, and therefore represents the total flux of oxygen. Integrating (2.36) with respect to x between $x = 0$ and $x = L$, we can express the total flux J in terms of boundary values of the two concentrations as

$$J = \frac{D_s}{L}(s_0 - s_L) + \frac{D_c}{L}(c_0 - c_L), \tag{2.37}$$

although the values c_0 and c_L are as yet unknown.

To further understand this system of equations, we introduce dimensionless variables, $\sigma = \frac{k_+}{k_-} s$, $u = c/e_0$, and $x = Ly$, in terms of which (2.32) and (2.34) become

$$\epsilon_1 \sigma_{yy} = \sigma(1 - u) - u = -\epsilon_2 u_{yy}, \tag{2.38}$$

where $\epsilon_1 = \frac{D_s}{e_0 k_+ L^2}, \epsilon_2 = \frac{D_c}{k_- L^2}$.

Reasonable numbers for the uptake of oxygen by myoglobin (Wittenberg, 1966) are $k_+ = 1.4 \times 10^{10} \text{cm}^3\,\text{M}^{-1}\text{s}^{-1}$, $k_- = 11\ \text{s}^{-1}$, and $L = 0.022$ cm in a solution with $e_0 = 1.2 \times 10^{-5}\ \text{M/cm}^3$. (These numbers are for an experimental setup in which the concentration of myoglobin was substantially higher than what naturally occurs in living tissue.) With these numbers we estimate that $\epsilon_1 = 1.5 \times 10^{-7}$, and $\epsilon_2 = 8.2 \times 10^{-5}$. Clearly, both of these numbers are small, suggesting that oxygen and myoglobin are at quasi-steady state throughout the medium, with

$$c = e_0 \frac{s}{K + s}, \tag{2.39}$$

where $K = k_-/k_+$. Now we substitute (2.39) into (2.37) to find the flux

$$\begin{aligned} J &= \frac{D_s}{L}(s_0 - s_L) + \frac{D_c}{L}e_0\left(\frac{s_0}{K+s_0} - \frac{s_L}{K+s_L}\right) \\ &= \frac{D_s}{L}(s_0 - s_L)\left(1 + \frac{D_c}{D_s}\frac{e_0 K}{(s_0+K)(s_L+K)}\right) \\ &= \frac{D_s}{L}(1+\mu\rho)(s_0 - s_L), \end{aligned} \tag{2.40}$$

where $\rho = \frac{D_c}{D_s}\frac{e_0}{K}$, $\mu = \frac{K^2}{(s_0+K)(s_L+K)}$.

In terms of dimensionless variables the full solution is given by

$$\sigma(y) + \rho u(y) = y[\sigma(1) + \rho u(1)] + (1-y)[\sigma(0) + \rho u(0)], \tag{2.41}$$

$$u(y) = \frac{\sigma(y)}{1+\sigma(y)}. \tag{2.42}$$

Now we see how diffusion can be facilitated by an enzymatic reaction. In the absence of a diffusing carrier, $\rho = 0$ and the flux is purely Fickian, as in (2.14). However, in the presence of carrier, diffusion is enhanced by the factor $\mu\rho$. The maximum enhancement possible is at zero concentration, when $\mu = 1$. With the above numbers for myoglobin, this maximum enhancement is substantial, being $\rho = 560$. If the oxygen supply is sufficiently high on the left side (near $x = 0$), then oxygen is stored as oxymyoglobin. Moving to the right, as the total oxygen content drops, oxygen is released by the myoglobin. Thus, even though the bound oxygen diffuses slowly compared to free oxygen, the quantity of bound oxygen is high (provided that e_0 is large compared to the half saturation level K), so that lots of oxygen is transported . We can also understand that to take advantage of the myoglobin-bound oxygen, the concentration of oxygen must drop to sufficiently low levels so that myoglobin releases its stored oxygen.

To explain it another way, note from (2.40) that J is the sum of two terms, the usual ohmic flux term and an additional term that depends on the diffusion coefficient of MbO_2. The total oxygen flux is the sum of the flux of free oxygen and the flux of oxygen bound to myoglobin. Clearly, if oxymyoglobin is free to diffuse, the total flux is thereby increased. But since oxymyoglobin can only diffuse down its gradient, the concentration of oxymyoglobin must be higher on one side than the other.

In Fig. 2.3A are shown the dimensionless free oxygen concentration σ and the dimensionless bound oxygen concentration u plotted as functions of position. Notice that the free oxygen content falls at first, indicating higher free oxygen flux, and the bound oxygen decreases more rapidly at larger y. Perhaps easier to interpret is Fig. 2.3B, where the dimensionless flux of free oxygen and the dimensionless flux of bound oxygen are shown as functions of position. Here we can see that as the free oxygen concentration drops, the flux of free oxygen also drops, but the flux of bound oxygen increases. For large y, most of the flux is due to the bound oxygen. For these figures, $\rho = 10$, $\sigma(0) = 2.0$, $\sigma(1) = 0.1$.

One mathematical detail that was ignored in this discussion is the validity of the quasi-steady-state solution (2.39) as an approximation of (2.38). Usually, when one

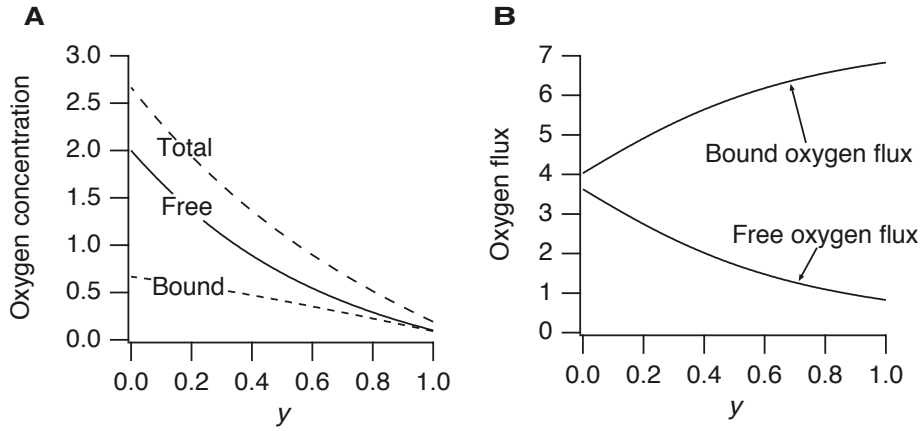


Figure 2.3 A: Free oxygen content $\sigma(y)$ and bound oxygen content $u(y)$ as functions of y. B: Free oxygen flux $-\sigma'(y)$ and bound oxygen flux $-\rho u'(y)$ as functions of y.

makes an approximation to boundary value problems in which the order of the system is reduced (as here where the order is four, and drops by two when ϵ_1 and ϵ_2 are ignored), there are difficulties with the solution at the boundary, because the boundary conditions cannot, in general, be met. Such problems, discussed briefly in Chapter 1 in the context of enzyme kinetics, are called *singular perturbation problems*, because the behavior of the solutions as functions of the small parameters is not regular, but singular (certain derivatives become infinitely large as the parameters approach zero). In this problem, however, there are no boundary layers, and the quasi-steady-state solution is a uniformly valid approximation to the solution. This occurs because the boundary conditions on c are of no-flux (Neumann) type, rather than of fixed (Dirichlet) type. That is, since the value of c is not specified by the boundary conditions, c is readily adjusted so that there are no boundary layers. Only a slight correction to the quasi-steady-state solution is needed to meet the no-flux boundary conditions, but this correction affects only the derivative, not the value, of c in a small region near the boundaries.

2.3.1 Facilitated Diffusion in Muscle Respiration

Even at rest, muscle fibers consume oxygen. This is because ATP is constantly consumed to maintain a nonzero membrane potential across a muscle cell wall, and this consumption of energy requires constant metabolizing of sugar, which consumes oxygen. Although sugar can be metabolized anaerobically, the waste product of this reaction is lactic acid, which is toxic to the cell. In humans, the oxygen consumption of live muscle tissue at rest is about 5×10^{-8} mol/cm^{3}s, and the concentration of myoglobin is about 2.8×10^{-7} mol/cm^3. Thus, when myoglobin is fully saturated, it contains only about a 5 s supply of oxygen. Further, the oxygen at the exterior of the muscle cell must

penetrate to the center of the cell to prevent the oxygen concentration at the center falling to zero, a condition called *oxygen debt*.

To explain how myoglobin aids in providing oxygen to a muscle cell and helps to prevent oxygen debt, we examine a model of oxygen consumption that includes the effects of diffusion of oxygen and myoglobin. We suppose that a muscle fiber is a long circular cylinder (radius $a = 2.5 \times 10^{-3}$ cm) and that diffusion takes place only in the radial direction. We suppose that the oxygen concentration at the boundary of the fiber is a fixed constant and that the distribution of chemical species is radially symmetric. With these assumptions, the steady-state equations governing the diffusion of oxygen and oxymyoglobin are

$$D_s \frac{1}{r}\frac{d}{dr}\left(r\frac{ds}{dr}\right) - f - g = 0, \tag{2.43}$$

$$D_c \frac{1}{r}\frac{d}{dr}\left(r\frac{dc}{dr}\right) + f = 0, \tag{2.44}$$

where, as before, $s = [O_2]$, $c = [MbO_2]$, and $f = -k_- c + k_+ se$. The coordinate r is in the radial direction. The new term in these equations is the constant g, corresponding to the constant consumption of oxygen. The boundary conditions are $s = s_a, dc/dr = 0$ at $r = a$, and $ds/dr = dc/dr = 0$ at $r = 0$. For muscle, s_a is typically 3.5×10^{-8} mol/cm^3 (corresponding to the partial pressure 20 mm Hg). Numerical values for the parameters in this model are difficult to obtain, but reasonable numbers are $D_s = 10^{-5}$ cm^2/s, $D_c = 5 \times 10^{-7}$ cm^2/s, $k_+ = 2.4 \times 10^{10}$ cm^3/mol $\cdot$ s, and $k_- = 65$/s (Wyman, 1966).

Introducing nondimensional variables $\sigma = \frac{k_+}{k_-}s, u = c/e_0$, and $r = ay$, we obtain the differential equations

$$\epsilon_1 \frac{1}{y}\frac{d}{dy}\left(y\frac{d\sigma}{dy}\right) - \gamma = \sigma(1-u) - u = -\epsilon_2 \frac{1}{y}\frac{d}{dy}\left(y\frac{du}{dy}\right), \tag{2.45}$$

where $\epsilon_1 = \frac{D_s}{e_0 k_+ a^2}, \epsilon_2 = \frac{D_c}{k_- a^2}, \gamma = g/k_-$. Using the parameters appropriate for muscle, we estimate that $\epsilon_1 = 2.3 \times 10^{-4}, \epsilon_2 = 1.2 \times 10^{-3}, \gamma = 3.3 \times 10^{-3}$. While these numbers are not as small as for the experimental slab described earlier, they are small enough to warrant the assumption that the quasi-steady-state approximation (2.39) holds in the interior of the muscle fiber.

It also follows from (2.45) that

$$\epsilon_1 \frac{1}{y}\frac{d}{dy}\left(y\frac{d\sigma}{dy}\right) + \epsilon_2 \frac{1}{y}\frac{d}{dy}\left(y\frac{du}{dy}\right) = \gamma. \tag{2.46}$$

We integrate (2.46) twice with respect to y to find

$$\epsilon_1 \sigma + \epsilon_2 u = A \ln y + B + \frac{\gamma}{4} y^2. \tag{2.47}$$

The constants A and B are determined by boundary conditions. Since we want the solution to be bounded at the origin, $A = 0$, and B is related to the concentration at the origin.

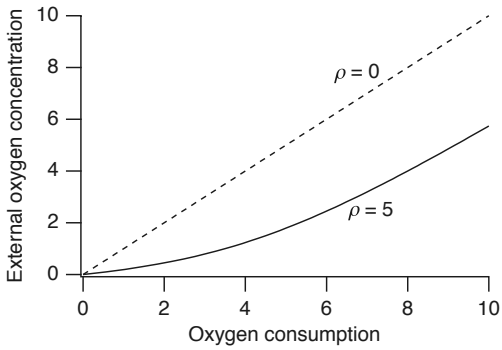


Figure 2.4 Critical concentration σ_0 plotted as a function of oxygen consumption $\frac{\gamma}{4\epsilon_1}$. The dashed curve is the critical concentration with no facilitated diffusion.

Now suppose that there is just enough oxygen at the boundary to prevent oxygen debt. In this model, oxygen debt occurs if σ falls to zero. Marginal oxygen debt occurs if $\sigma = u = 0$ at $y = 0$. For this boundary condition, we take $A = B = 0$. Then the concentration at the boundary must be at least as large as σ_0, where, using the quasi-steady state $\sigma(1-u) = u$,

$$\sigma_0 + \rho \frac{\sigma_0}{\sigma_0 + 1} = \frac{\gamma}{4\epsilon_1}, \tag{2.48}$$

and where $\rho = \epsilon_2/\epsilon_1$. Otherwise, the center of the muscle is in oxygen debt. Note also that σ_0 is a decreasing function of ρ, indicating a reduced need for external oxygen because of facilitated diffusion.

A plot of this critical concentration σ_0 as a function of the scaled consumption $\frac{\gamma}{4\epsilon_1}$ is shown in Fig. 2.4. For this plot $\rho = 5$, which is a reasonable estimate for muscle. The dashed curve is the critical concentration when there is no facilitated diffusion ($\rho = 0$). The easy lesson from this plot is that facilitated diffusion decreases the likelihood of oxygen debt, since the external oxygen concentration necessary to prevent oxygen debt is smaller in the presence of myoglobin than without.

A similar lesson comes from Fig. 2.5, where the internal free oxygen content σ is shown, plotted as a function of radius y. The solid curves show the internal free oxygen with facilitated diffusion, and the dashed curve is without. The smaller of the two solid curves and the dashed curve have exactly the critical external oxygen concentration, showing clearly that in the presence of myoglobin, oxygen debt is less likely at a given external oxygen concentration. The larger of the two solid curves has the same external oxygen concentration as the dashed curve, showing again the contribution of facilitation toward preventing oxygen debt. For this figure, $\rho = 5$, $\gamma/\epsilon_1 = 14$.

2.4 Carrier-Mediated Transport

Some substances are insoluble in the cell membrane and yet pass through by a process called *carrier-mediated transport*. It is also called *facilitated diffusion* in many physiology books, although we prefer to reserve this expression for the process described in the

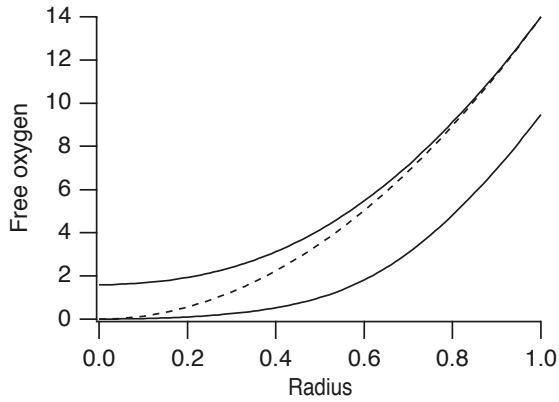


Figure 2.5 Free oxygen σ as a function of radius y. Solid curves show oxygen concentration in the presence of myoglobin ($\rho = 5$), the lower of the two having the critical external oxygen concentration. The dashed curve shows the oxygen concentration without facilitation at the critical external concentration level.

previous section. Carrier-mediated transport is the means by which some sugars cross the cell membrane to provide an energy source for the cell. For example, glucose, the most important of the sugars, combines with a carrier protein at the outer boundary of the membrane, and by means of a conformational change is released from the inner boundary of the membrane.

There are three types of carrier-mediated transport. Carrier proteins that transport a single solute from one side of the membrane to the other are called *uniports*. Other proteins function as coupled transporters by which the simultaneous transport of two solute molecules is accomplished, either in the same direction (called a *symport*) or in the opposite direction (called an *antiport*).

2.4.1 Glucose Transport

Although the details are not certain, the transport of glucose across the lipid bilayer of the cell membrane is thought to occur when the carrier molecule alternately exposes the solute binding site first on one side and then on the other side of the membrane. It is considered highly unlikely that the carrier molecule actually diffuses back and forth through the membrane.

We can model the process of glucose transport as follows: We suppose that the population of enzymatic carrier proteins C has two conformational states, C_i and C_e, with its glucose binding site exposed on the cell interior (subscript i) or exterior (subscript e) of the membrane, respectively. The glucose substrate on the interior S_i can bind with C_i and the glucose substrate on the exterior can bind with enzyme C_e to form the complex P_i or P_e, respectively. Finally, a conformational change transforms P_i into P_e and vice versa. These statements are summarized in Fig. 2.6.

The equations describing this model are

$$\frac{dp_i}{dt} = kp_e - kp_i + k_+ s_i c_i - k_- p_i, \tag{2.49}$$

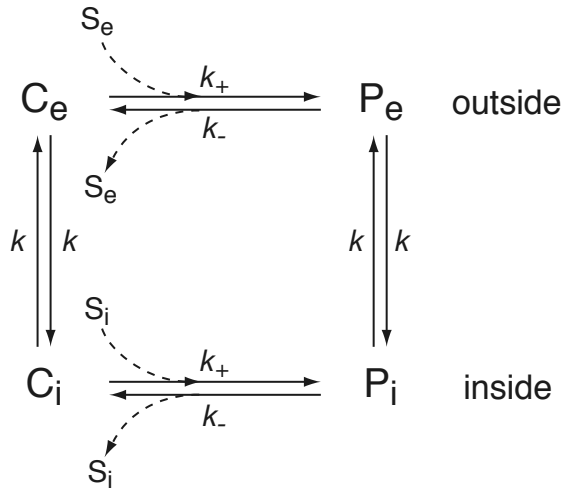


Figure 2.6 Schematic diagram of the glucose transporter described by (2.49)–(2.52).

$$\frac{dp_e}{dt} = kp_i - kp_e + k_+ s_e c_e - k_- p_e, \tag{2.50}$$

$$\frac{dc_i}{dt} = kc_e - kc_i + k_- p_i - k_+ s_i c_i, \tag{2.51}$$

$$\frac{dc_e}{dt} = kc_i - kc_e + k_- p_e - k_+ s_e c_e. \tag{2.52}$$

where $s_i = [\mathrm{S}_i]$, $p_i = [\mathrm{P}_i]$, etc. Since the total amount of receptor is conserved, we have $p_i + p_e + c_i + c_e = C_0$, where C_0 is a constant (the total transporter concentration). Hence there are only three independent equations, not four. The flux, J, is

$$J = k_- p_i - k_+ s_i c_i = k_+ s_e c_e - k_- p_e, \tag{2.53}$$

where we have defined a flux from outside to inside to be positive.

We find the steady-state flux by setting all derivatives to zero and solving the resulting algebraic system. It follows that

$$J = \frac{1}{2} K k C_0 \frac{s_e - s_i}{(s_i + K + K_d)(s_e + K + K_d) - K_d^2}, \tag{2.54}$$

where $K = k_-/k_+$ and $K_d = k/k_+$. Since k is the rate at which conformational change takes place, it acts like a diffusion coefficient in that it reflects the effect of random thermal activity at the molecular level.

The nondimensional flux is

$$j = \frac{\sigma_e - \sigma_i}{(\sigma_i + 1 + \kappa)(\sigma_e + 1 + \kappa) - \kappa^2}, \tag{2.55}$$

where $\sigma_i = s_i/K, \sigma_e = s_e/K, \kappa = K_d/K$. A plot of this nondimensional flux is shown in Fig. 2.7, plotted as a function of extracellular glucose σ_e, with fixed intracellular glucose and fixed κ. We can see that the rate of transport is limited by saturation of the enzyme kinetics (this saturation is observed experimentally) and thermal conformational

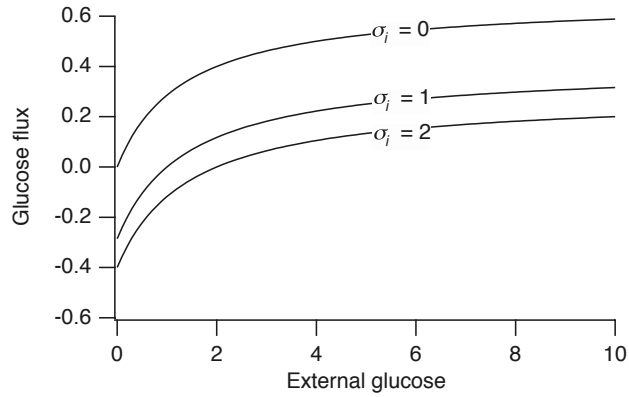


Figure 2.7 Plot of the (nondimensional) flux of glucose as a function of extracellular glucose, for three fixed intracellular glucose concentrations (σ_i), with $\kappa = K_d/K = 0.5$.

change is crucial to the transport process, as transport J drops to zero if $K_d = 0$. The binding affinity of the carrier protein for glucose (k_+), and hence the flux of glucose, is controlled by insulin.

It is important to recognize that the above expression for J is for the steady-state flux only. If the system is not at steady state, then the flux from the outside to the transporter, $J_{\text{on}} = k_+ s_e c_e - k_- p_e$, need not be the same as the flux off the transporter to the inside, $J_{\text{off}} = k_- p_i - k_+ s_i c_i$. Obviously, in this case the differential equations must be solved to obtain J_{on} and J_{off}.

It should be noted that there are two ways that the model of Fig. 2.6 can be understood. First, as we did here, we can let each variable represent the concentration of transporters in each of the four possible states. In this case, the conservation relationship is $s_i + p_i + s_e + p_e = C_0$. If each of the variables is scaled by C_0, the conservation relationship becomes $s_i + p_i + s_e + p_e = 1$, and each variable is then the fraction of the population in each state.

However, there is another way to interpret this second conservation relationship. If $s_i + p_i + s_e + p_e = 1$ we can interpret the model as referring to the behavior of a single exchanger, in which case the variables are probabilities of being in a given state, and the exchanger is modeled as a Markov process (see the Appendix to this chapter).

Markov models such as that shown in Fig. 2.6 can often be simplified by assuming that some of the transitions are much faster than others. The technique of reduction using a fast time scale is used in many places throughout this book; indeed, it is used in Chapter 1, in the equilibrium and quasi-steady-state approximations of enzyme kinetics; even though the technique is described in Chapter 1, it is sufficiently important that it warrants repeating.

The procedure can be simply illustrated with this model of the glucose transporter. Suppose that the binding and release of glucose is much faster than the change in conformation, i.e., that the transitions between C_e and P_e, and between C_i and P_i, are

much faster than those between C_i and C_e, or between P_e and P_i, so that $K_d \ll 1$. Assuming fast equilibrium between C_e and P_e, and between C_i and P_i, gives

$$s_e c_e = K p_e, \qquad s_i c_i = K p_i. \tag{2.56}$$

Now, we introduce two variables, $x = c_e + p_e$, $y = c_i + p_i = 1 - x$ (taking $C_0 = 1$). The differential equation for x is found by adding (2.50) and (2.52) to be

$$\begin{aligned} \frac{dx}{dt} &= k c_i + k p_i - k c_e - k p_e \\ &= k y - k x \\ &= k(1 - 2x), \end{aligned} \tag{2.57}$$

from which it follows that the steady value of x is $x_0 = 1/2$.

Next, from (2.56) we find that

$$x = c_e(1 + s_e/K) = p_e(1 + K/s_e), \tag{2.58}$$

with similar equations for y. Hence, at steady state, the flux through the transporter is given by

$$\begin{aligned} J &= k_+ s_e c_e - k_- p_e = k p_i - k p_e \\ &= \frac{k s_i x_0}{s_i + K} - \frac{k s_e x_0}{s_e + K} \\ &= \frac{k s_i \frac{1}{2}}{s_i + K} - \frac{k s_e \frac{1}{2}}{s_e + K} \\ &= kK \frac{1}{2} \frac{s_i - s_e}{(s_i + K)(s_e + K)}, \end{aligned} \tag{2.59}$$

where we have used (2.58) to replace p_e, and the analogous equation to replace p_i.

Notice that this answer is the same as found by letting $K_d \to 0$ in (2.54). However, while the two approaches give the same answer, the quasi-steady-state reduction of the full model is often preferable, especially when the solution of the full model is difficult to obtain.

Other examples of how to simplify Markov models with a fast time scale reduction are given in Exercises 12 and 13.

2.4.2 Symports and Antiports

Models of symport and antiport transporters follow in similar fashion. For a symport the protein carrier has multiple binding sites, which can be exposed to the intracellular or extracellular space. A change of conformation exchanges the location of all of the participating binding sites, from inside to outside, or vice versa. An example of a symport is the Na^+-driven glucose symport that transports glucose and Na^+ from the lumen of the gut to the intestinal epithelium. A similar process occurs in epithelial cells lining the proximal tubules in the kidney, to remove glucose and amino acids from the

filtrate (discussed in Chapter 17). Five different amino acid cotransporters have been identified.

If there are k binding sites that participate in the exchange, then there are 2^k possible combinations of bound and unbound sites. The key assumption that makes this model of transport work is that only the completely unbound or completely bound carrier participates in a conformational change. Thus, there is a carrier molecule, say C, with two conformations, C_i and C_e, and a fully bound complex P, also with two conformations, P_i and P_e, and possible transformation between the two conformations,

$$C_i \underset{k_{-c}}{\overset{k_c}{\rightleftarrows}} C_e, \qquad P_i \underset{k_{-p}}{\overset{k_p}{\rightleftarrows}} P_e. \tag{2.60}$$

In addition, there are 2^k possible combinations of binding and unbinding in each of the two conformations. For example, with two substrates S and T, and one binding site for each, we have the complexes C, SC, CT, and SCT = P. The possible reactions are summarized in Fig. 2.8.

Unfortunately, the analysis of this fully general reaction scheme is quite complicated. However, it simplifies significantly if we assume that the intermediates can be safely ignored and postulate the multi-molecular reaction scheme

$$m\mathrm{S} + n\mathrm{T} + \mathrm{C} \underset{k_-}{\overset{k_+}{\rightleftarrows}} \mathrm{P}. \tag{2.61}$$

Now the result for a symport is strikingly similar to the uniport flux, with

$$J = \frac{1}{2} K_d K k_+ C_0 \frac{s_e^m t_e^n - s_i^m t_i^n}{(s_i^m t_i^n + K + K_d)(s_e^m t_e^n + K + K_d) - K_d^2}, \tag{2.62}$$

where the flux of s is mJ and the flux of t is nJ. Here we have set $k_c = k_{-c} = k_p = k_{-p} = k$ and then $K = k_-/k_+$ and $K_d = k/k_+$.

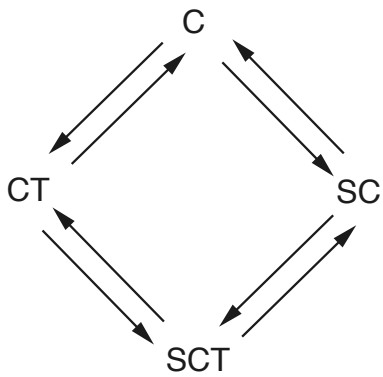


Figure 2.8 States and possible transitions of a transporter with two substrates, S and T, and one binding site for each.

For an antiport, the subscripts on one of the substances must be exchanged, to give

$$J = \frac{1}{2} K_d K k_+ C_0 \frac{s_e^m t_i^n - s_i^m t_e^n}{(s_i^m t_e^n + K + K_d)(s_e^m t_i^n + K + K_d) - K_d^2}. \tag{2.63}$$

The effectiveness of this type of exchanger is determined by the coefficients m and n. For this antiport, flux is positive (S flows inward and T flows outward) if

$$\left(\frac{s_e}{s_i}\right)^m > \left(\frac{t_e}{t_i}\right)^n. \tag{2.64}$$

For example, for the Na^+–Ca^{2+} exchanger (discussed in more detail in the next section) which exchanges three Na^+ ions for one Ca^{2+} ion, a ratio of extracellular to intracellular Na^+ of about 8 can be used to effectively pump Ca^{2+} out of a cell even when the ratio of extracellular to intracellular Ca^{2+} is 500.

2.4.3 Sodium–Calcium Exchange

For the glucose transporter described above, membrane flux is driven by a concentration difference of glucose across the membrane, and if glucose concentrations equilibrate, the transmembrane flux becomes zero. However, because it relies on two concentration differences, an antiport transporter such as the Na^+–Ca^{2+} exchanger can act as a pump. Although this transporter is a passive pump (because it consumes no chemical energy directly), it is often described as a secondarily active pump; it uses the Na^+ gradient to pump Ca^{2+} out of the cell against its concentration gradient, but energy is required to establish and maintain the Na^+ gradient. Na^+–Ca^{2+} exchange is an important mechanism for Ca^{2+} removal in a number of cell types, particularly cardiac ventricular cells, in which much of the Ca^{2+} that enters the cell during an action potential is removed from the cell by the Na^+–Ca^{2+} exchanger (Chapter 12). It has therefore been studied extensively, and a number of highly detailed models have been constructed (Hilgemann, 2004; Kang and Hilgemann, 2004). Here we describe a simple model of this important transporter.

In our model (see Fig. 2.9), E_i is the exchanger protein in the conformation for which the binding sites are exposed to the interior of the cell, and E_e is the conformation for which the binding sites are exposed to the exterior. Starting at state X_1 in the top left of the figure, the exchanger can bind Ca^{2+} inside the cell, simultaneously releasing three Na^+ ions to the interior. A change of conformation to E_e then allows the exchanger to release the Ca^{2+} to the outside and bind three external Na^+. A return to the E_i conformation completes the cycle. Of course, it is a crude approximation to assume that one Ca^{2+} and three Na^+ ions bind or unbind the exchanger simultaneously.

It is now straightforward to calculate the steady flux for this model. As with the previous transporter models, we first solve for the steady-state values of x_1, x_2, y_1, and y_2, the fraction of exchangers in the state X_1, X_2, Y_1, and Y_2, respectively. There are four equations: three differential equations for exchanger states and one conservation

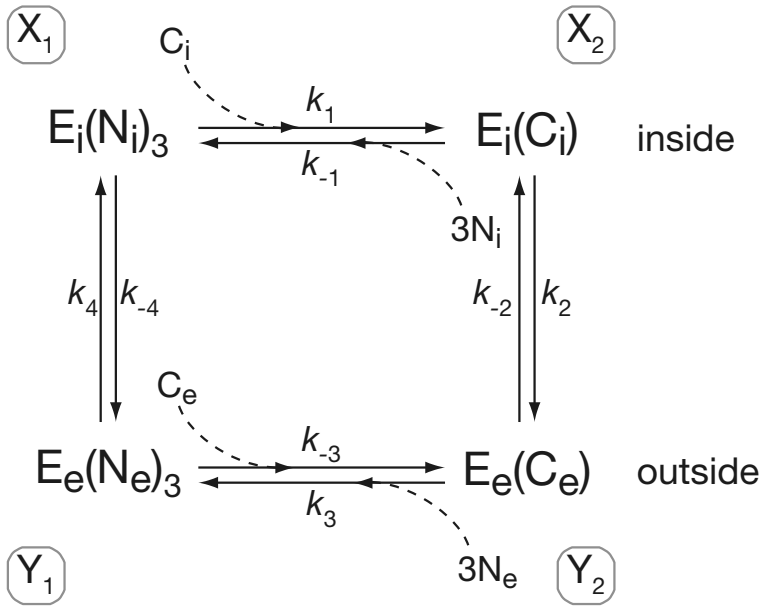


Figure 2.9 Schematic diagram of a simple model of the Na^+–Ca^{2+} exchanger.

equation. These are

$$\frac{dx_1}{dt} = k_{-1}n_i^3 x_2 + k_4 y_1 - (k_1 c_i + k_{-4})x_1, \tag{2.65}$$

$$\frac{dx_2}{dt} = k_{-2}y_2 + k_1 c_i x_1 - (k_2 + k_{-1}n_i^3)x_2, \tag{2.66}$$

$$\frac{dy_1}{dt} = k_{-4}x_1 + k_3 n_e^3 y_2 - (k_4 + k_{-3}c_e)y_1, \tag{2.67}$$

$$1 = x_1 + x_2 + y_1 + y_2. \tag{2.68}$$

Here c and n denote, respectively, Ca^{2+} and Na^+ concentration, and the subscripts e and i represent external and internal concentrations.

Using a symbolic package such as Maple, the steady-state solution of these equations is easily calculated. The flux, J, is found to be

$$J = k_4 y_1 - k_{-4}x_1 = \frac{k_1 k_2 k_3 k_4 (c_i n_e^3 - K_1 K_2 K_3 K_4 c_e n_i^3)}{16 \text{ positive terms}}, \tag{2.69}$$

where, as usual, $K_i = k_{-i}/k_i$.

Notice that the units of the flux J here (1/time) are different from those in the previous examples (concentration/time), because here the variables x_i and y_i are fractions of exchangers in a particular state (or probabilities) rather than concentrations of exchangers in a particular state. Hence, the flux in this model is a turnover rate, i.e., the

number of times the exchanger goes around the cycle per unit time. This can easily be converted to a concentration per time if the concentration of the exchangers is known.

An Electrogenic Exchanger

An important difference between the Na^+–Ca^{2+} exchange process and the transport processes discussed previously is that Na^+ and Ca^{2+} are ions. Since each cycle of the Na^+–Ca^{2+} exchanger transports two positive charges out and three positive charges in, it generates an electric current. Such exchangers are said to be *electrogenic*.

As is discussed in Section 2.6, all cells have an electrical potential difference across their membranes. Clearly, additional work is necessary for the exchanger to move electric current against a potential difference. To take this into account, consider a ligand, L, with a charge z, and suppose that there is a process that moves L from the cell interior with potential V_i to the cell exterior with potential V_e, i.e.,

$$L_i \rightarrow L_e. \tag{2.70}$$

The change in chemical potential (cf. Section 1.2) for this reaction is

$$\begin{aligned}\Delta G &= G^0_{L_e} + RT\ln([L_e]) + zFV_e - G^0_{L_i} - RT\ln([L_i]) - zFV_i \\ &= RT\ln\left(\frac{[L_e]}{[L_i]}\right) - zFV,\end{aligned} \tag{2.71}$$

where $V = V_i - V_e$ is the transmembrane potential. (The standard convention is to define the potential difference across the membrane as the internal potential minus the external potential, as discussed further in Section 2.6.1.) The standard free energy for L is the same on both sides of the membrane, so $G^0_{L_e} = G^0_{L_i}$. At equilibrium, $\Delta G = 0$, so that

$$K = \frac{[L_i]_{eq}}{[L_e]_{eq}} = \exp\left(\frac{-zFV}{RT}\right), \tag{2.72}$$

where K is the equilibrium constant for the reaction.

For the Na^+–Ca^{2+} exchanger, the overall reaction begins with three Na^+ outside the cell and one Ca^{2+} inside the cell, and ends with three Na^+ inside the cell and one Ca^{2+} outside. We can write this as

$$3Na^+_e + Ca^{2+}_i \longrightarrow 3Na^+_i + Ca^{2+}_e. \tag{2.73}$$

The change in chemical potential for this reaction is

$$\Delta G = RT\ln\left(\frac{n_i^3 c_e}{n_e^3 c_i}\right) + FV. \tag{2.74}$$

At equilibrium we must have $\Delta G = 0$, in which case

$$\frac{n^3_{i,eq} c_{e,eq}}{n^3_{e,eq} c_{i,eq}} = \exp\left(-\frac{FV}{RT}\right). \tag{2.75}$$

Recall that detailed balance requires that around any closed reaction loop the product of the forward rates must be the same as the product of the reverse rates. It follows that

$$k_1 c_{i,\mathrm{eq}} k_2 k_3 n_{e,\mathrm{eq}}^3 k_4 = n_{i,\mathrm{eq}}^3 k_{-1} k_{-4} c_{e,\mathrm{eq}} k_{-3} k_{-2}, \tag{2.76}$$

and thus

$$K_1 K_2 K_3 K_4 = \frac{c_{i,\mathrm{eq}}}{c_{e,\mathrm{eq}}} \frac{n_{e,\mathrm{eq}}^3}{n_{i,\mathrm{eq}}^3}. \tag{2.77}$$

Combining (2.76) and (2.77), we get

$$K_1 K_2 K_3 K_4 = \exp\left(\frac{FV}{RT}\right), \tag{2.78}$$

which, being independent of the concentrations, must hold in general.

It follows from (2.69) that the flux is given by

$$J = \frac{k_1 k_2 k_3 k_4 (c_i n_e^3 - e^{\frac{FV}{RT}} c_e n_i^3)}{16 \text{ positive terms}}. \tag{2.79}$$

All of the terms in the denominator are cubic products of rate constants, so that the flux J has units of inverse time. In general, the denominator of this expression also contains terms that depend on the membrane potential difference.

In writing (2.78), no assumption was made about where the charge transfer takes place. From Fig. 2.9 it might appear that the charge transfer takes place during the transitions $Y_1 \to X_1$ and $X_2 \to Y_2$. However, this is not necessarily the case. It could be that those conformational changes are accompanied by no charge transfer, but that the charge transfer occurs during other transitions. However, if we assume that one Ca^{2+} ion is transferred from inside to outside during the $X_2 \to Y_2$ transition, and three Na^+ ions are transferred during the $Y_1 \to X_1$ transition, free energy arguments yield the additional constraints

$$\frac{k_{-2}}{k_2} = \tilde{K}_2 \exp\left(\frac{-2FV}{RT}\right), \qquad \frac{k_4}{k_{-4}} = \tilde{K}_4^{-1} \exp\left(\frac{-3FV}{RT}\right), \tag{2.80}$$

where $\tilde{K}_2$ and $\tilde{K}_4$ are independent of voltage, and where $K_1 \tilde{K}_2 K_3 \tilde{K}_4 = 1$.

The most important observation is that for given n_i and n_e (set by other mechanisms such as the Na^+–K^+ ATPase discussed in the next section), a negative V enhances the rate at which the Na^+–Ca^{2+} exchanger removes Ca^{2+} from the cell. This makes sense; if V is negative, the potential inside the cell is negative compared to the outside and thus it is easier for the exchanger to move one positive charge into the cell. Since cells typically have a negative resting potential (Section 2.6), the electrogenic nature of the exchanger increases its ability to remove Ca^{2+} in resting conditions. To be specific, if the ratio of extracellular to intracellular Na^+ is 8, and the potential difference is

$V = -85$ mV (which is typical), then Ca^{2+} is removed, provided

$$\frac{c_i}{c_e} > \frac{n_i^3}{n_e^3} e^{\frac{FV}{RT}} = 7 \times 10^{-5}. \tag{2.81}$$

Notice that the difference in potential gives an improvement in the capability of the exchanger by a factor of 27 over an exchanger that is not electrogenic.

2.5 Active Transport

The carrier-mediated transport described above is always down electrochemical gradients, and so is identified with diffusion. Any process that works against gradients requires the expenditure of energy.

There are three primary means by which cells use energy to pump chemical species. The first is to keep the concentration of the cargo in the downstream domain small by binding or modifying it in some way. A binding protein in one compartment could sequester the transported cargo, or the cargo could be covalently modified in one compartment so that it no longer interacts with the transporter. For example, the flux of glucose is inward because intracellular glucose is quickly phosphorylated, thereby keeping the concentration of intracellular glucose low. However, phosphorylation of intracellular glucose requires the hydrolysis of an ATP molecule, from which the needed energy is extracted.

The second means is to use the gradient of one species to pump another species against its gradient. This is the mechanism of the Na^+–Ca^{2+} exchanger as well as numerous other exchangers that use to advantage the energy stored in the Na^+ gradient.

The third means is to regulate the binding of the cargo to the transporter in such a way that binding to the transporter is favored in one compartment and unbinding is favored in the other compartment. This change in affinity is driven by the hydrolysis of ATP or GTP. One important example of such an active (energy-consuming) exchanger is the Na^+–K^+ ATPase. This pump acts as an antiport, actively pumping Na^+ ions out of the cell and pumping K^+ ions in, each against a steep electrochemical gradient. It accomplishes this by using the energy released by the hydrolysis of ATP, and thus is called an ATPase. As is described later in this chapter, the Na^+–K^+ ATPase is important for regulating the cell volume and maintaining a membrane potential. Indeed, almost a third of the energy requirement of a typical animal cell is consumed in fueling this pump; in electrically active nerve cells, this figure approaches two-thirds of the cell's energy requirement. Other important ATPases are the sarco/endoplasmic reticulum calcium ATPase pumps (SERCA pumps) that pump Ca^{2+} into the endoplasmic or sarcoplasmic reticulum, or the plasma membrane Ca^{2+} ATPases, which pump Ca^{2+} out of the cell.

2.5.1 A Simple ATPase

We begin by considering a model of an ATPase that pumps a ligand, L, up its concentration gradient (Fig. 2.10). This hypothetical ATPase exists in six states. E is the base state; ATP can bind to E, followed by binding of the ligand L, to form the top line of states in Fig. 2.10. In each of these states the L binding site is exposed to the inside of the cell. Once ATP and L are bound, the ATPase changes conformation, exposing the L binding site to the outside, and at the same time decreasing the affinity of L for its binding site. Thus, L leaves the ATPase, followed by the hydrolysis of ATP and eventual return of the ATPase to its base state to complete the cycle. The overall cycle results in the transport of one L molecule from inside to outside. Although a realistic ATPase cycle is considerably more complicated than this one, this simple model serves to illustrate the basic principles.

If there is also a transition from E · ATP to E_e · ATP, as shown by the dashed line, then the overall cycle can break into two separate subcycles as indicated. This is called *slippage*, since each of the subcycles accomplishes nothing toward the goal of pumping L. The subcycle on the left goes naturally in a clockwise direction and hydrolyzes ATP to ADP and inorganic phosphate, P_i, without using this energy to pump L. Similarly,

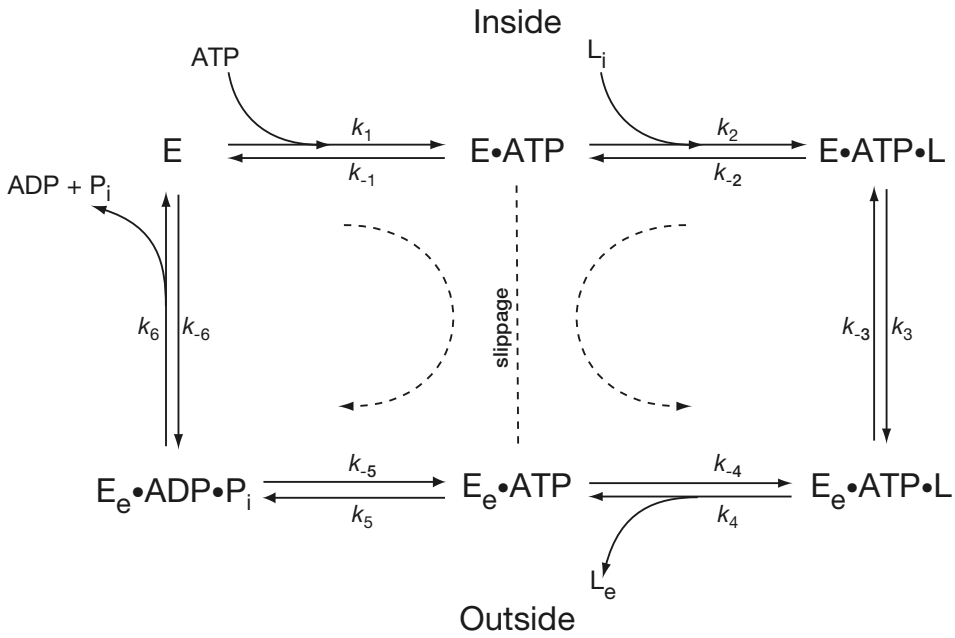


Figure 2.10 Schematic diagram of an ATPase pump that transports one ligand, L, from the inside to the outside against its concentration gradient. For each L transported, one molecule of ATP is hydrolyzed. A subscript e denotes the ATPase conformation in which the L binding sites are exposed to the exterior of the cell.

the subcycle on the right goes naturally in the direction that allows L to flow down its concentration gradient. The energy of the ATP is used to pump L against its gradient only when the ATPase proceeds around the entire cycle.

We use the law of mass action to write the differential equations for the six ATPase states. For example,

$$\frac{d[\mathrm{E}]}{dt} = k_{-1}[\mathrm{E}\cdot\mathrm{ATP}] + k_6[\mathrm{E}_e\cdot\mathrm{ADP}\cdot\mathrm{P_i}] - (k_1[\mathrm{ATP}] + k_{-6}[\mathrm{P_i}][\mathrm{ADP}])[\mathrm{E}], \tag{2.82}$$

with similar equations for each of the other states. The steady-state flux, J, is given by

$$J = k_1[\mathrm{ATP}][\mathrm{E}] - k_{-1}[\mathrm{E}\cdot\mathrm{ATP}]. \tag{2.83}$$

Even a relatively simple model of six states gives a long expression for the steady-state flux. In this case (with no slippage),

$$J = \frac{\frac{[\mathrm{ATP}][\mathrm{L}_i]}{[\mathrm{ADP}][\mathrm{P_i}][\mathrm{L}_e]} - K_1K_2K_3K_4K_5K_6}{\phi}, \tag{2.84}$$

where $\phi > 0$ is a complicated function of rate constants and concentrations, and where, as usual, $K_i = k_{-i}/k_i$. (Even though it is not obvious from the way it is written, the flux J has, as before, units of inverse time.)

Since detailed balance requires

$$\prod_{i=1}^{6} K_i = \frac{[\mathrm{L}_i]_{\mathrm{eq}}}{[\mathrm{L}_e]_{\mathrm{eq}}}\frac{[\mathrm{ATP}]_{\mathrm{eq}}}{[\mathrm{ADP}]_{\mathrm{eq}}[\mathrm{P_i}]_{\mathrm{eq}}}, \tag{2.85}$$

it follows that

$$J = \frac{\frac{[\mathrm{L}_i]}{[\mathrm{L}_e]}\frac{[\mathrm{ATP}]}{[\mathrm{ADP}][\mathrm{P_i}]} - \frac{[\mathrm{L}_i]_{\mathrm{eq}}}{[\mathrm{L}_e]_{\mathrm{eq}}}\frac{[\mathrm{ATP}]_{\mathrm{eq}}}{[\mathrm{ADP}]_{\mathrm{eq}}[\mathrm{P_i}]_{\mathrm{eq}}}}{\phi}. \tag{2.86}$$

We see from the numerator that the flux is either positive or negative depending on how far the concentrations of L, ATP, ADP, and $\mathrm{P_i}$ are from their equilibrium concentrations. In general, [ATP] is much higher than its equilibrium concentration (due to other processes in the cell that are continuously generating ATP), and it is this that causes a positive pump flux, pumping L against its gradient. However, if $[\mathrm{L}_e]$ is high enough it can force the pump to work in reverse, allowing L to move from the outside to the inside of the cell, generating ATP in the process.

To relate the rate constants to the change in free energy we use that the overall reaction is

$$\mathrm{L}_i + \mathrm{ATP} \longrightarrow \mathrm{L}_e + \mathrm{ADP} + \mathrm{P_i}. \tag{2.87}$$

The change in free energy is given by

$$\Delta G = G_{\mathrm{ADP}} + G_{\mathrm{P_i}} + G_{\mathrm{L}_e} - G_{\mathrm{ATP}} - G_{\mathrm{L}_i} \tag{2.88}$$

$$= G^0_{\mathrm{ADP}} + G^0_{\mathrm{P_i}} - G^0_{\mathrm{ATP}} - RT \ln \left(\frac{[\mathrm{ATP}][\mathrm{L}_i]}{[\mathrm{ADP}][\mathrm{P_i}][\mathrm{L}_e]} \right) \tag{2.89}$$

$$= \Delta G^0_{\mathrm{ATP}} - RT \ln \left(\frac{[\mathrm{ATP}][\mathrm{L}_i]}{[\mathrm{ADP}][\mathrm{P_i}][\mathrm{L}_e]} \right). \tag{2.90}$$

Note that the standard free energy of L is the same inside and outside the cell. At equilibrium $\Delta G = 0$ and thus

$$\Delta G^0_{\mathrm{ATP}} = RT \ln \left(\frac{[\mathrm{L}_i]_{\mathrm{eq}}}{[\mathrm{L}_e]_{\mathrm{eq}}} \frac{[\mathrm{ATP}]_{\mathrm{eq}}}{[\mathrm{ADP}]_{\mathrm{eq}}[\mathrm{P_i}]_{\mathrm{eq}}} \right). \tag{2.91}$$

Combining this with (2.85) gives

$$\prod_{i=1}^{6} K_i = e^{\frac{\Delta G^0_{\mathrm{ATP}}}{RT}} \mathrm{M}^{-1}, \tag{2.92}$$

which, because it is independent of concentrations, must hold in general. Notice from (2.91) that both sides of (2.92) must have units of liters per mole, or M^{-1}. In fact, since the free energy released by the hydrolysis of ATP is well known to be -31 kJ mole^{-1}, it follows that

$$e^{\frac{\Delta G^0_{\mathrm{ATP}}}{RT}} = 3.73 \times 10^{-6}. \tag{2.93}$$

Now from (2.84) it follows that

$$J = \frac{\frac{[\mathrm{L}_i]}{[\mathrm{L}_e]} \frac{[\mathrm{ATP}]}{[\mathrm{ADP}][\mathrm{P_i}]} - e^{\frac{\Delta G^0_{\mathrm{ATP}}}{RT}} \mathrm{M}^{-1}}{\phi}. \tag{2.94}$$

2.5.2 Active Transport of Charged Ions

Suppose that the interior of the cell has an electric potential of V_i while the exterior has a potential of V_e, and suppose further that L has a charge z. Then the change in potential of the ATPase cycle (2.87) is

$$\Delta G = \Delta G^0_{\mathrm{ATP}} - RT \ln \left(\frac{[\mathrm{ATP}][\mathrm{L}_i]}{[\mathrm{ADP}][\mathrm{P_i}][\mathrm{L}_e]} \right) - zFV, \tag{2.95}$$

where $V = V_i - V_e$ is, as usual, the membrane potential difference.

An identical argument to before shows that

$$\prod_{i=1}^{6} K_i = e^{\frac{\Delta^0 G_{\mathrm{ATP}}}{RT}} e^{\frac{-zFV}{RT}} \mathrm{l\ mole}^{-1}, \tag{2.96}$$

and thus

$$J = \frac{\frac{[L_i]}{[L_e]}\frac{[ATP]}{[ADP][P_i]} - e^{\frac{\Delta^0 G_{ATP}}{RT}} e^{\frac{-zFV}{RT}} M^{-1}}{\phi(V)}. \tag{2.97}$$

Note that the denominator is now a function of V also, since ϕ depends on the rate constants, which are themselves functions of V, and thus the precise dependence of J on V is not immediately clear. However, if $z > 0$ and $V > 0$, the flux is zero at lower concentrations of L_i, while if $V < 0$, the flux is zero at higher concentrations of L_i. Thus we conclude that a positive membrane potential makes it easier for the pump to move positive ions from inside to outside, while a negative membrane potential makes this more difficult. Although this is not a rigorous argument, a more detailed calculation shows that this result holds in general.

Although this thermodynamic argument shows that there must be some voltage-dependence in the rate constants, it does not tell us in which step (or steps) the voltage-dependence occurs. For example, in this model, the transition from $E \cdot ATP \cdot L$ to $E_e \cdot ATP \cdot L$ involves the net movement of the charge across the cell membrane, so that $\frac{k_{-3}}{k_3} = e^{\frac{-zFV}{RT}}$. (The argument here is identical to the argument used for the voltage-dependence of the Na^+–Ca^{2+} exchanger.) However, there are other possibilities. Although each model must have the same solution when $J = 0$, and the expressions for J must have the same sign, the models can have significantly different expressions for $\phi(V)$.

2.5.3 A Model of the Na^+–K^+ ATPase

One of the best-known ATPases is the Na^+–K^+ ATPase, which pumps K^+ into the cell and Na^+ out of the cell through the overall reaction scheme

$$ATP + 3Na_i^+ + 2K_e^+ \longrightarrow ADP + P_i + 3Na_e^+ + 2K_i^+. \tag{2.98}$$

It is an electrogenic pump (each pump cycle transfers one positive charge from inside to out) and a member of the family of P-type active cation transporters which includes the SERCA ATPases that are discussed at length in Chapter 7. A great deal of work has been done to determine the mechanisms that underlie Na^+ and K^+ transport by this ATPase; the most widely accepted model is the Post–Albers model which was developed by two independent groups in the 1960s (Albers et al., 1963; Charnock and Post, 1963). A more recent review is Apell (2004), while a history of investigations into the Na^+–K^+ ATPase is given by Glynn (2002). An excellent mathematical implementation of the Post–Albers scheme is that of Smith and Crampin (2004), and this is the model that we follow here.

In the Post–Albers scheme, phosphorylation of the pump (i.e., exchange of ATP for ADP) is associated with Na^+ efflux, while hydrolysis (i.e., loss of the additional phosphate group) is associated with K^+ influx. During the transition across the membrane each ion type is occluded, i.e., bound to the pump in a conformation in which

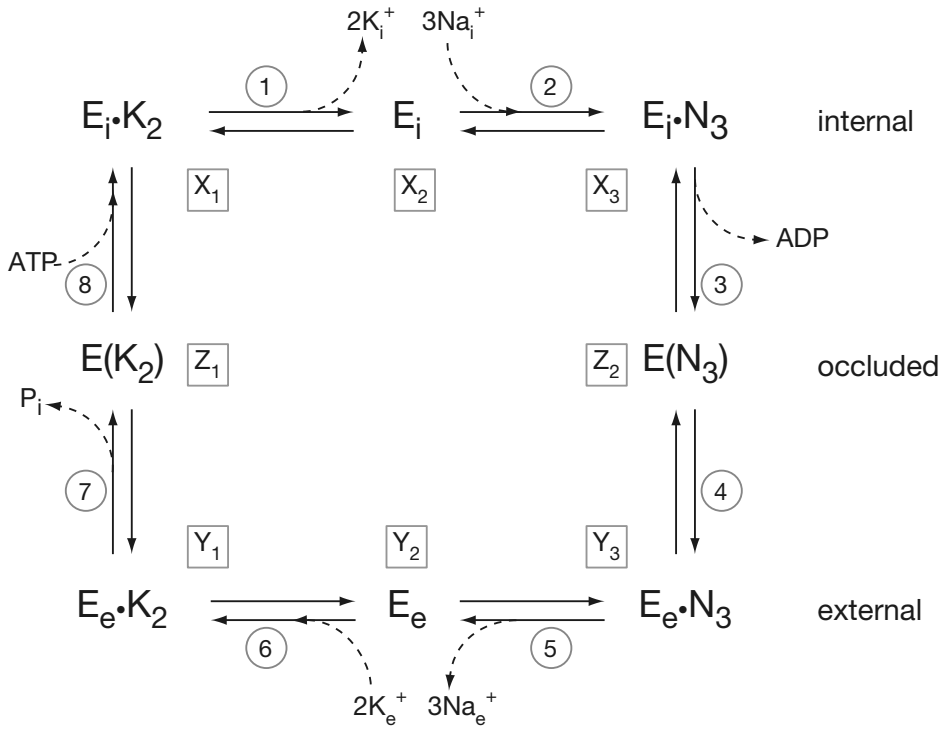


Figure 2.11 Model of the Na^+–K^+ ATPase based on the Post–Albers scheme.

it is accessible from neither side of the membrane. Occlusion prevents slippage, thus increasing the efficiency of the pump.

This scenario is illustrated in Fig. 2.11. Starting at the top left of the figure (state X_1), the ATPase begins in the conformation E_i, in which the binding sites for Na^+ and K^+ are exposed to the inside of the cell. The ATPase then loses two K^+ ions (which is assumed to occur in a single step) and gains three Na^+ ions, again in a single step, to move through states X_2 and X_3. ATP remains bound to the pump in each of the states X_1, X_2, and X_3, although this is not shown explicitly in the diagram. Loss of ADP then drives the ATPase to the occluded state Z_2, in which the three Na^+ ions are inaccessible to both the inside and outside of the cell. After another conformational change to the E_e state, in which the Na^+ and K^+ binding sites are exposed to the outside of the cell, the ATPase loses its three Na^+ to the outside, picks up another two K^+, and loses its extra phosphate to move through to the occluded state Z_1, in which the K^+ ions are shielded. Binding of ATP then returns the ATPase to the E_i conformation to complete the cycle. The rate constants are not shown explicitly, but each transition between states is labeled by a circled number. For each $i = 1, \ldots, 8$, transition i has two rate constants, k_i in the clockwise direction and k_{-i} in the counterclockwise direction.

From this diagram we can easily write the differential equations for each of the ATPase states. For example, letting a lowercase letter denote the fraction of the ATPase in that state, we have

$$\frac{dx_1}{dt} = k_{-1}[\mathrm{K}_i^+]^2 x_2 + k_8[\mathrm{ATP}]z_1 - (k_{-8} + k_1)x_1, \tag{2.99}$$

and so on, subject to the constraint $x_1 + x_2 + x_3 + y_1 + y_2 + y_3 + z_1 + z_2 = 1$. The resultant expression for the flux is long and unwieldy, of the form

$$J = \frac{[\mathrm{ATP}]n_i^3\kappa_e^2 - [\mathrm{ADP}][\mathrm{P_i}]\kappa_i^2 n_e^3(\prod_{i=1}^{8} K_i)}{\phi}, \tag{2.100}$$

where ϕ is the sum of a large number of terms involving products of the rate constants and concentrations. Here, n denotes the Na^+ concentration, and κ denotes the K^+ concentration. This expression for the flux is similar to that derived in the simpler model of Section 2.5.1. The same thermodynamic constraints apply, and so some of the rate constants are functions of the membrane potential. Smith and Crampin (2004), following the ideas of Apell (1989), incorporate voltage dependence into the rate constants for Na^+ binding and unbinding, i.e., K_2 and K_5 in this model.

A simplified version of this model is discussed in Exercise 13.

2.5.4 Nuclear Transport

The transport of proteins from the cytoplasm to the nucleus (or the reverse) is accomplished by means that combine features of each of the above transport mechanisms. The nuclear membrane contains protein structures called nuclear pore complexes (NPCs) that allow free diffusion of soluble carrier proteins. However, these carrier proteins can pass through the pore complex only when they are bound. These carrier proteins (called importins) recognize and readily bind cargo destined for translocation. The energy to transport cargo against its gradient is provided by the hydrolysis of GTP via a GTPase enzyme called Ran. Ran-GTP has a very high binding affinity for the carrier protein ($\Delta G^0 = -51$ kJ mol^{-1}), effectively excluding the cargo from binding. The transportin/Ran-GTP complex is disassembled by the hydrolysis of Ran-GTP ($\Delta G^0 = -33$ kJ mol^{-1}) to Ran-GDP, which has a binding affinity for the carrier protein that is 10,000-fold lower than that of Ran-GTP. The endogenous GTPase activity rate is extremely slow ($k_{\mathrm{cat}} = 1.5 \times 10^{-5}\mathrm{s}^{-1}$). However, the hydrolysis of GTP to GDP on Ran is catalyzed by a cytoplasmic GTPase-activating protein called RanGAP, which accelerates this rate by as much as 500,000-fold.

One cycle of transport works as follows. Cargo in the cytoplasm that is targeted for transport binds to the carrier molecule and moves via diffusion through the NPC. In the nucleus, when the cargo unbinds, Ran-GTP quickly binds to the carrier, preventing the cargo from rebinding. Ran-GTP is kept at high concentration in the nucleus by another mechanism, so the Ran-GTP carrier complex diffuses into the cytoplasm through the NPC. On the cytoplasmic side of the membrane, Ran-GTP is quickly hydrolyzed to

Ran-GDP, which because of its much lower binding affinity, unbinds from the carrier molecule, completing the cycle. Although all the reactions are reversible, the directionality is maintained by the free energy of GTP hydrolysis and the high concentration of GTP in the nucleus.

We leave the development of a model of this transport mechanism to the interested reader. In the absence of RanGAP, the model is similar to that of the Na^+–Ca^{2+} transporter described above. On the other hand, if the hydrolysis of Ran-GTP to RanGDP is assumed to be so fast that there is no unbinding of RanGTP in the cytoplasm, then the model is similar to that of the simple ATPase described above.

2.6 The Membrane Potential

The principal function of the active ATPase transport processes described above is to regulate the intracellular ionic composition of the cell. For example, the operation of the Na^+–K^+ ATPase results in high intracellular K^+ concentrations and low intracellular Na^+ concentrations. This is necessary for a cell's survival, as without such regulation, cells swell and burst. However, before we consider models of cell volume regulation, we consider the effects of ionic separation.

2.6.1 The Nernst Equilibrium Potential

One of the most important equations in electrophysiology is the Nernst equation, which describes how a difference in ionic concentration can result in a potential difference across the membrane separating the two concentrations.

Suppose there are two reservoirs containing the same ion S, but at different concentrations, as shown schematically in Fig. 2.12. The reservoirs are separated by a

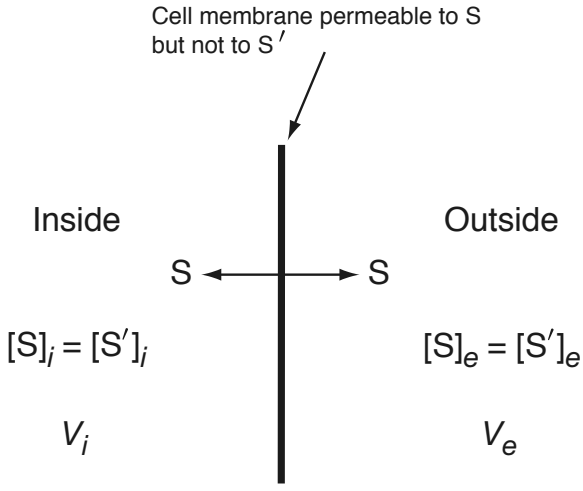


Figure 2.12 Schematic diagram of a membrane separating two solutions with different ionic concentrations. By convention, the potential difference, V, across the membrane is defined to be $V = V_i - V_e$.

semipermeable membrane. The solutions on each side of the membrane are assumed to be electrically neutral (at least initially), and thus each ion S is balanced by another ion, S′, with opposite charge. For example, S could be Na^+, while S′ could be Cl^-. Because we ultimately wish to apply the Nernst equation to cellular membranes, we call the left of the membrane the inside and the right the outside of the cell.

If the membrane is permeable to S but not to S′, the concentration difference across the membrane results in a net flow of S from one side to another, down its concentration gradient. However, because S′ cannot diffuse through the membrane, the diffusion of S causes a buildup of charge across the membrane. This charge imbalance, in turn, sets up an electric field that opposes the further net movement of S through the membrane. Equilibrium is reached when the electric field exactly balances the diffusion of S. Note that at steady state there are more S ions than S′ ions on one side and fewer S ions than S′ ions on the other, and thus neither side of the membrane is exactly electrically neutral. However, because the force from the charge buildup is so strong, only a small amount of S moves across the membrane. To a good approximation, the concentrations of S on either side of the membrane remain unchanged, the solutions on either side of the membrane remain electrically neutral, and the small excess charge accumulates near the interface. The region in which there is a charge imbalance is called the Debye layer, and is on the order of a few nanometers thick.

The chemical potential of S on the inside of the membrane is

$$G_{\mathrm{S},i} = G^0_{\mathrm{S}} + RT\ln([\mathrm{S}]_i) + zFV_i, \tag{2.101}$$

while on the outside it is

$$G_{\mathrm{S},e} = G^0_{\mathrm{S}} + RT\ln([\mathrm{S}]_e) + zFV_e. \tag{2.102}$$

The chemical potential difference is

$$\Delta G_{\mathrm{S}} = G_{\mathrm{S},i} - G_{\mathrm{S},e} = RT\ln\left(\frac{[\mathrm{S}]_i}{[\mathrm{S}]_e}\right) + zFV. \tag{2.103}$$

At equilibrium, it must be that $\Delta G_{\mathrm{S}} = 0$, and thus the equilibrium potential difference, V_{S}, across the membrane must be

$$V_{\mathrm{S}} = \frac{RT}{zF}\ln\left(\frac{[\mathrm{S}]_e}{[\mathrm{S}]_i}\right) = \frac{kT}{zq}\ln\left(\frac{[\mathrm{S}]_e}{[\mathrm{S}]_i}\right), \tag{2.104}$$

called the *Nernst potential*. Here k is Boltzmann's constant $k = \frac{R}{N_A}$, N_A is Avogadro's number, q is the charge on a proton, and z is the charge on the ion S. When $V = V_{\mathrm{S}}$, there is no net current of S across the membrane, as the tendency of ions to move down their gradient is exactly balanced by the electric potential difference.

Throughout this book we follow the usual convention and define the potential difference, V, across the membrane to be

$$V = V_i - V_e, \tag{2.105}$$

i.e., the intracellular potential minus the extracellular potential.

Typical concentrations (in this case, for squid axon) are 397, 50, and 40 mM for K^+, Na^+, and Cl^-, respectively, in the intracellular space, and 20, 437, and 556 mM in the extracellular space. With these concentrations, the Nernst potentials for squid nerve axon are $V_{Na} = 56$ mV, $V_K = -77$ mV, $V_{Cl} = -68$ mV (using $RT/F = 25.8$ mV at 27°C. See Table 2.1).

The Nernst equation is independent of how the ions move through the membrane and depends only on the ratio of concentrations. In this sense, it is a universal law (although because it was derived from an ideal, yet approximate, law, it too is approximate). Any equation that expresses the transmembrane current of S in terms of the membrane potential, no matter what its form, must have the reversal potential of V_S; i.e., the current must be zero at the Nernst potential $V = V_S$. However, although this is true when a single ion species crosses the membrane, the situation is considerably more complicated when more than one type of ion can cross the membrane. In this case, the membrane potential that generates zero total current does not necessarily have zero current for each individual ion. For example, a current of S in one direction might be balanced by a current of S′ in the same direction. Hence, when multiple ion types can diffuse through the membrane, the concentrations are not, in general, at equilibrium, even when there is no total current. Therefore, the arguments of chemical equilibrium used to derive the Nernst equation cannot be used, and there is no universal expression for the reversal potential in the multiple ion case. In this case, the reversal potential depends on the model used to describe the individual transmembrane ionic flows (see Chapter 3).

2.6.2 Gibbs–Donnan Equilibrium

Suppose one side of the membrane contains large charged macromolecules that cannot cross the membrane, but that both of the ion species S and S′ freely diffuse across the membrane. To be specific, suppose that the macromolecules are negatively charged with valence $-z_x$, S is positively charged with valence z, and S′ is negatively charged with valence $-z$. Note that both z and z_x are defined to be positive.

Outside the cell, S and S′ must have the same concentration, to maintain charge neutrality. Inside, charge neutrality requires more S than S′, in order to balance the negative charge on the macromolecules. At equilibrium, the membrane potential must be the Nernst potential for both S and S′, namely

$$V_S = \frac{RT}{zF} \ln\left(\frac{[S]_e}{[S]_i}\right) = -\frac{RT}{zF} \ln\left(\frac{[S']_e}{[S']_i}\right), \tag{2.106}$$

where

$$z_x[X] + z[S']_i = z[S]_i \quad \text{and} \quad [S']_e = [S]_e. \tag{2.107}$$

It follows that

$$[S']_e[S]_e = [S']_i[S]_i, \tag{2.108}$$

and thus

$$[\mathrm{S}]_i \left([\mathrm{S}]_i - \frac{z_x}{z}[\mathrm{X}] \right) - ([\mathrm{S}]_e)^2 = 0. \tag{2.109}$$

Now, we assume that the external concentration is fixed, and treat $[\mathrm{S}]_e$ as a known parameter. In this case, (2.109) can be solved to find a unique positive value for $[\mathrm{S}]_i$,

$$[\mathrm{S}]_i = \sigma[\mathrm{S}]_e, \qquad \sigma = \frac{1}{2}(Z + \sqrt{Z^2 + 4}), \tag{2.110}$$

where $Z = \frac{z_x[\mathrm{X}]}{z[\mathrm{S}]_e}$, and from this the transmembrane potential can be determined using (2.106).

If, instead, we have a fixed volume and a fixed total amount of S, say, then we use the constraint

$$v_i[\mathrm{S}]_i + v_e[\mathrm{S}]_e = [\mathrm{S}]_{\mathrm{tot}}, \tag{2.111}$$

where $[\mathrm{S}]_{\mathrm{tot}}$ is a constant, and v_i and v_e are the internal and external volumes, respectively. We can now solve (2.109) subject to this constraint to find $[\mathrm{S}]_i$ and the transmembrane potential. Note, however, that a physical solution is not always possible in this case, as there may be insufficient S or S′ to reach equilibrium.

This equilibrium is called the *Gibbs–Donnan equilibrium* (Exercise 15). The potential difference generated in this way is known to occur across cell membranes and also across the edge of a gel in aqueous solution. This potential drop occurs across the edge of a gel if the charged macromolecules are immobilized in the gel, and therefore unable to diffuse out of the gel.

2.6.3 Electrodiffusion: The Goldman–Hodgkin–Katz Equations

In general, the flow of ions through the membrane is driven by concentration gradients and also by the electric field. The contribution to the flow from the electric field is given by *Planck's equation*

$$\mathbf{J} = -u\frac{z}{|z|}c\nabla\phi, \tag{2.112}$$

where u is the *mobility* of the ion, defined as the velocity of the ion under a constant unit electric field; z is the valence of the ion, so that $z/|z|$ is the sign of the force on the ion; c is the concentration of S; and ϕ is the electrical potential, so that $-\nabla\phi$ is the electrical field.

There is a relationship, determined by Einstein, between the ionic mobility u and Fick's diffusion constant:

$$D = \frac{uRT}{|z|F}. \tag{2.113}$$

When the effects of concentration gradients and electrical gradients are combined, we obtain the *Nernst–Planck equation*

$$\mathbf{J} = -D\left(\nabla c + \frac{zF}{RT}c\nabla\phi\right). \tag{2.114}$$

If the flow of ions and the electric field are transverse to the membrane, (2.114) can be viewed as the one-dimensional relation

$$J = -D\left(\frac{dc}{dx} + \frac{zF}{RT}c\frac{d\phi}{dx}\right). \tag{2.115}$$

The Nernst Equation

The Nernst equation can also be derived from the Nernst–Planck equation (2.115). When the flux J is zero, we obtain

$$-D\left(\frac{dc}{dx} + \frac{zF}{RT}c\frac{d\phi}{dx}\right) = 0, \tag{2.116}$$

so that

$$\frac{1}{c}\frac{dc}{dx} + \frac{zF}{RT}\frac{d\phi}{dx} = 0. \tag{2.117}$$

Now suppose that the cell membrane extends from $x = 0$ (the inside) to $x = L$ (the outside), and let subscripts i and e denote internal and external quantities respectively. Then, integrating (2.117) from $x = 0$ to $x = L$ we get

$$\ln(c)\Big|_{c_i}^{c_e} = \frac{zF}{RT}(\phi_i - \phi_e), \tag{2.118}$$

and thus the potential difference across the membrane, $V = \phi_i - \phi_e$, is given by

$$V = \frac{RT}{zF}\ln\left(\frac{c_e}{c_i}\right), \tag{2.119}$$

which is the Nernst equation.

The Constant Field Approximation

In general, the electric potential ϕ is determined by the local charge density, and so, if it is not zero, J must be found by solving a coupled system of equations (discussed in detail in Chapter 3). However, a useful result is obtained by assuming that the electric field in the membrane is constant, and thus decoupled from the effects of charges moving through the membrane. Suppose two reservoirs are separated by a semipermeable membrane of thickness L, such that the potential difference across the membrane is V. On the left of the membrane (the inside) $[\mathrm{S}] = c_i$, and on the right (the outside) $[\mathrm{S}] = c_e$. If the electric field is constant through the membrane, $\partial\phi/\partial x = -V/L$, where $V = \phi(0) - \phi(L)$ is the membrane potential.

At steady state and with no production of ions, the flux must be constant. In this case, the Nernst–Planck equation (2.114) is an ordinary differential equation for the concentration c,

$$\frac{dc}{dx} - \frac{zFV}{RTL}c + \frac{J}{D} = 0, \tag{2.120}$$

whose solution is

$$\exp\left(\frac{-zVFx}{RTL}\right)c(x) = \frac{JRTL}{DzVF}\left[\exp\left(\frac{-zVFx}{RTL}\right) - 1\right] + c_i, \tag{2.121}$$

where we have used the left boundary condition $c(0) = c_i$. To satisfy the boundary condition $c(L) = c_e$, it must be that

$$J = \frac{D}{L}\frac{zFV}{RT}\frac{c_i - c_e\exp\left(\frac{-zVF}{RT}\right)}{1 - \exp\left(\frac{-zVF}{RT}\right)}, \tag{2.122}$$

where J is the flux density with units (typically) of moles per area per unit time. Note that these units are equivalent to units of concentration $\times$ speed. This flux density becomes an electrical current density (current per unit area) when multiplied by zF, the amount of charge carried per mole, and thus

$$I_S = P_S\frac{z^2F^2}{RT}V\frac{c_i - c_e\exp\left(\frac{-zFV}{RT}\right)}{1 - \exp\left(\frac{-zFV}{RT}\right)}, \tag{2.123}$$

where $P_S = D/L$ is the permeability of the membrane to S. This is the famous Goldman–Hodgkin–Katz (GHK) current equation, and plays an important role in models of cellular electrical activity. Notice that the GHK flux (2.122) reduces to Fick's law (2.14) in the limit $V \to 0$.

The current is zero if the diffusively driven flow and the electrically driven flow are in balance, which occurs, provided that $z \neq 0$, if

$$V = V_S = \frac{RT}{zF}\ln\left(\frac{c_e}{c_i}\right), \tag{2.124}$$

which is, as expected, the Nernst potential.

If there are several ions that are separated by the same membrane, then the flow of each of these is governed separately by its own current–voltage relationship. In general there is no potential at which these currents are all individually zero. However, the potential at which the net electrical current is zero is called the Goldman–Hodgkin–Katz potential. For a collection of ions all with valence $z = \pm 1$, we can calculate the GHK potential directly. For zero net electrical current, it must be that

$$0 = \sum_{z=1} P_j\frac{c_i^j - c_e^j\exp\left(\frac{-VF}{RT}\right)}{1 - \exp\left(\frac{-VF}{RT}\right)} + \sum_{z=-1} P_j\frac{c_i^j - c_e^j\exp\left(\frac{VF}{RT}\right)}{1 - \exp\left(\frac{VF}{RT}\right)}, \tag{2.125}$$

where $P_j = D_j/L$. This expression can be solved for V, to get

$$V = -\frac{RT}{F} \ln \left(\frac{\sum_{z=-1} P_j c_e^j + \sum_{z=1} P_j c_i^j}{\sum_{z=-1} P_j c_i^j + \sum_{z=1} P_j c_e^j} \right). \tag{2.126}$$

For example, if the membrane separates Na^+ ($z = 1$), K^+ ($z = 1$), and Cl^- ($z = -1$) ions, then the GHK potential is

$$V_r = -\frac{RT}{F} \ln \left(\frac{P_{Na}[Na^+]_i + P_K[K^+]_i + P_{Cl}[Cl^-]_e}{P_{Na}[Na^+]_e + P_K[K^+]_e + P_{Cl}[Cl^-]_i} \right). \tag{2.127}$$

It is important to emphasize that neither the GHK potential nor the GHK current equation are universal expressions like the Nernst equation. Both depend on the assumption of a constant electric field, and other models give different expressions for the transmembrane current and reversal potential. In Chapter 3 we discuss other models of ionic current, and compare them to the GHK equations. However, the importance of the GHK equations is so great, and their use so widespread, that their separate presentation here is justified.

2.6.4 Electrical Circuit Model of the Cell Membrane

Since the cell membrane separates charge, it can be viewed as a capacitor. The capacitance of any insulator is defined as the ratio of the charge across the capacitor to the voltage potential necessary to hold that charge, and is denoted by

$$C_m = \frac{Q}{V}. \tag{2.128}$$

From standard electrostatics (Coulomb's law), one can derive the fact that for two parallel conducting plates separated by an insulator of thickness d, the capacitance is

$$C_m = \frac{k\epsilon_0}{d}, \tag{2.129}$$

where k is the dielectric constant for the insulator and ϵ_0 is the permittivity of free space. The capacitance of cell membrane is typically 1.0 μF/cm^2. Using that $\epsilon_0 = (10^{-9}/(36\pi))$F/m, we calculate that the dielectric constant for cell membrane is about 8.5, compared to $k = 3$ for oil.

A simple electrical circuit model of the cell membrane is shown in Fig. 2.13. It is assumed that the membrane acts like a capacitor in parallel with a resistor (although not necessarily ohmic). Since the current is dQ/dt, it follows from (2.128) that the capacitive current is $C_m dV/dt$, provided that C_m is constant. Since there can be no net buildup of charge on either side of the membrane, the sum of the ionic and capacitive currents must be zero, and so

$$C_m \frac{dV}{dt} + I_{ion} = 0, \tag{2.130}$$

where, as usual, $V = V_i - V_e$.

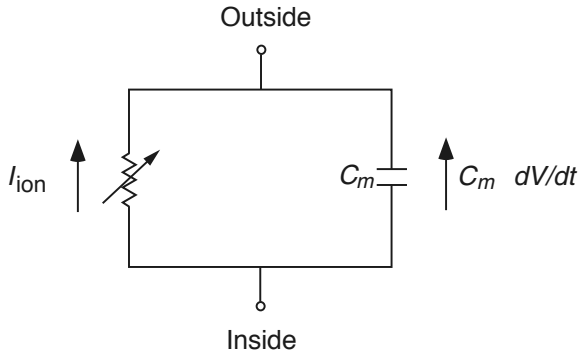


Figure 2.13 Electrical circuit model of the cell membrane.

This equation appears many times in this book, as it is the basis for much of theoretical electrophysiology. A significant challenge is to determine the form of I_{ion}. We have already derived one possible choice, the GHK current equation (2.123), and others are discussed in Chapter 3.

Another common model describes I_{ion} as a linear function of the membrane potential. In Chapter 3 we show how a linear I–V curve can be derived from more realistic models; however, because it is used so widely, we present a brief, heuristic, derivation here. Consider the movement of an ion S across a membrane. We assume that the potential drop across the membrane has two components. First, the potential drop due to concentration differences is given by the Nernst equation

$$V_{\text{S}} = \frac{RT}{zF} \ln\left(\frac{[\text{S}]_e}{[\text{S}]_i}\right), \tag{2.131}$$

and, second, if the channel is ohmic, the potential drop due to an electrical current is rI_{S}, where r is the channel resistance and I_{S} is the transmembrane current (positive outward) of S. Summing these two contributions we obtain

$$V = rI_{\text{S}} + V_{\text{S}}, \tag{2.132}$$

and solving for the current, we get the current–voltage relationship

$$I_{\text{S}} = g(V - V_{\text{S}}), \tag{2.133}$$

where $g = 1/r$ is the *membrane conductance*. The current I_{S} and conductance g are usually specified per unit area of membrane, being the product of the single channel conductance times the number of channels per unit area of membrane.

Notice that this current–voltage relationship also has zero current when $V = V_S$, the Nernst potential, as it must.

2.7 Osmosis

Suppose two chambers of water are separated by a rigid porous membrane. Because it is porous, water can flow between the two chambers. If the two chambers are topped by pistons, then water can be driven between the two chambers by applying different pressures to the two pistons. In general there is a linear relationship between the pressure difference and the flux of water through the membrane, given by

$$rQ = P_1 - P_2, \tag{2.134}$$

where Q is the flux (volume per unit time) of water from chamber one to chamber two, P_1 and P_2 are the applied pressures for chambers one and two, respectively, and r is the flow resistance of the membrane (not the same as the resistance to flow of ions). The expression (2.134) is actually a definition of the flow resistance r, and this linear relationship is analogous to Ohm's law relating current and voltage in a conductor. It is useful but not universally correct.

Suppose that a solute is added to chamber one, and that the membrane is impermeable to the solute. The difference in free energy per mole (or chemical potential) of solvent (i.e., water) between the two chambers is

$$\Delta G = RT \ln \frac{S_1}{S_2}, \tag{2.135}$$

where S_i is the mole fraction of solvent in the ith chamber. Note that because this expression involves the ratio of S_1 to S_2, we can use whatever units are most convenient. Hence we use mole fraction rather than concentration, which is standard. Because it dilutes the solvent ($S_1 < S_2$), the presence of a solute lowers the chemical potential of the solvent and induces a flow of solvent from chamber two to chamber one. In other words, the solvent diffuses from a region of higher concentration to one of lower concentration.

At constant temperature, equilibrium can be attained either by diluting the solution until it is pure solvent, or by increasing the pressure on the solution. The *osmotic pressure* π_s is defined to be the pressure that must be applied to chamber 1 to bring the free energy back to the free energy of the pure solvent. It follows that

$$RT \ln \frac{S_1}{S_2} + \pi_s v_s = 0, \tag{2.136}$$

where v_s is the molar volume (liters per mole) of the solvent. Note that, from the ideal gas law $PV = nRT$, we see that $\pi_s v_s$ has the same units as RT.

Since $S_2 = 1$ it now follows that

$$\pi_s = -\frac{RT}{v_s} \ln(S_1) = -\frac{RT}{v_s} \ln(1 - N) \approx \frac{RT}{v_s} N, \tag{2.137}$$

where N is the mole fraction of solvent. Also, since $N = \frac{n}{n+n_s} \approx \frac{n}{n_s}$, where n and n_s are the number of moles of solute and solvent, respectively, we have that

$$\pi_s = \frac{RT}{v_s}\frac{n}{n_s} \approx RcT, \tag{2.138}$$

since $n_s v_s$ is quite close to the volume, v, of solution. Here c is the concentration of solvent in units of moles per liter. Using that $c = n/v$, (2.138) becomes

$$\pi_s v = nRT, \tag{2.139}$$

which is the same as the ideal gas law. Equation (2.138) was first found empirically by van't Hoff.

If n has the units of numbers of molecules per liter, rather than moles per liter, as above, then (2.139) becomes

$$\pi_s v = nkT. \tag{2.140}$$

As with all things derived from ideal properties, the expression (2.139) is an approximation, and for real solutions at physiological concentrations, the deviation can be significant. The formula

$$\pi_s v = \phi nRT, \tag{2.141}$$

works much better, where ϕ is a concentration-dependent correction factor found experimentally. For all solutes, ϕ approaches one for sufficiently small concentrations. At concentrations typical of extracellular fluids in mammals, $\phi = 1.01$ for glucose and lactose, whereas for NaCl and KCl, $\phi = 0.93$ and 0.92, respectively. Deviation from ideality is even more significant for proteins, and is typically more concentration dependent as well. In spite of this, in the remainder of this book we use van't Hoff's law (2.138) to calculate osmotic pressure.

Notice that π_s is not the pressure of the solute but rather the pressure that must be applied to the solution to prevent solvent from flowing in through the semipermeable membrane. Thus, the flow rate of solvent is modified by osmotic pressure to be

$$rQ = P_1 - \pi_s - P_2, \tag{2.142}$$

The flux of water due to osmotic pressure is called *osmosis*. The effect of the osmotic pressure is to draw water into chamber one, causing an increase in its volume and thereby to decrease the concentration of solute.

Osmotic pressure is determined by the number of particles per unit volume of fluid, and not the mass of the particles. The unit that expresses the concentration in terms of number of particles is called the *osmole*. One osmole is 1 gram molecular weight (that is, one mole) of an undissociated solute. Thus, 180 grams of glucose (1 gram molecular weight) is 1 osmole of glucose, since glucose does not dissociate in water. On the other hand, 1 gram molecular weight of sodium chloride, 58.5 grams, is 2 osmoles, since it dissociates into 2 moles of osmotically active ions in water.

A solution with 1 osmole of solute dissolved in a kilogram of water is said to have osmolality of 1 osmole per kilogram. Since it is difficult to measure the amount of water

in a solution, a more common unit of measure is osmolarity, which is the osmoles per liter of aqueous solution. In dilute conditions, such as in the human body, osmolarity and osmolality differ by less than one percent. At body temperature, 37° C, a concentration of 1 osmole per liter of water has an osmotic pressure of 19,300 mm Hg, which corresponds to a column of water over 250 meters high. Clearly, osmotic pressures can be very large. It is for this reason that red blood cells burst when the blood serum is diluted with pure water, and this is known to have been the cause of death in hospital patients when pure water was accidentally injected into the veins.

Suppose two columns (of equal cross-section) of water are separated at the bottom by a rigid porous membrane. If n molecules of sugar are dissolved in column one, what will be the height difference between the two columns after they achieve steady state? At steady state there is no flux between the two columns, so at the level of the membrane, $P_1 - \pi_s = P_2$. Since P_1 and P_2 are related to the height of the column of water through $P = \rho g h$, where ρ is the density of the fluid, g is the gravitational constant, and h is the height of the column. We suppose that the density of the two columns is the same, unaffected by the presence of the dissolved molecule, so we have

$$\rho g h_2 = \rho g h_1 - \frac{nkT}{h_1 A}, \tag{2.143}$$

where A is the cross-sectional area of the columns. Since fluid is conserved, $h_1 + h_2 = 2h_0$, where h_0 is the height of the two columns of water before the sugar was added. From these, we find a single quadratic equation for h_1:

$$h_1^2 - h_0 h_1 - \frac{nkT}{2\rho g A} = 0. \tag{2.144}$$

The positive root of this equation is $h_1 = h_0/2 + \frac{1}{2}\sqrt{h_0^2 + \frac{2nkT}{\rho g A}}$, so that

$$h_1 - h_2 = \sqrt{h_0^2 + \frac{2nkT}{\rho g A}} - h_0. \tag{2.145}$$

When the solute is at a high enough concentration, physical solutions of (2.145) are not possible. Specifically, if the solute is too concentrated with $\frac{nkT}{\rho g A} > 4h_0^2$, the weight of a column of water of height $2h_0$ is insufficient to balance the osmotic pressure, in which case there is not enough water to reach equilibrium.

2.8 Control of Cell Volume

The principal function of the ionic pumps is to set up and maintain concentration differences across the cell membrane, concentration differences that are necessary for the cell to control its volume. In this section we describe how this works by means of a simple model in which the volume of the cell is regulated by the balance between ionic pumping and ionic flow down concentration gradients (Tosteson and Hoffman, 1960; Jakobsson, 1980; Hoppensteadt and Peskin, 2001).

Because the cell membrane is a thin lipid bilayer, it is incapable of withstanding any hydrostatic pressure differences. This is a potentially fatal weakness. For a cell to survive, it must contain a large number of intracellular proteins and ions, but if their concentrations become too large, osmosis causes the entry of water into the cell, causing it to swell and burst (this is what happens to many cells when their pumping machinery is disabled). Thus, for cells to survive, they must regulate their intracellular ionic composition (Macknight, 1988).

An even more difficult problem for some cells is to transport large quantities of water, ions, or other molecules while maintaining a steady volume. For example, Na^+-transporting epithelial cells, found (among other places) in the bladder, the colon, and nephrons of the kidney, are designed to transport large quantities of Na^+ from the lumen of the gut or the nephron to the blood. Indeed, these cells can transport an amount of Na^+ equal to their entire intracellular contents in one minute. However, the rate of transport varies widely, depending on the concentration of Na^+ on the mucosal side. Thus, these cells must regulate their volume and ionic composition under a wide variety of conditions and transport rates (Schultz, 1981).

2.8.1 A Pump–Leak Model

We begin by modeling the active and passive transport of ionic species across the cell membrane. We have already derived two equations for ionic current as a function of membrane potential: the GHK current equation (2.123) and the linear relationship (2.133). For our present purposes it is convenient to use the linear expression for ionic currents. Active transport of Na^+ and K^+ is performed primarily by the Na^+–K^+ ATPase (see Section 2.5.3).

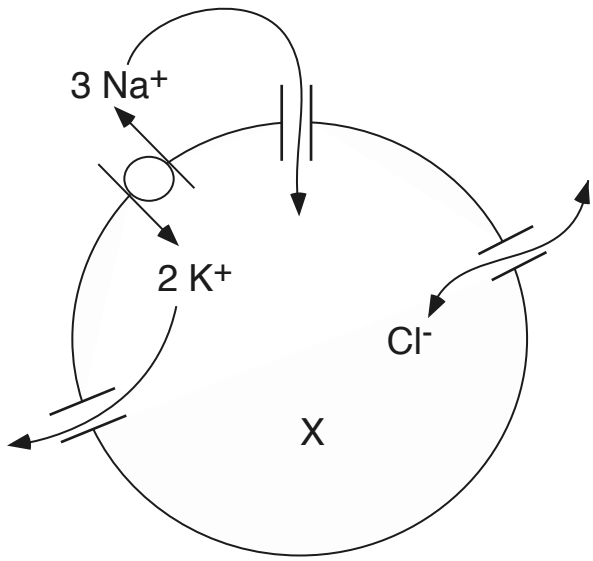


Figure 2.14 Schematic diagram of the pump–leak model.

Combining the expressions for active and passive ion transport, we find that the Na^+, K^+, and Cl^- currents are given by

$$I_{Na} = g_{Na}\left[V - \frac{RT}{F}\ln\left(\frac{[Na^+]_e}{[Na^+]_i}\right)\right] + 3pq, \tag{2.146}$$

$$I_K = g_K\left[V - \frac{RT}{F}\ln\left(\frac{[K^+]_e}{[K^+]_i}\right)\right] - 2pq, \tag{2.147}$$

$$I_{Cl} = g_{Cl}\left[V + \frac{RT}{F}\ln\left(\frac{[Cl^-]_e}{[Cl^-]_i}\right)\right], \tag{2.148}$$

where p is the rate at which the ion exchange pump works and q is the charge of a single ion.

We can express these current–voltage equations as differential equations by noting that an outward ionic current of ion A^{z+} affects the intracellular concentration of that ion through

$$I_A = -\frac{d}{dt}(zFw[A^{z+}]), \tag{2.149}$$

with w denoting the cell volume. (We use w rather than v to denote the cell volume to prevent confusion with V, the membrane potential.) Thus we have

$$-\frac{d}{dt}(Fw[Na^+]_i) = g_{Na}\left[V - \frac{RT}{F}\ln\left(\frac{[Na^+]_e}{[Na^+]_i}\right)\right] + 3pq, \tag{2.150}$$

$$-\frac{d}{dt}(Fw[K^+]_i) = g_K\left[V - \frac{RT}{F}\ln\left(\frac{[K^+]_e}{[K^+]_i}\right)\right] - 2pq, \tag{2.151}$$

$$\frac{d}{dt}(Fw[Cl^-]_i) = g_{Cl}\left[V + \frac{RT}{F}\ln\left(\frac{[Cl^-]_e}{[Cl^-]_i}\right)\right]. \tag{2.152}$$

Next, we let X denote the number of moles of large negatively charged molecules (with valence $z_x \leq -1$) that are trapped inside the cell. The flow of water across the membrane is driven by osmotic pressure, so that the change of cell volume is given by

$$r\frac{dw}{dt} = RT\left([Na^+]_i - [Na^+]_e + [K^+]_i - [K^+]_e + [Cl^-]_i - [Cl^-]_e + \frac{X}{w}\right). \tag{2.153}$$

Here we have assumed that the mechanical (hydrostatic) pressure difference across the membrane is zero, and we have also assumed that the elastic restoring force for the membrane is negligible.

Now to determine the membrane potential, we could use the electrical circuit model of the cell membrane, and write

$$C_m\frac{dV}{dt} + I_{Na} + I_K + I_{Cl} = 0. \tag{2.154}$$

However, the system of equations (2.150)–(2.154) has an infinite number of steady states. This can be seen from the fact that, at steady state, we must have $I_{Na} = I_K = I_{Cl} = 0$, from which it follows that dV/dt must also necessarily be zero. Since the

solution is thus determined by the choice of initial condition, it is better to use the integrated form of (2.154), i.e.,

$$C_m V = Q_i - Q_e, \tag{2.155}$$

where Q_i and Q_e are the total positive charge in the intracellular and extracellular spaces, respectively. (Note that here C_m refers to the total cell capacitance, assumed to be independent of cell volume. In other chapters of this book, C_m refers to membrane capacitance per unit area.) Since total charge is the difference between total number of positive and negative charges, we take

$$Q_i = qw([\mathrm{Na}^+]_i + [\mathrm{K}^+]_i - [\mathrm{Cl}^-]_i) + z_x q X, \tag{2.156}$$

$$Q_e = qw_e([\mathrm{Na}^+]_e + [\mathrm{K}^+]_e - [\mathrm{Cl}^-]_e). \tag{2.157}$$

This expression would be correct if the concentration of an ion was defined as the total number of ions in a region divided by the volume of that region, or if the distribution of ions was spatially homogeneous. But such is not the case here. This is because excess charge always accumulates in a thin region at the boundary of the domain (the Debye layer). However, this excess charge is quite small. To see this, consider a cylindrical piece of squid axon of typical radius 500 μm. With a capacitance of 1 μF/cm^2 and a typical membrane potential of 100 mV, the total charge is $Q = C_m V = \pi \times 10^{-8}$ C/cm. In comparison, the charge associated with intracellular K^+ ions at 400 mM is about 0.1 π C/cm, showing a relative charge deflection of about 10^{-7}.

Thus, since Q_i and Q_e are so small compared to the charges of each ion, it is a excellent approximation to assume that both the extracellular and intracellular media are electroneutral. Thus, Na^+, K^+, and Cl^- are assumed to be in electrical balance in the extracellular space. In view of the numbers for squid axon, this assumption is not quite correct, indicating that there must be other ions around to maintain electrical balance. In the intracellular region, Na^+, K^+, and Cl^- are not even close to being in electrical balance, but here, electroneutrality is maintained by the large negatively charged proteins trapped within the cell's interior. It is, of course, precisely the presence of these proteins in the interior of the cell that makes this whole exercise necessary. If a cell were not full of proteins (negatively charged or otherwise), it could avoid excessive osmotic pressures simply by allowing ions to cross the plasma membrane freely.

The assumption of electroneutrality gives the two equations

$$[\mathrm{Na}^+]_e + [\mathrm{K}^+]_e - [\mathrm{Cl}^-]_e = 0, \tag{2.158}$$

$$[\mathrm{Na}^+]_i + [\mathrm{K}^+]_i - [\mathrm{Cl}^-]_i + z_x \frac{X}{w} = 0. \tag{2.159}$$

It is convenient to assume that the cell is in an infinite bath, so that ionic currents do not change the external concentrations, which are thus assumed to be fixed and known, and to satisfy (2.158).

The differential equations (2.150), (2.151), (2.152), and (2.153) together with the requirement of intracellular electroneutrality (2.159) describe the changes of cell volume and membrane potential as functions of time. Note that we have 4 differential

equations and one algebraic equation for the five unknowns ($[Na^+]_i$, $[K^+]_i$, $[Cl^-]_i$, w and V).

Even though we formulated this model as a system of differential equations, we are interested, for the moment, only in their steady-state solution. Time-dependent currents and potentials become important in Chapter 5 for the discussion of excitability.

To understand these equations, we introduce the nondimensional variables $v = \frac{FV}{RT}, P = \frac{pFq}{RTg_{Na}}, \mu = \frac{w}{X}[Cl^-]_e$ and set $y = e^{-v}$. Then, the equation of intracellular electroneutrality becomes

$$\alpha y - \frac{1}{y} + \frac{z_x}{\mu} = 0, \tag{2.160}$$

and the equation of osmotic pressure balance becomes

$$\alpha y + \frac{1}{y} + \frac{1}{\mu} - 2 = 0, \tag{2.161}$$

where $\alpha = \frac{[Na^+]_e e^{-3P} + [K^+]_e e^{2P\gamma}}{[Na^+]_e + [K^+]_e}$ and $\gamma = g_{Na}/g_K$. In terms of these nondimensional variables, the ion concentrations are

$$\frac{[Na^+]_i}{[Na^+]_e} = e^{-3P} y, \tag{2.162}$$

$$\frac{[K^+]_i}{[K^+]_e} = e^{2P\gamma} y, \tag{2.163}$$

$$\frac{[Cl^-]_i}{[Cl^-]_e} = \frac{1}{y}. \tag{2.164}$$

Solving (2.160) for its unique positive root, we obtain

$$y = \frac{-z_x + \sqrt{z_x^2 + 4\alpha\mu^2}}{2\alpha\mu}, \tag{2.165}$$

and when we substitute for y back into (2.161), we find the quadratic equation for μ:

$$4(1-\alpha)\mu^2 - 4\mu + 1 - z_x^2 = 0. \tag{2.166}$$

For $z_x \leq -1$, this quadratic equation has one positive root if and only if $\alpha < 1$. Expressed in terms of concentrations, the condition $\alpha < 1$ is

$$\rho(P) = \frac{[Na^+]_e e^{-3P} + [K^+]_e e^{2P\gamma}}{[Na^+]_e + [K^+]_e} < 1. \tag{2.167}$$

One can easily see that $\rho(0) = 1$ and that for large P, $\rho(P)$ is exponentially large and positive. Thus, the only hope for $\rho(P)$ to be less than one is if $\rho'(0) < 0$. This occurs if and only if

$$\frac{3[Na^+]_e}{g_{Na}} > \frac{2[K^+]_e}{g_K}, \tag{2.168}$$

in which case there is a range of values of P for which a finite, positive cell volume is possible and for which there is a corresponding nontrivial membrane potential.

To decide if this condition is ever satisfied we must determine "typical" values for g_{Na} and g_K. This is difficult to do, because, as is described in Chapter 5, excitability of nerve tissue depends strongly on the fact that conductances are voltage-dependent and can vary rapidly over a large range of values. However, at rest, in squid axon, reasonable values are $g_K = 0.367$ mS/cm^2 and $g_{Na} = 0.01$ mS/cm^2. For these values, and at the extracellular concentrations of 437 and 20 mM for Na^+ and K^+, respectively, the condition (2.168) is readily met.

One important property of the model is that the resting value of V is equal to the Nernst potential of Cl^-, as can be seen from (2.152) or (2.164). Thus, the membrane potential is set by the activity of the Na^+–K^+ ATPase, and the intracellular Cl^- concentration is set by the membrane potential.

In Figs. 2.15 and 2.16 the nondimensional volume μ and the potential V (assuming $RT/F = 25.8$ mV) are plotted as functions of the pump rate P. In addition, in Fig. 2.16 are shown the Na^+ and K^+ equilibrium potentials. For these plots, γ was chosen to be 0.11, and $z_x = -1$. Then, at $P = 1.6$, the Na^+ and K^+ equilibrium potentials and the membrane potentials are close to their observed values for squid axon, of 56, −77 and −68 mV, respectively.

From these plots we can see the effect of changing pump rate on cell volume and membrane potential. At zero pump rate, the membrane potential is zero and the cell volume is infinite (dead cells swell). As the pump rate increases from zero, the cell volume and membrane potential rapidly decrease to their minimal values and then gradually increase until at some upper limit for pump rate, the volume and potential become infinite. The K^+ equilibrium potential is seen to decrease rapidly as a function of pump rate until it reaches a plateau at a minimum value. The Na^+ equilibrium potential increases monotonically.

Physically realistic values of the membrane potential are achieved fairly close to the local minimum. Clearly, there is little advantage for a higher pump rate, and since

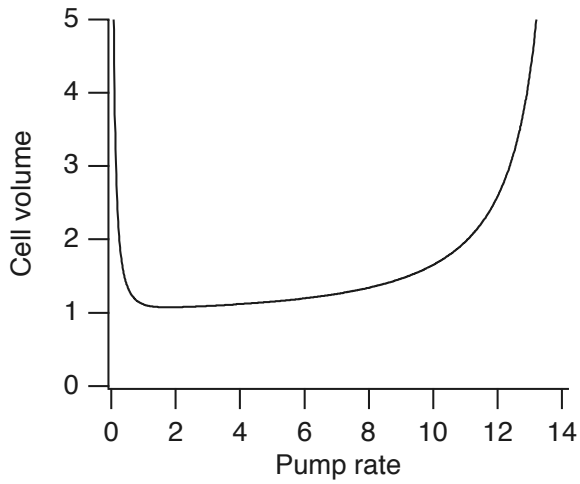


Figure 2.15 Cell volume as a function of the pump rate.

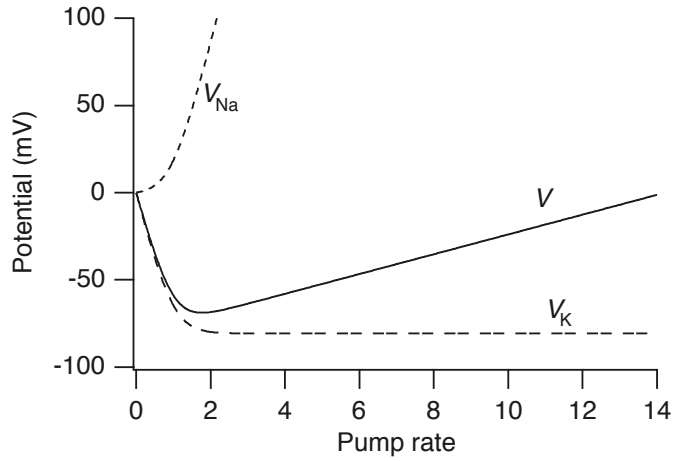


Figure 2.16 Membrane potential, Na^+ equilibrium potential, and K^+ equilibrium potential as functions of the pump rate.

Table 2.4 Resting potentials in some typical excitable cells.

Cell Type	Resting Potential (mV)
Neuron	−70
Skeletal muscle (mammalian)	−80
Skeletal muscle (frog)	−90
Cardiac muscle (atrial and ventricular)	−80
Cardiac Purkinje fiber	−90
Atrioventricular nodal cell	−65
Sinoatrial nodal cell	−55
Smooth muscle cell	−55

the pump rate is proportional to energy expenditure, it would seem that the pump rate is chosen approximately to minimize cell volume, membrane potential, and energy expenditure. However, no mechanism for the regulation of energy expenditure is suggested.

Generalizations

While the above model of volume control and membrane potential is useful and gives some insight into the control mechanisms, as with most models there are important features that have been ignored but that might lead to substantially different behavior.

There are (at least) two significant simplifications in the model presented here. First, the conductances g_{Na} and g_K are treated as constants. In Chapter 5 we show that the ability of cells to generate an electrical signal results from voltage and time dependence of the conductances. In fact, the discovery that ion channels have differing

properties of voltage sensitivity was of fundamental importance to the understanding of neurons. The second simplification relates to the operation of the ion exchange pump. Figure 2.16 suggests that the minimal membrane potential is achieved at a particular pump rate and suggests the need for a tight control of pump rate that maintains the potential near this minimum. If indeed, such a tight control is required, it is natural to ask what that control mechanism might be. There is also the difficulty that in this simple model there is nothing preventing the complete depletion of Na^+ ions.

A different model of the pump activity might be beneficial. Recall from (2.98) that with each cycle of the ion exchange pump, three intracellular Na^+ ions are exchanged for two extracellular K^+ ions. Our previous analysis of the Na^+–K^+ ATPase (see (2.100)) suggests that at low internal Na^+ concentrations, the pump rate can be represented in nondimensional variables as

$$P = \rho u^3, \tag{2.169}$$

where $u = [Na^+]_i/[Na^+]_e$. This representation is appropriate at high pump rates, where effects of saturation are of no concern. Notice that P is proportional to the rate of ATP hydrolysis, and hence to energy consumption. Thus, as u decreases, so also does the rate of energy consumption. With this change, the equation for the Na^+ concentration becomes

$$u \exp(3\rho u^3) = y, \tag{2.170}$$

and this must be solved together with the quadratic polynomials (2.160) and (2.161), with (2.170) replacing (2.162).

In Fig. 2.17 are shown the membrane potential, and the Na^+ and K^+ equilibrium potentials, plotted as functions of the nondimensional reaction rate ρ. Here we

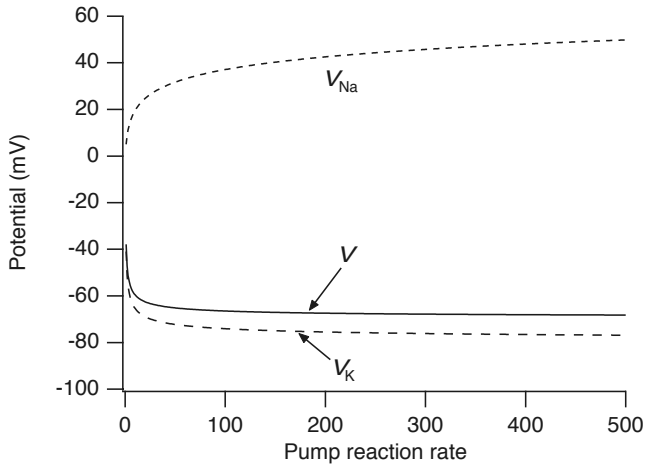


Figure 2.17 Membrane potential, Na^+ equilibrium potential, and K^+ equilibrium potential as functions of the pump rate, for the modified pump rate (2.169).

see something qualitatively different from what is depicted in Fig. 2.16. There the membrane potential had a noticeable local minimum and was sensitive to changes in pump rate. In this modified model, the membrane potential is insensitive to changes in the pump rate. The reason for this difference is clear. Since the effectiveness of the pump depends on the internal Na^+ concentration, increasing the speed of the pumping rate has little effect when the internal Na^+ is depleted, because of the diminished number of Na^+ ions available to be pumped.

While the pump rate is certainly ATP dependent, there are a number of drugs and hormones that are known to affect the pump rate. Catecholamines rapidly increase the activity of the pump in skeletal muscle, thereby preserving proper K^+ during strenuous exercise. Within minutes, insulin stimulates pump activity in the liver, muscle, and fat tissues, whereas over a period of hours, aldosterone and corticosterones increase activity in the intestine.

On the other hand, digitalis (clinically known as digoxin) is known to suppress pump activity. Digitalis is an important drug used in the treatment of congestive heart failure and during the 1980s was the fourth most widely prescribed drug in the United States. At therapeutic concentrations, digitalis inhibits a moderate fraction (say, 30–40%) of the Na^+–K^+ ATPase, by binding with the Na^+ binding site on the extracellular side. This causes an increase in internal Na^+, which has an inhibitory effect on the Na^+–Ca^{2+} exchanger, slowing the rate by which Ca^{2+} exits the cells. Increased levels of Ca^{2+} result in increased myocardial contractility, a positive and useful effect. However, it is also clear that at higher levels, the effect of digitalis is toxic.

2.8.2 Volume Regulation and Ionic Transport

Many cells have a more difficult problem to solve, that of maintaining their cell volume in widely varying conditions, while transporting large quantities of ions through the cell. Here we present a simplified model of transport and volume regulation in a Na^+-transporting epithelial cell.

As are virtually all models of transporting epithelia, the model is based on that of Koefoed-Johnsen and Ussing (1958), the so-called KJU model. In the KJU model, an epithelial cell is modeled as a single cell layer separating a mucosal solution from the serosal solution (Fig. 2.18). (The mucosal side of an epithelial cell is that side on which mucus is secreted and from which various chemicals are withdrawn, for example, from the stomach. The serosal side is the side of the epithelial cell facing the interstitium, wherein lie capillaries, etc.) Na^+ transport is achieved by separating the Na^+ pumping machinery from the channels that allow Na^+ entry into the cell. Thus, the mucosal membrane contains Na^+ channels that allow Na^+ to diffuse down its concentration gradient into the cell, while the serosal membrane contains the Na^+–K^+ ATPases which remove Na^+ from the cell. The overall result is the transport of Na^+ from the mucosal side of the cell to the serosal side. The important question is whether the cell can maintain a steady volume under widely varying concentrations of Na^+ on the mucosal side.

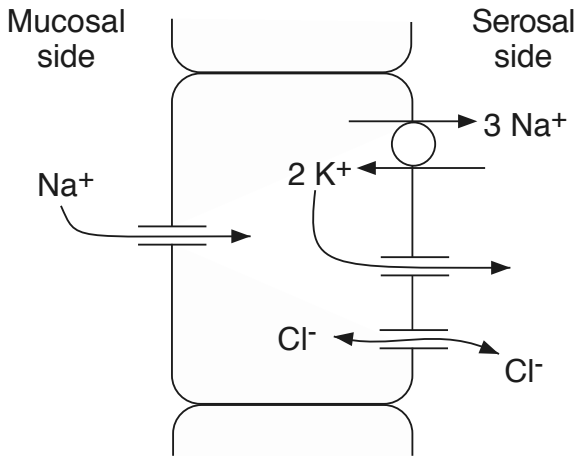


Figure 2.18 Schematic diagram of the model of a Na^+-transporting epithelial cell, based on the model of Koefoed-Johnsen and Ussing (1958).

We begin by letting N, K, and C denote Na^+, K^+, and Cl^- concentrations respectively, and letting subscripts m, i, and s denote mucosal, intracellular and serosal concentrations. Thus, for example, N_i is the intracellular Na^+ concentration, and N_m is the mucosal Na^+ concentration. We now write the conservation equations for Na^+, K^+, and Cl^- at steady state. The conservation equations are the same as those of the pump–leak model with some minor exceptions. First, instead of the linear I–V curve used in the pump–leak model, we use the GHK formulation to represent the ionic currents. This makes little qualitative change to the results but is more convenient because it simplifies the analysis that follows. Second, we assume that the rate of the Na^+–K^+ ATPase is proportional to the intracellular Na^+ concentration, N_i, rather than N_i^3, as was assumed in the generalized version of the pump–leak model. Thus,

$$P_{\mathrm{Na}} v \frac{N_i - N_m e^{-v}}{1 - e^{-v}} + 3qpN_i = 0, \tag{2.171}$$

$$P_{\mathrm{K}} v \frac{K_i - K_s e^{-v}}{1 - e^{-v}} - 2qpN_i = 0, \tag{2.172}$$

$$P_{\mathrm{Cl}} v \frac{C_i - C_s e^{v}}{1 - e^{v}} = 0. \tag{2.173}$$

Note that the voltage, v, is nondimensional, having been scaled by $\frac{F}{RT}$, and that the rate of the Na^+–K^+ ATPase is pN_i. Also note that the inward Na^+ current is assumed to enter from the mucosal side, and thus N_m appears in the GHK current expression, but that no other ions enter from the mucosa. Here the membrane potential is assumed to be the same across the lumenal membrane and across the basal membrane. This is not quite correct, as the potential across the lumenal membrane is typically −67 mV while across the basal membrane it is about −70 mV.

There are two further equations to describe the electroneutrality of the intracellular space and the osmotic balance. In steady state, these are, respectively,

$$w(N_i + K_i - C_i) + z_x X = 0, \tag{2.174}$$

$$N_i + K_i + C_i + \frac{X}{w} = N_s + K_s + C_s, \tag{2.175}$$

where X is the number of moles of protein, each with a charge of $z_x \leq -1$, that are trapped inside the cell, and w is the cell volume. Finally, the serosal solution is assumed to be electrically neutral, and so in specifying N_s, K_s, and C_s we must ensure that

$$N_s + K_s = C_s. \tag{2.176}$$

Since the mucosal and serosal concentrations are assumed to be known, we now have a system of 5 equations to solve for the 5 unknowns, N_i, K_i, C_i, v, and $\mu = w/X$. First, notice that (2.171), (2.172), and (2.173) can be solved for N_i, K_i, and C_i, respectively, to get

$$N_i(v) = \frac{vN_m e^{-v}}{v + 3\rho_n(1 - e^{-v})}, \tag{2.177}$$

$$K_i(v) = 2\rho_k N_i(v)\frac{1 - e^{-v}}{v} + K_s e^{-v}, \tag{2.178}$$

$$C_i(v) = C_s e^{v}, \tag{2.179}$$

where $\rho_n = pq/P_{\mathrm{Na}}$ and $\rho_k = pq/P_{\mathrm{K}}$.

Next, eliminating $N_i + K_i$ between (2.174) and (2.175), we find that

$$2\mu(C_i - C_s) = z_x - 1. \tag{2.180}$$

We now use (2.179) to find that

$$z_x - 1 = 2\mu C_s(e^{v} - 1), \tag{2.181}$$

and thus, using (2.181) to eliminate μ from (2.174), we get

$$N_i(v) + K_i(v) = \frac{C_s}{1 - z_x}[-2z_x + e^{v}(1 + z_x)] \equiv \phi(v). \tag{2.182}$$

Since $z_x - 1 < 0$, it must be (from (2.181)) that $v < 0$, and as $v \to 0$, the cell volume $w = \mu X$ becomes infinite. Thus, we wish to find a negative solution of (2.182), with $N_i(v)$ and $K_i(v)$ specified by (2.177) and (2.178).

It is instructive to consider when solutions for v (with $v < 0$) exist. First, notice that $\phi(0) = C_s$. Further, since $z_x \leq -1$, ϕ is a decreasing function of v, bounded above, with decreasing slope (i.e., concave down), as sketched in Fig. 2.19. Next, from (2.177) and (2.178) we determine that $N_i(v) + K_i(v)$ is a decreasing function of v that approaches ∞ as $v \to -\infty$ and approaches zero as $v \to \infty$. It follows that a negative solution for v exists if $N_i(0) + K_i(0) < C_s$, i.e., if

$$\frac{N_m}{1 + 3\rho_n} + \frac{2\rho_k N_m}{1 + 3\rho_n} + K_s < C_s. \tag{2.183}$$

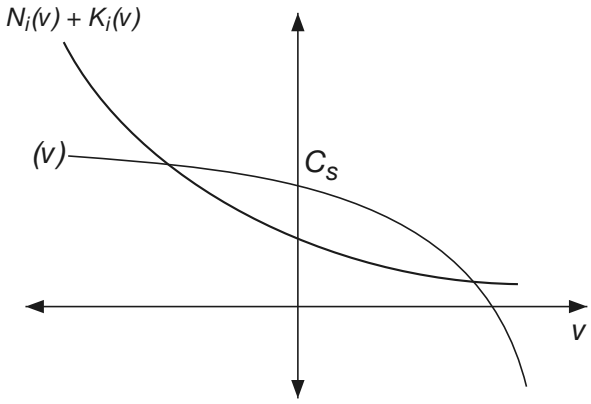


Figure 2.19 Sketch (not to scale) of the function $\phi(v)$, defined as the right-hand side of (2.182), and of $N_i(v) + K_i(v)$, where N_i and K_i are defined by (2.177) and (2.178). $\phi(v)$ is sketched for $z_x < -1$.

Since $K_s + N_s = C_s$, this becomes

$$\frac{N_m}{N_s} < \frac{1 + 3\rho_n}{1 + 2\rho_k}. \tag{2.184}$$

This condition is sufficient for the existence of a solution, but not necessary. That is, if this condition is satisfied, we are assured that a solution exists, but if this condition fails to hold, it is not certain that a solution fails to exist. The problem, of course, is that negative solutions are not necessarily unique, nor is it guaranteed that increasing N_m through $N_s \frac{1+3\rho_n}{1+2\rho_k}$ causes a negative solution to disappear. It is apparent from (2.177) and (2.178) that $N_i(v)$ and $K_i(v)$ are monotone increasing functions of the parameter N_m, so that no negative solutions exist for N_m sufficiently large. However, for $N_m = N_s \frac{1+3\rho_n}{1+2\rho_k}$ to be the value at which the cell bursts by *increasing* N_m, it must also be true that

$$N_i'(0) + K_i'(0) < \phi'(0), \tag{2.185}$$

or that

$$4(1 + 3\rho_n)C_s + N_s(1 - z_x)\frac{3\rho_n - 2\rho_k}{1 + 2\rho_k} > 0. \tag{2.186}$$

For the remainder of this discussion we assume that this condition holds, so that the failure of (2.184) also implies that the cell bursts.

According to (2.184), a transporting epithelial cell can maintain its cell volume, provided the ratio of mucosal to serosal concentrations is not too large. When N_m/N_s becomes too large, μ becomes unbounded, and the cell bursts. Typical solutions for the cell volume and membrane potential, as functions of the mucosal Na^+ concentration, are shown in Fig. 2.20.

Obviously, this state of affairs is unsatisfactory. In fact, some epithelial cells, such as those in the loop of Henle in the nephron (Chapter 17), must work in environments with extremely high mucosal Na^+ concentrations. To do so, these Na^+-transporting epithelial cells have mechanisms to allow operation over a much wider range of mucosal Na^+ concentrations than suggested by this simple model.

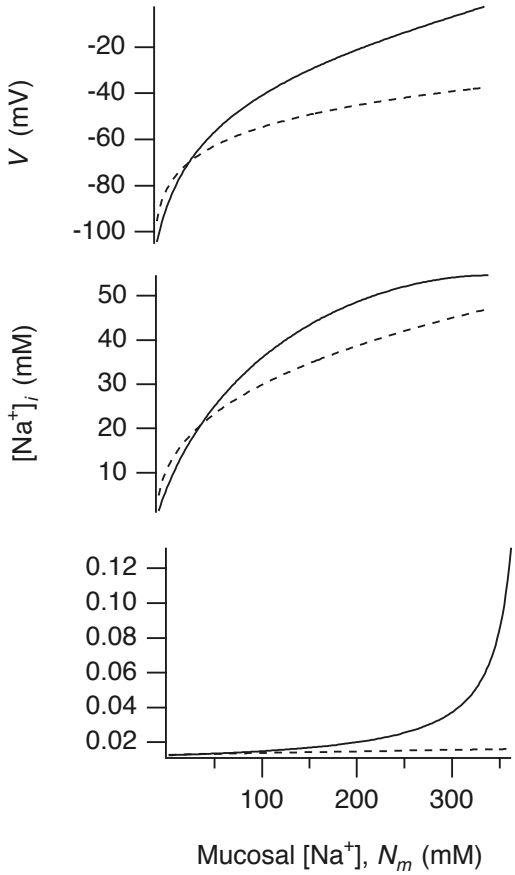


Figure 2.20 Numerical solutions of the model of epithelial cell volume regulation and Na^+ transport. The membrane potential, V, the scaled cell volume, μ, and the intracellular Na^+ concentration, $[Na^+]_i$, are plotted as functions of the mucosal Na^+ concentration. The solid lines are the solutions of the simpler version of the model, where P_{Na} and P_K are assumed to be constant. The dashed lines are the solutions of the model when P_{Na} is assumed to be a decreasing function of N_i, and P_K is assumed to be an increasing function of w, as described in the text. Parameter values are $K_s = 2.5$, $N_s = 120$, $C_s = 122.5$, $P = 2$, $\gamma = 0.3$, $z_x = -2$. All concentrations are in mM.

From (2.184) we can suggest some mechanisms by which a cell might avoid bursting at high mucosal concentrations. For example, the possibility of bursting is decreased if ρ_n is increased or if ρ_k is decreased. The reasons for this are apparent from (2.177) and (2.178), since $N_i(v) + K_i(v)$ is a decreasing function of ρ_n and an increasing function of ρ_k. From a physical perspective, increasing N_m causes an increase in N_i, which increases the osmotic pressure, inducing swelling. Decreasing the conductance of Na^+ ions from the mucosal side helps to control this swelling. Similarly, increasing the conductance of K^+ ions allows more K^+ ions to flow out of the cell, thereby decreasing the osmotic pressure from K^+ ions and counteracting the tendency to swell.

It has been conjectured for some time that epithelial cells use both of these mechanisms to control their volume (Schultz, 1981; Dawson and Richards, 1990; Beck et al., 1994). There is evidence that as N_i increases, epithelial cells decrease the Na^+ conductance on the mucosal side of the cell, thus restricting Na^+ entry. There is also evidence that as the cell swells, the K^+ conductance is increased, possibly by means of stretch-activated K^+ channels (Ussing, 1982. This assumption was used in the modeling work of Strieter et al., 1990).

To investigate the effects of these mechanisms in our simple model, we replace P_{Na} by $P_{\mathrm{Na}}20/N_i$ (20 is a scale factor, so that when $N_i = 20$ mM, P_{Na} has the same value as in the original version of the model) and replace P_{K} by $P_{\mathrm{K}}w/w_0$, where w_0 is the volume of the cell when $N_m = 100$ mM. As before, we can solve for v and μ as functions of N_m, and the results are shown in Fig. 2.20. Clearly the incorporation of these mechanisms decreases the variation of cell volume and allows the cell to survive over a much wider range of mucosal Na^+ concentrations.

The model of control of ion conductance used here is extremely simplistic, as for example, there is no parametric control of sensitivity, and the model is heuristic, not mechanistic. More realistic and mechanistic models have been constructed and analyzed in detail (Lew et al., 1979; Civan and Bookman, 1982; Strieter et al., 1990; Weinstein, 1992, 1994, 1996; Tang and Stephenson, 1996).

2.9 Appendix: Stochastic Processes

Although all of the models that have been presented so far in this text have been deterministic, the reality is that biological processes are fundamentally noisy. Furthermore, many of the assumptions underlying deterministic models are questionable, largely because of significant stochastic effects.

The purpose of this appendix is to outline some of the basic ideas of stochastic processes that play an important role in mathematical modeling of biological phenomena. Furthermore, there are a number of sections in the remainder of this book where stochastic models are crucial, and we hope that this appendix provides the necessary background for these. Of course, this is not a detailed treatment of these topics; for that one needs to consult a text on stochastic processes such as Gardiner (2004) or van Kampen (2007).

2.9.1 Markov Processes

A *Markov process* is any stochastic process that has no memory. More precisely, if the value of the state variable x is known at two times, say $t_1 < t_2$, to be x_1 and x_2, respectively, then

$$P(x, t \mid x_1, t_1, x_2, t_2) = P(x, t \mid x_2, t_2), \tag{2.187}$$

where $P(x \mid y)$ denotes the conditional probability of x given y. In words, the conditional probability for the value of the state variable depends only on the most recent condition and not on any previous conditions.

A simple example of a Markov process is radioactive decay. For example, an atom of carbon-14 may lose an electron and decay to nitrogen-14. This is a Markov process, because the probability of decay does not depend on how old the carbon-14 atom is. A newly formed carbon-14 atom has exactly the same probability of decay in the next second as a very old carbon-14 atom.

To model this process, we suppose that the state variable has two possible values, say C (for carbon) and N (for nitrogen), and we denote the probability that the molecule is in state S at time t by $P(\mathrm{S}, t)$. Now suppose that the probability of radioactive decay in a small interval of time dt is $\lambda\, dt$. (Note that λ need not be independent of time for this to be a Markov process.) Then,

$$P(\mathrm{N}, t + \Delta t) = P(\mathrm{N}, t) + P(\mathrm{C}, t)\lambda \Delta t. \tag{2.188}$$

For this problem we assume that $P(\mathrm{C}, t) + P(\mathrm{N}, t) = 1$. It follows that in the limit $\Delta t \to 0$,

$$\frac{dP(\mathrm{N}, t)}{dt} = \lambda(1 - P(\mathrm{N}, t)). \tag{2.189}$$

By similar arguments, many of the differential equations in this chapter describing concentrations or fractions of molecules in a given state can be reinterpreted as equations for the probability that a single molecule is in a particular state. For example, for a transporter molecule that has two states,

$$\mathrm{C}_i \underset{k}{\overset{k}{\rightleftarrows}} \mathrm{C}_e, \tag{2.190}$$

under the assumption that transitions between states do not depend on how long the molecule has been in a given state, the probability P of being in state C_i at time t, $P(\mathrm{C}_i, t)$, is given by

$$\frac{dP(\mathrm{C}_i, t)}{dt} = k(1 - P(\mathrm{C}_e, t)) - kP(\mathrm{C}_i, t). \tag{2.191}$$

Similarly, for the chemical reaction

$$\mathrm{A} \underset{k_-}{\overset{k_+}{\rightleftarrows}} \mathrm{B}, \tag{2.192}$$

the probability $P(\mathrm{A}, t)$ of a molecule being in state A at time t is determined by

$$\frac{dP(\mathrm{A}, t)}{dt} = k_- P(\mathrm{B}, t) - k_+ P(\mathrm{A}, t). \tag{2.193}$$

Of course, this is the same as the equation found using the law of mass action for the conversion of A to B, namely,

$$\frac{da}{dt} = k_- b - k_+ a. \tag{2.194}$$

Even though these equations are identical, their interpretations are quite different. For example, at steady state, $\frac{a}{b} = \frac{k_-}{k_+}$ implies that the ratio of the number of molecules in state A to the number of those in state B is $\frac{k_-}{k_+}$. However, this cannot be the probabilistic interpretation, since there is no fractional state. Instead, the probabilistic interpretation is that the ratio of the time a single molecule spends in state A to the time it spends in state B is $\frac{k_-}{k_+}$.

2.9.2 Discrete-State Markov Processes

Often, the state space of a Markov process is discrete; for example, the model of the Na^+–K^+ ATPase (Fig. 2.11) assumes that the ATPase can be in only a small number of different states. The simple chemical reaction in the previous section is also a model of this type, since the molecule can exist in only two states, A or B. Since time is a continuous variable, such models are called *discrete-space continuous-time Markov processes*. Such Markov processes play an enormously important role in modeling; indeed, a large fraction of the models of chemical reactions, exchangers, pumps, and ion channels we discuss in this book are discrete-space continuous-time Markov processes. As is described in Chapter 3, one of the most important applications of such models is the analysis of data from single ion channels.

To describe the stochastic behavior of a discrete-space continuous-time Markov process with n possible states (open, closed, inactivated, etc.), we introduce a discrete random variable $S(t) \in 1, 2, \ldots, n$ defined so that $S(t) = i$ if the model is in state i at time t. Further, we suppose that the probability that the model changes from state i to state j in the time interval $(t, t + dt)$ is $k_{ij}dt$. In more condensed notation,

$$P(S(t + dt) = j \mid S(t) = i) = k_{ij}dt. \tag{2.195}$$

Note that (2.195) is valid only in the limit of small dt, since for sufficiently large dt and k_{ij} nonzero, this probability exceeds 1. Also, the probability that the model does not change state in the time interval $(t, t + dt)$ is given by

$$P(S(t + dt) = i \mid S(t) = i) = 1 - K_i dt, \tag{2.196}$$

where $K_i = \sum_{j \neq i} k_{ij}$.

Now we let $\Phi_j(t) = P(S(t) = j)$, i.e., $\Phi_j(t)$ is the probability that the model is in state j at time t. In words, the probability that the model is in state j at time $t + dt$ is the probability that it was in state j at time t and did not leave state j between times t and $t + dt$ plus the probability that at time t it was in another state and switched into state j in the time interval t to $t + dt$. In mathematical language,

$$\begin{aligned}\Phi_j(t + dt) &= \Phi_j(t)P(S(t + dt) = i \mid S(t) = i) + \sum_{l \neq j} P(S(t + dt) = j \mid S(t) = l)\Phi_l(t) \\ &= \Phi_j(t)\left(1 - K_j dt\right) + \sum_{l \neq j} k_{lj}\Phi_l(t)dt,\end{aligned} \tag{2.197}$$

and thus, taking the limit $dt \to 0$,

$$\frac{d\Phi_j}{dt} = -K_j\Phi_j(t) + \sum_{l \neq j} k_{lj}\Phi_l(t). \tag{2.198}$$

In vector notation, with Φ the vector having elements Φ_j, we have

$$\frac{d\Phi}{dt} = A^T\Phi(t), \tag{2.199}$$

where A is the matrix

$$A = \begin{pmatrix} -K_1 & k_{12} & \cdots & k_{1n} \\ k_{21} & -K_2 & \cdots & k_{2n} \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}. \tag{2.200}$$

Of course, this is exactly the system of differential equations we would have written down for a system of first-order chemical reactions with reaction rates between species k_{ij}. Here, however, the variables Φ_j are probabilities and so must sum to one. In the terminology of stochastic processes, (2.199) is called the *master equation* for this Markov process.

The Waiting Time

One important question is how long a Markov model stays in a particular state before switching. This is called the waiting-time problem. To solve this we let T_i be the random time at which the model switches from state i to some other state, and we let $P_i(t) = P(T_i < t)$, that is, $P_i(t)$ is the probability that by time t the switch has occurred. In words, the probability that the switch has occurred by time $t + dt$ is the probability that it has occurred by time t plus the probability that the switch has not occurred by time t and occurs in the time interval between t and $t + dt$. In mathematical language,

$$P_i(t + dt) = P_i(t) + (1 - P_i)K_i dt. \tag{2.201}$$

Taking the limit $dt \to 0$ we obtain the differential equation for the waiting-time probability

$$\frac{dP_i}{dt} = K_i(1 - P_i), \tag{2.202}$$

which, since $P_i(0) = 0$ (assuming that the switch has not occurred at time $t = 0$), and if K_i is independent of time, yields

$$P_i(t) = 1 - \exp(-K_i t). \tag{2.203}$$

The function $P_i(t)$ is a cumulative probability distribution function, with probability density function

$$p_i(t) = \frac{dP_i}{dt} = K_i \exp(-K_i t). \tag{2.204}$$

The probability that the switch occurs between two specified times, t_1 and t_2, is

$$P(t_1 < T_i < t_2) = \int_{t_1}^{t_2} p_i(s)\, ds = P_i(t_2) - P_i(t_1), \tag{2.205}$$

and the expected switching time is

$$E(T_i) = \int_0^\infty t p_i(t)\, dt. \tag{2.206}$$

If K_i is time-independent,

$$E(T_i) = \frac{1}{K_i}. \tag{2.207}$$

The Transition Time

The waiting-time problem can be generalized to calculate the probability density function, ϕ_{ij}, for the time it takes to move from state i to state j.

To do this we make state j an absorbing state and then solve the master equations starting in state i. That is, we set $k_{jl} = 0$ for every l, and then solve for $\Phi_j(t)$, with the initial condition $\Phi_i(0) = 1$, $\Phi_l(0) = 0$ for $l \neq i$. Then, $\Phi_j(t)$ is the probability that the process is in state j at time t, given that it started in state i at time 0. Hence, $\Phi_j(t)$ is the cumulative probability that the transition to state j occurred at some time previous to t. Of course, this relies on the assumption that once state j is reached, it cannot be left. The probability density function for the transition time is the derivative of the cumulative probability. Hence,

$$\phi_{ij}(t) = \frac{d\Phi_j}{dt} = \sum_{l \neq j} k_{lj} \Phi_l. \tag{2.208}$$

Note that

$$\int_0^\infty \phi_{ij}\, dt = \Phi_j\big|_0^\infty = 1, \tag{2.209}$$

and thus ϕ_{ij} is indeed a probability density as claimed.

2.9.3 Numerical Simulation of Discrete Markov Processes

It is becoming increasingly important and valuable to do numerical simulations of discrete stochastic processes. The definition (2.195) provides a natural numerical algorithm for such a simulation. For a fixed small time step of size dt, divide the unit interval into $n-1$ regions of length $k_{ij}dt$ and one remaining region of length $1 - K_i dt$. Then, at each time step, pick a random number that is uniformly distributed on the unit interval and determine the next state by the subinterval in which the random number falls.

While this method of numerical simulation is simple and direct, it is not particularly efficient. Furthermore, it converges (i.e., gives the correct statistics) only in the limit that $dt \to 0$.

A method that is much more efficient is known as Gillespie's method (Gillespie, 1977). The idea of this method is to calculate the sequence of random switching times. For example, suppose that at time $t = 0$, the state variable is $S(0) = i$. We know from (2.204) that the probability that the first transition out of state i occurs in the interval $(t, t+dt)$ is $\int_t^{t+dt} K_i \exp(-K_i s)\, ds$. Hence, we can calculate the time to the next transition by selecting a random number from the exponential waiting-time distribution, $p_i(t)$. We do this by selecting a random number x uniformly distributed on the unit interval,

and transforming that number via some transformation $t = f(x)$ to get the time at which the transition occurs. The transformation we use should preserve probabilities,

$$p_i(t)dt = q(x)dx, \tag{2.210}$$

where $q(x) = 1$ is the uniform distribution. Integrating gives $x = \int_0^t p_i(s)\,ds = 1 - \exp(-K_i y) = P_i(t)$, and solving for t, we get

$$t = -\frac{1}{K_i}\ln(1-x). \tag{2.211}$$

However, since x is uniformly distributed on the unit interval, it is equivalent to replace $1 - x$ by x to get

$$t = -\frac{1}{K_i}\ln(x). \tag{2.212}$$

Therefore, to calculate the next switching time, pick a random number that is uniformly distributed on the unit interval, say ξ, and then pick the time interval, T, to the next switch to be

$$T = -\frac{1}{K_i}\ln(\xi). \tag{2.213}$$

To determine the state into which to switch, divide the unit interval into segments of length $\frac{k_{ij}}{K_i}$, and select the next interval to be the subinterval in which another uniformly distributed random number η resides. The reasoning for this is that if a switch is to occur, then the probability that the switch is into state j is $\frac{k_{ij}}{K_i}$.

There are numerous advantages to Gillespie's method. First, it is maximally efficient and it is exact. That is, since there is no time discretization step dt, accuracy does not require taking a limit $dt \to 0$. However, other nice features of the method are that it is easier to collect statistics such as closed time and open time distributions, because these quantities are directly, not indirectly, calculated.

A word of caution, however, is that while this method works well for time-independent processes (which we assumed in this discussion), for time-dependent processes, or processes in which the transition rates depend on other time-varying variables, determination of the next transition time requires a more sophisticated calculation (Alfonsi et al., 2005).

Further difficulties with stochastic simulations occur when there are reactions with vastly different time scales. Then, it is often the case that the rapid reactions achieve quasi-equilibrium, but most of the computational time for a stochastic simulation is taken in calculating the many fast transitions of the fast reactions. It is beyond the scope of this text to describe what to do in these situations.

Gillespie's method is the method of choice in many situations. It has been implemented by Adalsteinsson et al. (2004) in a useful software package that is readily available.

2.9.4 Diffusion

The diffusion equation (2.7) was derived to describe the evolution of a chemical concentration, under the assumption that the concentration is a continuous variable, even though the number of molecules involved is necessarily an integer. Einstein recognized that the solution of the diffusion equation could also be interpreted as the probability distribution function for the location of a single particle undergoing some kind of a random walk. That is, if $p(x,t)$ is the solution of the diffusion equation

$$\frac{\partial p}{\partial t} = D\nabla^2 p, \tag{2.214}$$

then $\int_\Omega p(x,t)\,dx$ could be identified as the probability that a particle is in the region Ω at time t. More specifically, if $p(x,t \mid x_0,t_0)$ is the probability distribution function for the particle to be at position x at time t, given that it was at position x_0 at time t_0, then

$$p(x,t_0 \mid x_0,t_0) = \delta(x-x_0), \tag{2.215}$$

and solving the diffusion equation (in one spatial dimension) gives

$$p(x,t \mid x_0,t_0) = \frac{1}{2\sqrt{\pi D(t-t_0)}} \exp\left(-\frac{(x-x_0)^2}{4D(t-t_0)}\right), \tag{2.216}$$

provided $t > t_0$. It follows immediately that the mean and variance of this distribution are

$$\langle x \rangle = \int_{-\infty}^{\infty} x p(x,t \mid 0,0)\,dx = 0 \tag{2.217}$$

and

$$\langle x^2 \rangle = \int_{-\infty}^{\infty} x^2 p(x,t \mid 0,0)\,dx = 2Dt. \tag{2.218}$$

The conditional probability $p(x,t \mid x_0,t_0)$ is the Green's function of the diffusion equation on an infinite domain (where we require, for physical reasons, that the solution and all its derivatives vanish at infinity). As a side issue (but a particularly interesting one), note that, from the transitive nature of Green's functions, it follows that

$$p(x_1,t_1 \mid x_3,t_3) = \int_{x_2} p(x_1,t_1 \mid x_2,t_2) p(x_2,t_2 \mid x_3,t_3)\,dx_2, \quad t_3 < t_2 < t_1, \tag{2.219}$$

which is known as the Chapman–Kolmogorov equation. From the point of view of conditional probabilities, the Chapman–Kolmogorov equation makes intuitive sense; for Markov processes, the probability of x_1 given x_3 is the sum of the probabilities of each path by which one can get from x_3 to x_1.

Now suppose we let $X(t)$ represent the position as a function of time of a sample path. We can readily calculate that $X(t)$ is continuous, since for any $\epsilon > 0$, the

probability of escaping from a region of size ϵ in time Δt is

$$\int_{|x-z|>\epsilon} p(x, t+\Delta t \mid z, t)\, dx = 2\int_{\epsilon}^{\infty} \frac{1}{2\sqrt{\pi D \Delta t}} \exp\left(-\frac{x^2}{4D\Delta t}\right) dx$$
$$= \int_{\frac{\epsilon}{2\sqrt{D\Delta t}}}^{\infty} \exp\left(-x^2\right) dx,$$

which approaches zero in the limit $\Delta t \to 0$. On the other hand, the velocity of the particle is likely to be extremely large, since

$$\text{Prob}\left(\frac{1}{\Delta t}(X(t+\Delta t) - X(t)) > k\right) = \int_{k\Delta t}^{\infty} \frac{1}{2\sqrt{\pi D \Delta t}} \exp\left(-\frac{x^2}{4D\Delta t}\right) dx$$
$$= \int_{\frac{k}{2}\sqrt{\frac{\Delta t}{\pi D}}}^{\infty} \exp\left(-x^2\right) dx \to \frac{1}{2}, \tag{2.220}$$

in the limit that $\Delta t \to 0$. In other words, with probability 1, the absolute value of the velocity is larger than any number k, hence infinite.

If $D = 1$, the stochastic process $X(t)$ is known as a *Wiener process*, is usually denoted by $W(t)$, and is a model of Brownian motion.

Diffusion as a Markov Process

A popular derivation of the diffusion equation is based on a Markovian random walk on a grid, as follows. We suppose that a particle moves along a one-dimensional line in discrete steps of length Δx at discrete times with time step Δt. At each step, however, the direction of motion is random, with probability $\frac{1}{2}$ of going to the left and probability $\frac{1}{2}$ of going to the right. If $p(x, t)$ is the probability of being at position x at time t, then

$$p(x, t+\Delta t) = \frac{1}{2}p(x+\Delta x, t) + \frac{1}{2}p(x-\Delta x, t). \tag{2.221}$$

Now we make the assumption that $p(x, t)$ is a smooth function of both x and t and obtain the Taylor series expansion of (2.221),

$$\Delta t \frac{\partial p}{\partial t} + O(\Delta t^2) = \frac{\Delta x^2}{2}\frac{\partial^2 p}{\partial x^2} + O(\Delta x^4). \tag{2.222}$$

In the limit that Δt and Δx both approach zero, keeping $\frac{\Delta x^2}{\Delta t} = 1$, we obtain the diffusion equation with diffusion coefficient $\frac{1}{2}$.

2.9.5 Sample Paths; the Langevin Equation

The diffusion equation describes the probability distribution that a particle is at a particular place at some time, but does not describe how the particle actually moves. The challenge, of course, is to write (and solve) an equation for motion that is random and continuous, but nowhere differentiable. Obviously, one cannot use a standard differential equation to describe the motion of such a particle. So, instead of writing

$\frac{dx}{dt}$ = something (which does not make sense, since the velocity $\frac{dx}{dt}$ of a Brownian particle is not finite), it is typical to write

$$dx = \sqrt{2D}\, dW. \tag{2.223}$$

To make careful mathematical sense of this expression requires a discussion of the Ito or Stratonovich calculus, topics that are beyond the scope of this text. However, a reasonable verbal description of what this means in practical terms is as follows. The term dW is intended to represent the fact that the displacement of a particle after a very short time interval, say dt, is a random variable having three properties, namely, it is uncorrelated with previous displacements (it has no memory and is therefore Markovian), it has zero mean, and it has variance dt, in the limit $dt \to 0$. This is also referred to as uncorrelated *Gaussian white noise*. In fact, this definition is rigged so that the probability distribution for this particle is described by the diffusion equation.

For this text, it is important to know how to numerically calculate representative sample paths, and to this end we write

$$dx = \sqrt{2D\, dt}N(0,1), \tag{2.224}$$

where $N(0,1)$ represents the Gaussian (normal) distribution with zero mean and variance 1. The interpretation of this is that at any given time one randomly chooses a number n from a normal distribution, takes a step of size $dx = \sqrt{2Ddt}n$, and then increments time by dt. It can be shown that in the limit that $dt \to 0$, this converges to the Wiener process (2.223).

Equation (2.223) is an example of a stochastic differential equation, also called a *Langevin equation*. More generally, Langevin equations are of the form

$$dx = a(x,t)\, dt + \sqrt{2b(x,t)}\, dW, \tag{2.225}$$

or, in a form that suggests a numerical algorithm,

$$dx = a(x,t)\, dt + \sqrt{2b(x,t)\, dt}N(0,1). \tag{2.226}$$

Here $a(x,t)$ represents the deterministic part of the velocity, since if there were no noise ($b(x,t) = 0$), this would be the same as the deterministic equation

$$\frac{dx}{dt} = a(x,t). \tag{2.227}$$

Thus, the displacement dx is a random variable, with mean value $a(x,t)dt$ and variance $2b(x,t)dt$, in the limit $dt \to 0$.

The special case $a(x,t) = -x$, $b(x,t) = 1$, called an *Ornstein–Uhlenbeck* process, is important in the study of molecular motors, described in Chapter 15.

2.9.6 The Fokker–Planck Equation and the Mean First Exit Time

The diffusion equation is the simplest example of an equation describing the evolution of the probability distribution function for the position of a particle. More generally,

if we suppose that the position of the particle is continuous in time (no finite jumps are possible), that the Chapman–Kolmogorov equation (2.219) holds, and that the displacement of the particle in time dt has mean $a(x,t)dt$ and variance $2b(x,t)dt$, then one can derive that the probability distribution $p(x,t)$ for the position, x, of the particle at time t is governed by

$$\frac{\partial p}{\partial t} = -\frac{\partial}{\partial x}(a(x,t)p) + \frac{\partial^2}{\partial x^2}(b(x,t)p), \tag{2.228}$$

called the *Fokker–Planck* equation. Note that, since this equation models the motion of a particle which must be at a single position y at the starting time t_0, the initial condition must be $p(x,t_0 \mid y,t_0) = \delta(x-y)$. Thus, the probability distribution for the position of the particle is the Green's function of (2.228).

More generally, it is possible to start with the Chapman–Kolmogorov equation (2.219) and derive a general version of the Fokker–Planck equation that includes discrete jump processes. This is the point of view usually taken in the stochastic processes literature, which treats the Chapman–Kolmogorov equation as a fundamental requirement of a Markov process.

An extremely important problem is the so-called *mean first exit time problem*, in which we wish to determine how long a particle stays in a particular region of space. Before we can solve this problem we must first determine the equation for the conditional probability, $p(x,t \mid y,\tau)$, as a function of y and $\tau < t$, with x and t fixed. That is, we want to know the probability distribution function for a particle with known position x at time t to have been at the location y at time $\tau < t$.

The equation governing this conditional probability is most easily derived by using the fact that $p(x,t \mid y,\tau)$ is the Green's function of the Fokker–Planck equation. It follows from the properties of Green's functions (Keener, 1998) that

$$p(x,t \mid y,\tau) = p^*(y,\tau \mid x,t), \tag{2.229}$$

where p^* is the Green's function of the adjoint equation. The adjoint equation is easily calculated using integration by parts, and assuming that p and all its derivatives vanish at infinity. It follows that p, considered as a function of y and τ, satisfies the adjoint equation

$$\frac{\partial p}{\partial \tau} = -a(y,\tau)\frac{\partial p}{\partial y} - b(y,\tau)\frac{\partial^2 p}{\partial y^2}, \tag{2.230}$$

subject to the condition $p(x,t \mid y,t) = \delta(x-y)$. This equation for the backward conditional probability is called the *backward Fokker–Planck equation*. Notice that this is a *backward* diffusion equation, which in forward time is ill posed. However, it is well posed when solved for backward times $\tau < t$.

Armed with the backward Fokker–Planck equation, we now turn our attention to the mean first exit time problem. Suppose a particle is initially at position y, inside a one-dimensional region $\alpha < y < \beta$, and that the wall at $y = \alpha$ is impermeable, but the particle can leave the region freely at $y = \beta$. If we let $\tau(y)$ represent the time at which

the particle first leaves the region having started from position y, then

$$P(\tau(y) > t) = G(y,t) = \int_\alpha^\beta p(x,t \mid y,0)\,dx, \tag{2.231}$$

which is the probability that the particle is inside the region at time t. Notice that since $p(x,0 \mid y,0) = \delta(x-y)$, $G(y,0) = 1$.

Since

$$P(\tau(y) > t) = G(y,t) = -\int_t^\infty G_t(y,s)\,ds, \tag{2.232}$$

it follows that $G_t(y,t)$ is the probability density function for the random variable $\tau(y)$. Thus, the expected value of $\tau(y)$ is

$$T(y) = E(\tau(y)) = -\int_0^\infty tG_t(y,t)\,dt = \int_0^\infty G(y,t)\,dt, \tag{2.233}$$

where we have integrated by parts to get the final expression. Note that $T(y)$ is the mean time at which a particle leaves the domain, given that it starts at y at time $t = 0$.

For a time-independent process (i.e., $a(y,t) = a(y)$, $b(y,t) = b(y)$) we have $p(x,t \mid y,0) = p(x,0 \mid y,-t)$. Hence, substituting $-t$ for t in (2.230), and writing $q(y,t \mid x) = p(x,t \mid y,0)$, it follows that q satisfies the negative backward Fokker–Planck equation

$$\frac{\partial q}{\partial t} = a(y)\frac{\partial q}{\partial y} + b(y)\frac{\partial^2 q}{\partial y^2}, \tag{2.234}$$

with $q(y,0 \mid x) = \delta(x-y)$. Integrating with respect to x from $x = \alpha$ to $x = \beta$ (since $G(y,t) = \int_\alpha^\beta p(x,t \mid y,0)\,dx$) then gives

$$\frac{\partial G}{\partial t} = a(y)\frac{\partial G}{\partial y} + b(y)\frac{\partial^2 G}{\partial y^2}, \tag{2.235}$$

with $G(y,0) = 1$. Finally, we can determine the equation for the expected value of $\tau(y)$ by integrating (2.235) in time to find

$$-1 = a(y)\frac{\partial T}{\partial y} + b(y)\frac{\partial^2 T}{\partial y^2}. \tag{2.236}$$

To completely specify the problem, we must specify boundary conditions. At impermeable boundaries, we require $\frac{\partial T}{\partial y} = 0$, while at absorbing boundaries (boundaries through which exit is allowed but reentry is not permitted) we require $T = 0$.

As an example, consider a pure diffusion process on a domain of length L with a reflecting boundary at $x = 0$ and an absorbing boundary at $x = L$. The mean first exit time satisfies the differential equation

$$DT_{xx} = -1, \tag{2.237}$$

subject to boundary conditions $T_x(0) = 0$, $T(L) = 0$. This has solution

$$T(x) = \frac{-x^2 + L^2}{2D}. \tag{2.238}$$

We readily calculate that $T(0) = \frac{L^2}{2D}$, as might be expected from (2.218). In addition, as x increases, $T(x)$ decreases, which again makes intuitive sense. The closer the particle starts to the absorbing boundary, the shorter is the mean first exit time.

2.9.7 Diffusion and Fick's Law

The Fokker–Planck equation describes the evolution of the probability distribution function for single particle diffusion. However, it also applies to the concentration of a dilute chemical species under the assumption that the chemical particles have no self-interaction. If the diffusion coefficient is homogeneous in space, then the Fokker–Planck equation and the diffusion equation are the same. However, if diffusion is not homogeneous in space, then the diffusion equation, derived using Fick's law, and the Fokker–Planck equation are not the same. With Fick's law, the diffusion equation is

$$\frac{\partial c}{\partial t} = \frac{\partial}{\partial x}\left(D\frac{\partial c}{\partial x}\right) \tag{2.239}$$

and the Fokker–Planck equation is

$$\frac{\partial c}{\partial t} = \frac{\partial^2 (Dc)}{\partial x^2}. \tag{2.240}$$

Which of these is correct?

There is a simple observation that can help answer this question. If Fick's law is correct, then at steady state the flux in a closed container is zero, so that

$$D\frac{\partial c}{\partial x} = 0, \tag{2.241}$$

implying that $c(x)$ is a uniform constant. On the other hand, if the Fokker–Planck equation is correct, then at steady state

$$D(x)c(x) = \text{constant}. \tag{2.242}$$

Notice that in this solution, the concentration varies inversely with D. That is, if D is low then c should be high, and vice versa. Further, if c is initially uniform, Fick's law predicts no change in the solution as a function of time, whereas the Fokker–Planck equation predicts transient behavior leading to a nonuniform distribution of c.

Van Milligen et al. (2006) reported a simple experimental test of this observation. They added green food coloring to water and then added differing amounts of gelatin to small quantities of the colored water. They then created gel bilayers consisting of two layers of the colored gelatin with differing gelatin concentrations under the assumption that diffusion of food coloring varies inversely with the amount of gelatin. They recorded the color intensity at several later times and compared these recordings with the solution of the partial differential equations.

The first observation is that the initially uniform concentration of food coloring did not remain uniform, but increased in regions where the gelatin density was highest. Even more striking, they were able to find very good fits of the data to the numerical

solution of the Fokker–Planck equation. These observations lead to the conclusion that Fick's law is not the correct description of chemical diffusion in media where the diffusion coefficient is not constant. The more appropriate description, coming from the Fokker–Planck equation, is that

$$J = -\nabla(Dc). \tag{2.243}$$

2.10 EXERCISES

1. A rule of thumb (derived by Einstein) is that the diffusion coefficient for a globular molecule satisfies $D \sim M^{-1/3}$, where M is the molecular weight. Determine how well this relationship holds for the substances listed in Table 2.2 by plotting D and M on a log-log plot.

2. A fluorescent dye with a diffusion coefficient $D = 10^{-7}$ cm^2/s and binding equilibrium $K_{\text{eq}} = 30$ mM is used to track the spread of hydrogen ($D_h = 4.4 \times 10^{-5}$cm^2/s). Under these conditions the measured diffusion coefficient is 8×10^{-6}cm^2/s. How much dye is present? (Assume that the dye is a fast buffer of hydrogen and that the amount of hydrogen is much less that K_{eq}.)

3. Segel, Chet and Henis (1977) used (2.18) to estimate the diffusion coefficient for bacteria. With the external concentration C_0 at 7×10^7 ml^{-1}, at times $t = 2$, 5, 10, 12.5, 15, and 20 minutes, they counted N of 1,800, 3,700, 4,800, 5,500, 6,700, and 8,000 bacteria, respectively, in a capillary of length 32 mm with 1 μl total capacity. In addition, with external concentrations C_0 of 2.5, 4.6, 5.0, and 12.0 $\times 10^7$ bacteria per milliliter, counts of 1,350, 2,300, 3,400, and 6,200 were found at $t = 10$ minutes. Estimate D.

4. Calculate the effective diffusion coefficient of oxygen in a solution containing 1.2×10^{-5} M/cm^3 myoglobin. Assume that the rate constants for the uptake of oxygen by myoglobin are $k_+ = 1.4 \times 10^{10}$cm^3 M^{-1}s^{-1} and $k_- = 11$ s^{-1}.

5. Find the maximal enhancement for diffusive transport of carbon dioxide via binding with myoglobin using $D_s = 1.92 \times 10^{-5}$ cm^2/s, $k_+ = 2 \times 10^8$ cm^3/M $\cdot$ s, $k_- = 1.7 \times 10^{-2}$/s. Compare the amount of facilitation of carbon dioxide transport with that of oxygen at similar concentration levels.

6. Devise a model to determine the rate of production of product for a "one-dimensional" enzyme capsule of length L in a bath of substrate at concentration S_0. Assume that the enzyme is confined to the domain $0 \le x \le L$ and there is no flux through the boundary at $x = 0$. Assume that the enzyme cannot diffuse within the capsule but that the substrate and product can freely diffuse into, within, and out of the capsule. Show that the steady-state production per unit volume of enzyme is less than the production rate of a reactor of the same size in which substrate is homogeneously mixed (infinite diffusion).

7. Devise a model to determine the rate of production of product for a spherical enzyme capsule of radius R_0 in a bath of substrate at concentration S_0. Assume that the enzyme cannot diffuse within the capsule but that the substrate and product can freely diffuse into, within, and out of the capsule. Show that spheres of small radius have a larger rate of production than spheres of large radius.

Hint: Reduce the problem to the nondimensional boundary value problem

$$\frac{1}{y^2}(y^2\sigma')' - \alpha^2\frac{\sigma}{\sigma+1} = 0, \tag{2.244}$$

$$\sigma'(0) = 0, \tag{2.245}$$

$$\sigma(1) = \sigma_0, \tag{2.246}$$

and solve numerically as a function of α. How does the radius of the sphere enter the parameter α?

8. Red blood cells have a passive exchanger that exchanges a single Cl^- ion for a bicarbonate (HCO_3^-) ion. Develop a model of this exchanger and find the flux.

9. Almost immediately upon entering a cell, glucose is phosphorylated in the first reaction step of glycolysis. How does this rapid and nearly unidirectional reaction affect the transmembrane flux of glucose as represented by (2.54)? How is this reaction affected by the concentration of ATP?

10. In the model of the glucose transporter (Fig. 2.6) the reaction diagram was simplified by assuming that each conformation of the transporter is equally likely, and that the affinity of the glucose binding site is unaffected by a change in conformation.

    (a) Construct a more detailed model in which these assumptions are relaxed, and calculate the flux through the model.

    (b) What is the total change in chemical potential after one cycle of the exchanger? What is the equilibrium condition?

    (c) Apply detailed balance to obtain a relationship between the rate constants.

11. Consider the model of a nonelectrogenic, 3 for 1, Na^+–Ca^{2+} exchanger. At equilibrium, the concentrations on either side of the membrane are related by the equation

$$\frac{n_e^3 c_i}{n_i^3 c_e} = 1. \tag{2.247}$$

Assume that the membrane separates two equal volumes. For a given set of initial concentrations and assuming there are no other exchange processes, what three additional conservation equations must be used to determine the equilibrium concentrations? Prove that there is a unique equilibrium solution. Hint: Give a graphical proof.

12. Simplify the model of the Na^+–Ca^{2+} exchanger (Fig. 2.9) by assuming that the binding and unbinding of Na^+ and Ca^{2+} are fast compared to the exchange processes between the inside and the outside of the cell. Write the new model equations and calculate the steady-state flux. Hint: The assumption of fast equilibrium gives

$$k_1 c_i x_1 = k_{-1} n_i^3 x_2, \tag{2.248}$$

$$k_3 n_e^3 y_2 = k_{-3} c_e y_1. \tag{2.249}$$

Then introduce the new variables $X = X_1 + X_2$ and $Y = Y_1 + Y_2$ and derive the equations for X and Y.

13. Simplify the model of Fig. 2.11 by assuming fast binding of Na^+ and K^+, and draw the reaction diagram of the simplified model. Calculate the expression for the steady-state flux.

Hint: Combine the states X_1, X_2, and X_3 into a single state, X, and similarly for Y_1, Y_2, and Y_3. Then use the equilibrium conditions

$$x_1 = K_1 x_2 \kappa_i^2, \tag{2.250}$$

$$n_i^3 x_2 = K_2 x_3, \tag{2.251}$$

$$y_3 = K_5 n_e^3 y_2, \tag{2.252}$$

$$\kappa_e^2 y_2 = K_6 y_1, \tag{2.253}$$

where $K_i = k_{-i}/k_i$, n denotes $[Na^+]$, and κ denotes $[K^+]$, to derive the differential equations for X and Y.

14. Calculate the flux of the Ran-GTP nuclear transporter. Use the information given in the text to estimate the concentrating ability of this transporter, assuming there is no difference in potential across the nuclear membrane.

15. Suppose that two compartments, each of one liter in volume, are connected by a membrane that is permeable to both K^+ and Cl^-, but not to water or protein (X). Suppose further that, as illustrated in Fig. 2.21, the compartment on the left initially contains 300 mM K^+ and 300 mM Cl^-, while the compartment on the right initially contains 200 mM protein, with valence -2, and 400 mM K^+.

    (a) Is the starting configuration electrically and osmotically balanced?

    (b) Find the concentrations at equilibrium.

    (c) Why is $[K^+]_i$ at equilibrium greater than its starting value, even though $[K^+]_i > [K^+]_e$ initially? Why does K^+ not diffuse from right to left to equalize the concentrations?

    (d) What is the equilibrium potential difference?

    (e) What would happen if the connecting membrane were suddenly made permeable to water when the system is at equilibrium? How large would the osmotic pressure be?

16. The derivation of the Gibbs–Donnan equilibrium for the case when $[S]_e$ is not fixed requires an additional constraint. Show that it is equivalent to use either $v_i[S]_i + v_e[S]_e = [S]_{tot}$ or $v_i[S']_i + v_e[S']_e = [S']_{tot}$. How must $[S]_{tot}$ and $[S']_{tot}$ be related so that the answers for the two are the same?

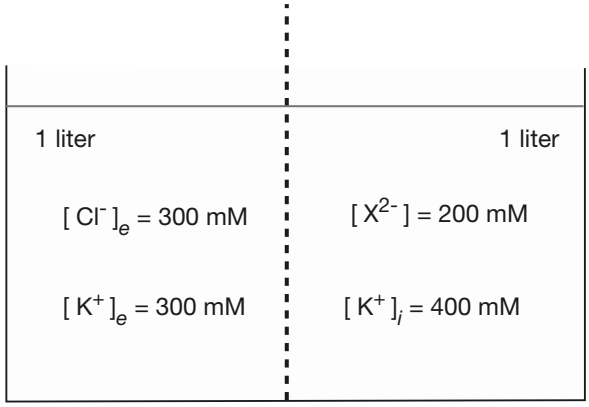


Figure 2.21 The initial configuration for Exercise 15.

17. Suppose the intracellular macromolecule X can bind b molecules of the ion S via $X + bS \rightleftarrows XS_b$. What is the effect of this buffering on the Gibbs–Donnan equilibrium potential?

18. A 1.5 oz bag of potato chips (a typical single serving) contains about 200 mg of Na^+. When eaten and absorbed into the body, how many osmoles does this bag of potato chips represent?

19. (a) Confirm that π_s in (2.138) has units of pressure.

    (b) Confirm the statement that a pressure of 25 atm corresponds to a column of water over 250 meters high.

    (c) Consider a vertical tube with a cross-sectional area of 1 cm^2. The bottom of the tube is closed with a semipermeable membrane, and 1 gram of sugar is placed in the tube. The membrane-closed end of the tube is then put into an inexhaustible supply of pure water at $T = 300$ K. What will be the height of the water in the tube at equilibrium? (The weight of a sugar molecule is 3×10^{-22} gm, and the density of water is 1 gm/cm^3).

    (d) Two columns with cross-sectional area 1 cm^2 are initially filled to a height of one meter with water at $T = 300$ K. Suppose 0.001 gm of sugar is dissolved in one of the two columns. How high will the sugary column be when equilibrium is reached?

    (e) Suppose that, in the previous question, 1 gm of sugar is dissolved in one of the two columns. What is the equilibrium height of the two columns?

20. Suppose an otherwise normal cell is placed in a bath of high extracellular K^+. What happens to the cell volume and resting potentials?

21. Based on what you know about glycolysis from Chapter 1, how would you expect anoxia (insufficient oxygen) to affect the volume of the cell? How might you incorporate this into a model of cell volume? Hint: Lactic acid does not diffuse out of a cell as does carbon dioxide.

22. Suppose 90% of the Na^+ in the bath of a squid axon is replaced by inert choline, preserving electroneutrality. What happens to the equilibrium potentials and membrane potentials?

23. Determine the effect of temperature (through the Nernst equation) on cell volume and membrane potential.

24. Simulate the time-dependent differential equations governing cell volume and ionic concentrations. What happens if the extracellular ionic concentrations are suddenly increased or decreased?

25. Ouabain is known to compete with K^+ for external K^+ binding sites of the Na^+–K^+ ATPase. Many animal cells swell and burst when treated with the drug ouabain. Why? Hint: How would you include this effect in a model of cell volume control?

26. Since the Na^+–K^+ ATPase is electrogenic, the pump rate P in the pump-leak model must also include effects from the membrane potential. What effect does membrane potential have on the expression (2.169) and how does this modification affect the solution?

27. Use (2.224) to simulate a diffusion process, and verify that the mean and variance of the process are 0 and $2Dt$, respectively, as expected.

28. Find the steady-state probability distribution for the Ornstein–Uhlenbeck process.

29. A small particle (with diffusion constant D and viscosity ν) experiences a constant load F directed to the left but is not permitted to move into the region with $x < 0$. Suppose that the particle is initially at $x = 0$. What is the expected time to first reach location $x = \delta$?

30. Suppose that a molecule enters a spherical cell of radius 5 μm at its boundary. How long will it take for the molecule to move by diffusion to find a binding target of radius 0.5 nm located at the center of the cell? Use a diffusion coefficient of 10^{-6} cm^2/s. (Hint: In higher dimensions the differential equation for the mean first exit time is $\nabla^2 T = -1$.)

CHAPTER 3

Membrane Ion Channels

Every cell membrane contains ion channels, macromolecular pores that allow specific ions to travel through the channels by a passive process, driven by their concentration gradient and the membrane potential. One of the most extensively studied problems in physiology is the regulation of such ionic currents. Indeed, in practically every chapter of this book there are examples of how the control of ionic current is vital for cellular function. Already we have seen how the cell membrane uses ion channels and pumps to maintain an intracellular environment that is different from the extracellular environment, and we have seen how such ionic separation results in a membrane potential. In subsequent chapters we will see that modulation of the membrane potential is one of the most important ways in which cells control their behavior or communicate with other cells. However, to understand the role played by ion channels in the control of membrane potential, it is first necessary to understand how membrane ionic currents depend on the voltage and ionic concentrations.

There is a vast literature, both theoretical and experimental, on the properties of ion channels. One of the best books on the subject is that of Hille (2001), to which the reader is referred for a more detailed presentation than that given here. The bibliography provided there also serves as a starting point for more detailed studies.

3.1 Current–Voltage Relations

Before we discuss specific models of ion channels, we emphasize an important fact that can be a source of confusion. Although the Nernst equation (2.104) for the equilibrium voltage generated by ionic separation can be derived from thermodynamic considerations and is thus universally applicable, there is no universal expression for the ionic

current. An expression for, say, the Na^+ current cannot be derived from thermodynamic first principles and depends on the particular model used to describe the Na^+ channels. Already we have seen two different models of ionic currents. In the previous chapter we described two common models of Na^+ current as a function of the membrane potential and the internal and external Na^+ concentrations. In the simpler model, the Na^+ current across the cell membrane was assumed to be a linear function of the membrane potential, with a driving force given by the Na^+ Nernst potential. Thus,

$$I_{Na} = g_{Na}(V - V_{Na}), \tag{3.1}$$

where $V_{Na} = (RT/F)\ln([Na^+]_e/[Na^+]_i)$ is the Nernst potential of Na^+, and where $V = V_i - V_e$. (As usual, a subscript e denotes the external concentration, while a subscript i denotes the internal concentration.) Note that the Na^+ current is zero when V is the Nernst potential, as it must be. However, we also discussed an alternative model, where integration of the Nernst–Planck equation (2.114), assuming a constant electric field, gave the Goldman–Hodgkin–Katz (GHK), or constant-field, current equation:

$$I_{Na} = P_{Na}\frac{F^2}{RT}V\left[\frac{[Na^+]_i - [Na^+]_e \exp\left(\frac{-VF}{RT}\right)}{1 - \exp\left(\frac{-VF}{RT}\right)}\right]. \tag{3.2}$$

As before, the Na^+ current is zero when V equals the Nernst potential, but here the current is a nonlinear function of the voltage and linear in the ionic concentrations. In Fig. 3.1A we compare the linear and GHK I–V curves when there is only a single ion present.

There is no one "correct" expression for the Na^+ current, or any other ionic current for that matter. Different cells have different types of ion channels, each of which may have different current–voltage relations. The challenge is to determine the current–voltage, or I–V, curve for a given ion channel and relate it to underlying biophysical mechanisms.

Our choice of these two models as examples was not coincidental, as they are the two most commonly used in theoretical models of cellular electrical activity. Not only are they relatively simple (at least compared to some of the other models discussed later in this chapter), they also provide good quantitative descriptions of many ion channels. For example, the I–V curves of open Na^+ and K^+ channels in the squid giant axon are approximately linear, and thus the linear model was used by Hodgkin and Huxley in their classic model of the squid giant axon (discussed in detail in Chapter 5). However, the I–V curves of open Na^+ and K^+ channels in vertebrate axons are better described by the GHK equation, and so nonlinear I–V curves are often used for vertebrate models (Frankenhaeuser, 1960a,b, 1963; Campbell and Hille, 1976).

Because of the importance of these two models, we illustrate another way in which they differ. This also serves to illustrate the fact that although the Nernst potential is universal when there is only one ion present, the situation is more complicated when two or more species of ion can pass through the membrane. If both Na^+ and K^+ ions

are present and both obey the GHK current equation, we showed in (2.127) that the reversal potential V_r at which there is no net current flow is

$$V_r = \frac{RT}{F} \ln \left(\frac{P_{\mathrm{Na}}[\mathrm{Na}^+]_e + P_{\mathrm{K}}[\mathrm{K}^+]_e}{P_{\mathrm{Na}}[\mathrm{Na}^+]_i + P_{\mathrm{K}}[\mathrm{K}^+]_i} \right). \tag{3.3}$$

However, if instead that the I–V curves for Na^+ and K^+ are assumed to be linear, then the reversal potential is

$$V_r = \frac{g_{\mathrm{Na}} V_{\mathrm{Na}} + g_{\mathrm{K}} V_{\mathrm{K}}}{g_{\mathrm{Na}} + g_{\mathrm{K}}}, \tag{3.4}$$

where V_{K} is the Nernst potential of K^+. Clearly, the reversal potential is model-dependent. This is due to the fact that at the reversal potential the net current flow is zero, but the individual Na^+ and K^+ currents are not. Thus, the equilibrium arguments used to derive the Nernst equation do not apply, and a universal form for the reversal potential does not exist. As an illustration of this, in Fig. 3.1B we plot the reversal potentials V_r from (3.3) and (3.4) as functions of $[\mathrm{K}^+]_e$. Although the linear and GHK I–V curves predict different reversal potentials, the overall qualitative behavior is similar, making it difficult to distinguish between a linear and a GHK I–V curve on the basis of reversal potential measurements alone.

3.1.1 Steady-State and Instantaneous Current–Voltage Relations

Measurement of I–V curves is complicated by the fact that ion channels can open or close in response to changes in the membrane potential. Suppose that in a population of ion channels, I increases as V increases. This increase could be the result of two different factors. One possibility is that more channels open as V increases while the current through an individual channel remains unchanged. It is also possible that the same number of channels remain open but the current through each one increases. To understand how each channel operates, it is necessary to separate these two factors to determine the I–V curve of a single open channel. This has motivated the definition of *steady-state* and *instantaneous* I–V curves.

If channels open or close in response to a change in voltage, but this response is slower than the change in current in a channel that is already open, it should be possible to measure the I–V curve of a single open channel by changing the voltage quickly and measuring the channel current soon after the change. Presumably, if the measurement is performed fast enough, no channels in the population have time to open or close in response to the voltage change, and thus the observed current change reflects the current change through the open channels. Of course, this relies on the assumption that the current through each open channel changes instantaneously. The I–V curve measured in this way (at least in principle) is called the instantaneous I–V curve and reflects properties of the individual open channels. If the current measurement is performed after channels have had time to open or close, then the current change reflects

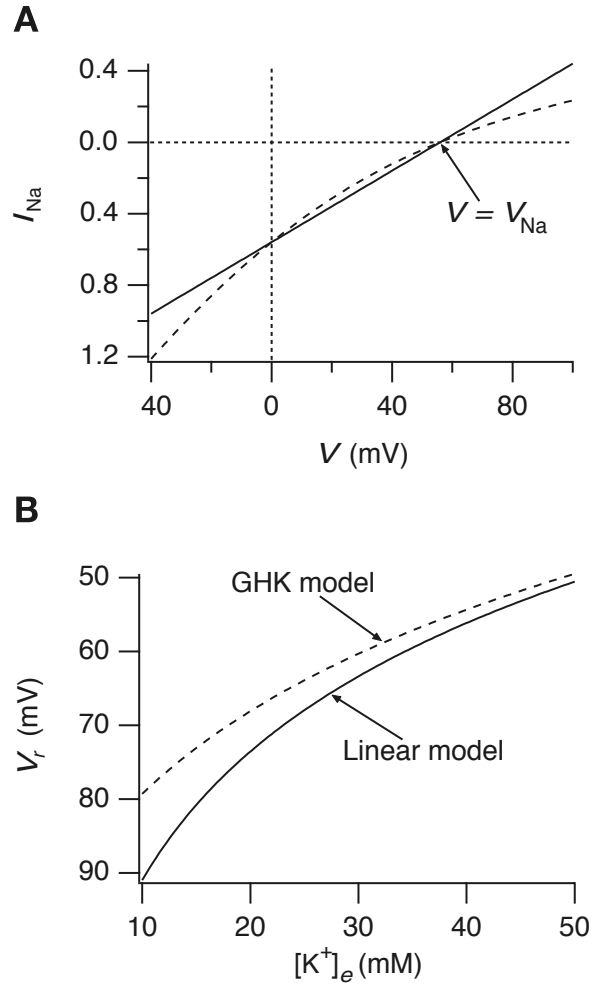


Figure 3.1 A: *I–V* curves of the linear and GHK models for Na^+ flux through a membrane. Both curves have the same reversal potential as expected, but the GHK model (dashed curve) gives a nonlinear *I–V* curve. Typical concentrations and conductances of the squid axon were used: $[Na^+]_i = 50$ mM, $[Na^+]_e = 437$ mM, and $g_{Na} = 0.01$ mS/cm^2. P_{Na} was chosen so that the GHK *I–V* curve intersects the linear *I–V* curve at $V = 0$. B: Reversal potentials of the linear and GHK models as functions of $[K^+]_e$. The membrane is permeable to both Na^+ and K^+. The same parameters as A, with $[K^+]_i = 397$ mM and $g_K = 0.367$ mS/cm^2. P_K was chosen so that the GHK *I–V* curve for K^+, with $[K^+]_e = 20$ mM, intersects the linear *I–V* curve for K^+ at $V = 0$.

the *I–V* curve of a single channel as well as the proportion of open channels. In this way one obtains a steady-state *I–V* curve.

There are two basic types of model that are used to describe ion flow through open channels, and we discuss simple versions of each. In the first type of model, the channel is described as a continuous medium, and the ionic current is determined by

the Nernst–Planck electrodiffusion equation, coupled to the electric field by means of the Poisson equation. In more complex models of this type, channel geometry and the effects of induced charge on the channel wall are incorporated. In the second type of model the channel is modeled as a sequence of binding sites, separated by barriers that impede the ion's progress: the passage of an ion through the channel is described as a process of "hopping" over barriers from one binding site to another. The height of each barrier is determined by the properties of the channel, as well as by the membrane potential. Thus, the rate at which an ion traverses the channel is a function both of the membrane potential and of the channel type. An excellent summary of the advantages and disadvantages of the two model types is given by Dani and Levitt (1990).

We also discuss simple models of the kinetics of channel gating, and the stochastic behavior of a single channel. These models are of fundamental importance in Chapter 5, where we use an early model of the voltage-dependent gating of ion channels proposed by Hodgkin and Huxley as part of their model of the action potential in the squid giant axon. More detailed recent models of channel gating are not discussed at any length. The interested reader is referred to Hille (2001), Armstrong (1981), Armstrong and Bezanilla (1973, 1974, 1977), Aldrich et al. (1983), and Finkelstein and Peskin (1984) for a selection of models of how channels can open and close in response to changes in membrane potential. An important question that we do not consider here is how channels can discriminate between different ions. Detailed discussions of this and related issues are in Hille (2001) and the references therein.

3.2 Independence, Saturation, and the Ussing Flux Ratio

One of the most fundamental questions to be answered about an ion channel is whether the passage of an ion through the channel is independent of other ions. If so, the channel is said to obey the *independence principle*.

Suppose a membrane separates two solutions containing an ion species S with external concentration c_e and internal concentration c_i. If the independence principle is satisfied, the flow of S is proportional to its local concentration, independent of the concentration on the opposite side of the membrane, and thus the flux from outside to inside, J_{in}, is

$$J_{\text{in}} = k_e c_e, \tag{3.5}$$

for some constant k_e. Similarly, the outward flux is given by

$$J_{\text{out}} = k_i c_i, \tag{3.6}$$

where in general, $k_e \neq k_i$. We let V_{S} denote the Nernst potential of the ion S, and let V denote the potential difference across the membrane. Now we let c_e^* be the external concentration for which V is the Nernst potential. Thus,

$$\frac{c_e}{c_i} = \exp\left(\frac{zV_{\text{S}}F}{RT}\right), \tag{3.7}$$

and

$$\frac{c_e^*}{c_i} = \exp\left(\frac{zVF}{RT}\right). \tag{3.8}$$

If the external concentration were c_e^* with internal concentration c_i, then there would be no net flux across the membrane; i.e., the outward flux equals the inward flux, and so

$$k_e c_e^* = k_i c_i. \tag{3.9}$$

It follows that the flux ratio is given by

$$\begin{aligned}\frac{J_{\text{in}}}{J_{\text{out}}} &= \frac{k_e c_e}{k_i c_i} = \frac{k_e c_e}{k_e c_e^*} = \frac{c_e}{c_e^*} \\ &= \frac{\exp\left(\frac{zV_S F}{RT}\right)}{\exp\left(\frac{zVF}{RT}\right)} \\ &= \exp\left[\frac{z(V_S - V)F}{RT}\right].\end{aligned} \tag{3.10}$$

This expression for the ratio of the inward to the outward flux is usually called the *Ussing flux ratio*. It was first derived by Ussing (1949), although the derivation given here is due to Hodgkin and Huxley (1952a). Alternatively, the Ussing flux ratio can be written as

$$\frac{J_{\text{in}}}{J_{\text{out}}} = \frac{c_e}{c_i} \exp\left(\frac{-zVF}{RT}\right). \tag{3.11}$$

Note that when $V = 0$, the ratio of the fluxes is equal to the ratio of the concentrations, as might be expected intuitively.

Although many ion channels follow the independence principle approximately over a range of ionic concentrations, most show deviations from independence when the ionic concentrations are sufficiently large. This has motivated the development of models that show saturation at high ionic concentrations. For example, one could assume that ion flow through the channel can be described by a barrier-type model, in which the ion jumps from one binding site to another as it moves through the channel. If there are only a limited number of binding sites available for ion passage through the channel, and each binding site can bind only one ion, then as the ionic concentration increases there are fewer binding sites available, and so the flux is not proportional to the concentration. Equivalently, one could say that each channel has a single binding site for ion transfer, but there are only a limited number of channels. However, in many of these models the Ussing flux ratio is obeyed, even though independence is not. Hence, although any ion channel obeying the independence principle must also satisfy the Ussing flux ratio, the converse is not true. We discuss saturating models later in this chapter.

Another way in which channels show deviations from independence is in flux-coupling. If ions can interact within a channel so that, for example, a group of ions

must move through the channel together, then the Ussing flux ratio is not satisfied. The most common type of model used to describe such behavior is the so-called *multi-ion model*, in which it is assumed that there are a number of binding sites within a single channel and that the channel can bind multiple ions at the same time. The consequent interactions between the ions in the channel can result in deviations from the Ussing flux ratio. A more detailed consideration of multi-ion models is given later in this chapter. However, it is instructive to consider how the Ussing flux ratio is modified by a simple multi-ion channel mechanism in which the ions progress through the channel in single file (Hodgkin and Keynes, 1955).

Suppose a membrane separates two solutions, the external one (on the right) containing an ion S at concentration c_e, and the internal one (on the left) at concentration c_i. To keep track of where each S ion has come from, all the S ions on the left are labeled A, while those on the right are labeled B. Suppose also that the membrane contains n binding sites and that S ions traverse the membrane by binding sequentially to the binding sites and moving across in single file. For simplicity we assume that there are no vacancies in the chain of binding sites. It follows that the possible configurations of the chain of binding sites are $[A_r, B_{n-r}]$, for $r = 0, \ldots, n$, where $[A_r, B_{n-r}]$ denotes the configuration such that the r leftmost sites are occupied by A ions, while the rightmost $n - r$ sites are occupied by B ions. Notice that the only configuration that can result in the transfer of an A ion to the right-hand side is $[A_n B_0]$, i.e., if the chain of binding sites is completely filled with A ions.

Now we let α denote the total rate at which S ions are transferred from left to right. Since α denotes the total rate, irrespective of labeling, it does not take into account whether an A ion or a B ion is moved out of the channel from left to right. For this reason, α is not the same as the flux of labeled ions. Similarly, let β denote the total flux of S ions, irrespective of labeling, from right to left. It follows that the rate at which $[A_r B_{n-r}]$ is converted to $[A_{r+1} B_{n-r-1}]$ is $\alpha[A_r B_{n-r}]$, and the rate of the reverse conversion is $\beta[A_{r+1} B_{n-r-1}]$. According to Hodgkin and Keynes, it is reasonable to assume that if there is a potential difference V across the membrane, then the total flux ratio obeys the Ussing flux ratio,

$$\frac{\alpha}{\beta} = \frac{c_e}{c_i} \exp\left(\frac{-VF}{RT}\right). \tag{3.12}$$

This assumption is justified by the fact that a flux of one ion involves the movement of a single charge (assuming $z = 1$) through the membrane (as in the independent case treated above) and thus should have the same voltage dependence. We emphasize that α/β is not the flux ratio of labeled ions, but the total flux ratio.

To obtain the flux ratio of labeled ions, notice that the rate at which A ions are transferred to the right-hand side is $\alpha[A_n B_0]$, and the rate at which B ions are transferred to the left-hand side is $\beta[A_0 B_n]$. Thus, the flux ratio of labeled ions is

$$\frac{J_{\text{in}}}{J_{\text{out}}} = \frac{\alpha}{\beta} \frac{[A_n B_0]}{[A_0 B_n]}. \tag{3.13}$$

At steady state there can be no net change in the distribution of configurations, so that

$$\frac{[\mathrm{A}_{r+1}\mathrm{B}_{n-r-1}]}{[\mathrm{A}_r\mathrm{B}_{n-r}]} = \frac{\alpha}{\beta}. \tag{3.14}$$

Thus,

$$\frac{J_{\text{in}}}{J_{\text{out}}} = \frac{\alpha}{\beta}\frac{[\mathrm{A}_n\mathrm{B}_0]}{[\mathrm{A}_0\mathrm{B}_n]} = \left(\frac{\alpha}{\beta}\right)^2 \frac{[\mathrm{A}_{n-1}\mathrm{B}_1]}{[\mathrm{A}_0\mathrm{B}_n]} = \cdots = \left(\frac{\alpha}{\beta}\right)^{n+1}, \tag{3.15}$$

so that

$$\frac{J_{\text{in}}}{J_{\text{out}}} = \left[\frac{c_e}{c_i}\exp\left(\frac{-VF}{RT}\right)\right]^{n+1}. \tag{3.16}$$

A similar argument, taking into account the fact that occasional vacancies in the chain arise when ions at the two ends dissociate and that these vacancies propagate through the chain, gives

$$\frac{J_{\text{in}}}{J_{\text{out}}} = \left[\frac{c_e}{c_i}\exp\left(\frac{-VF}{RT}\right)\right]^{n}. \tag{3.17}$$

Experimental data confirm this theoretical prediction (although historically, the theory was motivated by the experimental result, as is often the case). Hodgkin and Keynes (1955) showed that flux ratios in the K^+ channel of the *Sepia* giant axon could be described by the Ussing flux ratio raised to the power 2.5. Their result, as presented in modified form by Hille (2001), is shown in Fig. 3.2. Unidirectional K^+ fluxes were measured with radioactive K^+, and the ratio of the outward to the inward flux was plotted as a function of $V - V_{\text{K}}$. The best-fit line on a semilogarithmic plot has a slope of 2.5, which suggests that at least 2 K^+ ions traverse the K^+ channel simultaneously.

3.3 Electrodiffusion Models

Most early work on ion channels was based on the theory of electrodiffusion. Recall from Chapter 2 that the movement of ions in response to a concentration gradient and an electric field is described by the Nernst–Planck equation,

$$J = -D\left(\frac{dc}{dx} + \frac{zF}{RT}c\frac{d\phi}{dx}\right), \tag{3.18}$$

where J denotes the flux density, c is the concentration of the ion under consideration, and ϕ is the electrical potential. If we make the simplifying assumption that the field $d\phi/dx$ is constant through the membrane, then (3.18) can be solved to give the Goldman–Hodgkin–Katz current and voltage equations (2.123) and (2.126). However, in general there is no reason to believe that the potential has a constant gradient in the membrane. Ions moving through the channel affect the local electric field, and this local

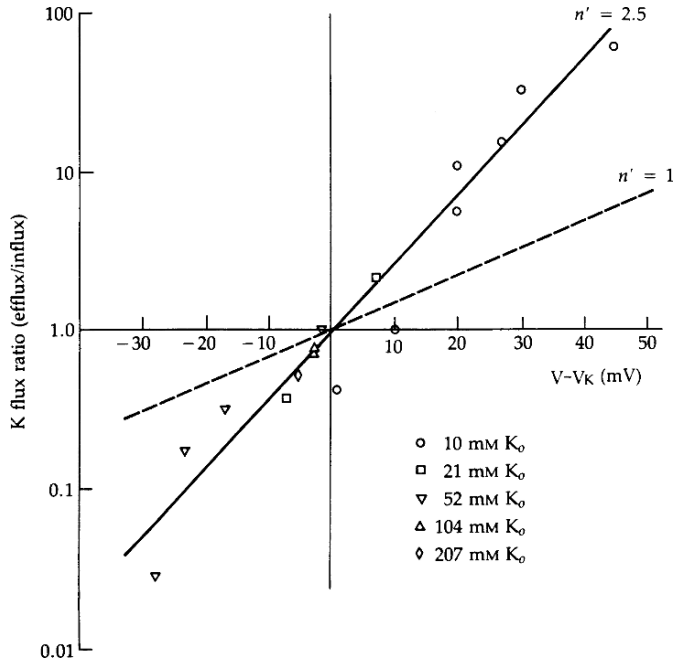


Figure 3.2 K^+ flux ratios as measured by Hodgkin and Keynes (1955), Fig. 7. Slightly modified into modern conventions by Hille (2001), page 487. K_o is the external K^+ concentration, and n' is the flux-ratio exponent, denoted by n in (3.17). (Hille, 2001, Fig. 15.7, p. 487.)

field in turn affects ionic fluxes. Thus, to determine the electric field and consequent ionic fluxes, one must solve a coupled problem.

3.3.1 Multi-Ion Flux: The Poisson–Nernst–Planck Equations

Suppose there are two types of ions, S_1 and S_2, with concentrations c_1 and c_2, passing through an ion channel, as shown schematically in Fig. 3.3.

For convenience, we assume that the valence of the first ion is $z > 0$ and that of the second is $-z$. Then, the potential in the channel $\phi(x)$ must satisfy Poisson's equation,

$$\frac{d^2\phi}{dx^2} = -\frac{zq}{\epsilon} N_a (c_1 - c_2), \tag{3.19}$$

where q is the unit electric charge, ϵ is the dielectric constant of the channel medium (usually assumed to be an aqueous solution), and N_a is Avogadro's number, necessary to convert units of concentration in moles per liter into number of molecules per liter. The flux densities J_1 and J_2 of S_1 and S_2 satisfy the Nernst–Planck equation, and at steady state dJ_1/dx and dJ_2/dx must both be zero to prevent charge buildup within the channel. Hence, the steady-state flux through the channel is described by (3.19)

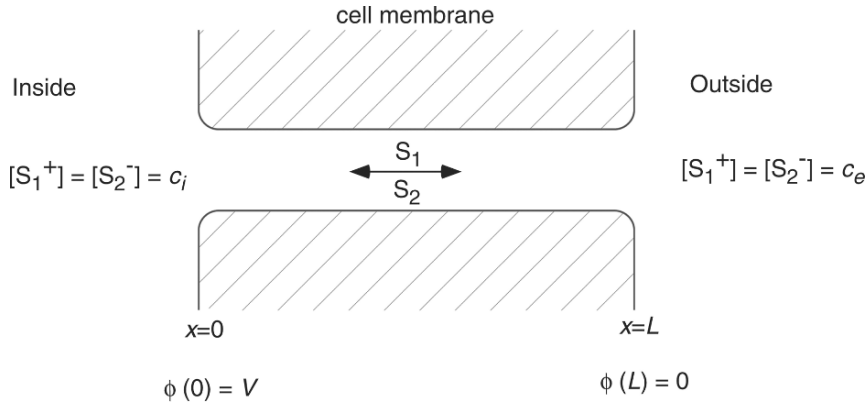


Figure 3.3 Schematic diagram of the electrodiffusion model of current through an ionic channel. Each side of the channel is electrically neutral, and both ion types can diffuse through the channel.

coupled with

$$J_1 = -D_1\left(\frac{dc_1}{dx} + \frac{zF}{RT}c_1\frac{d\phi}{dx}\right), \tag{3.20}$$

$$J_2 = -D_2\left(\frac{dc_2}{dx} - \frac{zF}{RT}c_2\frac{d\phi}{dx}\right), \tag{3.21}$$

where J_1 and J_2 are constants. To complete the specification of the problem, it is necessary to specify boundary conditions for c_1, c_2, and ϕ. We assume that the channel has length L, and that $x = 0$ denotes the left border, or inside, of the membrane. Then,

$$\begin{aligned} c_1(0) &= c_i, & c_1(L) &= c_e, \\ c_2(0) &= c_i, & c_2(L) &= c_e, \\ \phi(0) &= V, & \phi(L) &= 0. \end{aligned} \tag{3.22}$$

Note that the solutions on both sides of the membrane are electrically neutral. V is the potential difference across the membrane, defined, as usual, as the internal potential minus the external potential. While at first glance it might appear that there are too many boundary conditions for the differential equations, this is in fact not so, as the constants J_1 and J_2 are additional unknowns to be determined.

In general, it is not possible to obtain an exact solution to the Poisson–Nernst–Planck (PNP) equations (3.19)–(3.22). However, some simplified cases can be solved approximately. A great deal of work on the PNP equations has been done by Eisenberg and his colleagues (Chen et al., 1992; Barcilon, 1992; Barcilon et al., 1992; Chen and Eisenberg, 1993). Here we present simplified versions of their models, ignoring, for example, the charge induced on the channel wall by the presence of ions in the channel, and considering only the movement of two ion types through the channel. Similar models have also been discussed by Peskin (1991).

It is convenient first to nondimensionalize the PNP equations. We let $y = x/L$, $\psi = \phi zF/RT$, $v = VF/RT$, $u_k = c_k/\tilde{c}$, for $k = 1, 2, i$ and e, where $\tilde{c} = c_e + c_i$. Substituting into (3.19)–(3.21), we find

$$-j_1 = \frac{du_1}{dy} + u_1 \frac{d\psi}{dy}, \tag{3.23}$$

$$-j_2 = \frac{du_2}{dy} - u_2 \frac{d\psi}{dy}, \tag{3.24}$$

$$\frac{d^2\psi}{dy^2} = -\lambda^2(u_1 - u_2), \tag{3.25}$$

where $\lambda^2 = L^2 qFN_a\tilde{c}/(\epsilon RT)$, $j_1 = J_1 L/(\tilde{c}D_1)$, and $j_2 = J_2 L/(\tilde{c}D_1)$. The boundary conditions are

$$\begin{aligned} u_1(0) &= u_i, & u_1(1) &= u_e, \\ u_2(0) &= u_i, & u_2(1) &= u_e, \\ \psi(0) &= v, & \psi(1) &= 0. \end{aligned} \tag{3.26}$$

The Short-Channel or Low Concentration Limit

If the channel is short or the ionic concentrations on either side of the membrane are small, so that $\lambda \ll 1$, we can find an approximate solution to the PNP equations by setting $\lambda = 0$. This gives

$$\frac{d^2\psi}{dy^2} = 0, \tag{3.27}$$

and thus

$$\frac{d\psi}{dy} = -v. \tag{3.28}$$

Hence, $\lambda \approx 0$ implies that the electric potential has a constant gradient in the membrane, which is exactly the constant field assumption that was made in the derivation of the GHK equations (Chapter 2). The equation for u_1 is then

$$\frac{du_1}{dy} - vu_1 = -j_1, \tag{3.29}$$

and thus

$$u_1 = \frac{j_1}{v} + K_1 e^{vy}. \tag{3.30}$$

From the boundary conditions $u_1(0) = u_i$, $u_1(1) = u_e$ it follows that

$$j_1 = v\frac{u_i - u_e e^{-v}}{1 - e^{-v}}. \tag{3.31}$$

In dimensional form, this is

$$I_1 = FJ_1 = \frac{D_1}{L}\frac{F^2}{RT}V\left(\frac{c_i - c_e \exp(\frac{-zVF}{RT})}{1 - \exp(\frac{-zVF}{RT})}\right), \tag{3.32}$$

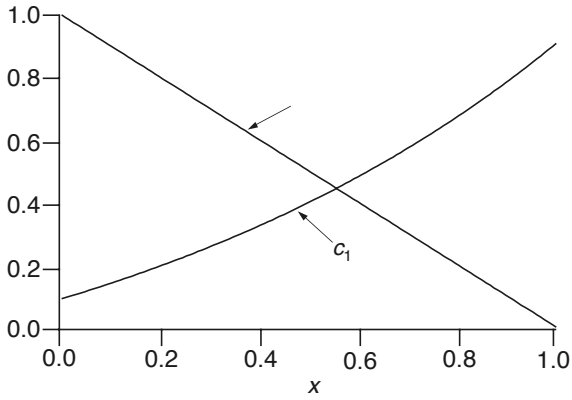


Figure 3.4 The concentration and potential profiles for the short-channel limit of the Poisson–Nernst–Planck equations. Dimensionless parameters were set arbitrarily at $u_i = 50/550 = 0.091$, $u_e = 500/550 = 0.909$, $v = 1$. In this limit the electric field is constant through the channel (the potential has a constant slope), the concentration profile is nonlinear, and the GHK *I–V* curve is obtained.

which is, as expected, the GHK current equation. Graphs of the concentration and voltage profiles through the membrane are shown in Fig. 3.4. It is reassuring that the widely used GHK equation for the ionic flux can be derived as a limiting case of a more general model.

The Long-Channel Limit

Another interesting limit is obtained by letting the length of the channel go to infinity. If we let $\eta = 1/\lambda$ denote a small parameter, the model equations are

$$-j_1 = \frac{du_1}{dy} + u_1 \frac{d\psi}{dy}, \tag{3.33}$$

$$-j_2 = \frac{du_2}{dy} - u_2 \frac{d\psi}{dy}, \tag{3.34}$$

$$-\eta^2 \frac{d^2\psi}{dy^2} = (u_1 - u_2). \tag{3.35}$$

Since there is a small parameter multiplying the highest derivative, this is a singular perturbation problem. The solution obtained by setting $\eta = 0$ does not, in general, satisfy all the boundary conditions, as the degree of the differential equation has been reduced, resulting in an overdetermined system. In the present case, however, this reduction of order is not a problem.

Setting $\eta = 0$ in (3.35) gives $u_1 = u_2$, which happens to satisfy both the left and right boundary conditions. Thus, u_1 and u_2 are identical throughout the channel. From (3.33) and (3.34) it follows that

$$\frac{d}{dy}(u_1 + u_2) = -j_1 - j_2. \tag{3.36}$$

Since both j_1 and j_2 are constants, it follows that du_1/dy is a constant, and hence, from the boundary conditions,

$$u_1 = u_2 = u_i + (u_e - u_i)y. \tag{3.37}$$

We are now able to solve for ψ. Subtracting (3.35) from (3.34) gives

$$2u_1 \frac{d\psi}{dy} = 2j, \tag{3.38}$$

where $2j = j_2 - j_1$, and hence

$$\psi = \frac{j}{u_e - u_i} \ln[u_i + (u_e - u_i)y] + K, \tag{3.39}$$

for some other constant K. Applying the boundary conditions $\psi(0) = v$, $\psi(1) = 0$ we determine j and K, with the result that

$$\psi = -\frac{v}{v_1} \ln\left[\frac{u_i}{u_e} + \left(1 - \frac{u_i}{u_e}\right) y\right], \tag{3.40}$$

where $v_1 = \ln(u_e/u_i)$ is the dimensionless Nernst potential of ion S_1. The flux density of one of the ions, say S_1, is obtained by substituting the expressions for u_1 and ψ into (3.33) to get

$$j_1 = \frac{u_e - u_i}{v_1}(v - v_1), \tag{3.41}$$

or in dimensional form,

$$J_1 = \frac{D_1}{L} \frac{zF}{RT} \frac{c_e - c_i}{\ln \frac{c_e}{c_i}} (V - \frac{RT}{zF} \ln \frac{c_e}{c_i}), \tag{3.42}$$

which is the linear I–V curve that we met previously. Graphs of the corresponding concentration and voltage profiles through the channel are shown in Fig. 3.5.

In summary, by taking two different limits of the PNP equations we obtain either the GHK I–V curve or a linear I–V curve. In the short-channel limit, ψ has a constant gradient through the membrane, but the concentration does not. In the long-channel limit the reverse is true, with a constant gradient for the concentration through the channel, but not for the potential. It is left as an exercise to prove that although the GHK equation obeys the independence principle and the Ussing flux ratio, the linear

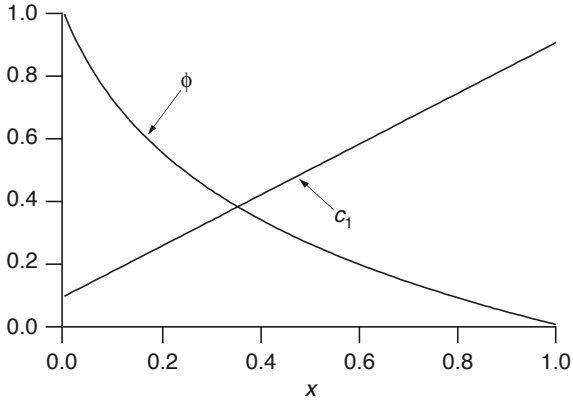


Figure 3.5 The concentration and potential profiles for the long-channel limit of the Poisson–Nernst–Planck equations. Dimensionless parameters were set arbitrarily at $u_i = 50/550 = 0.091$, $u_e = 500/550 = 0.909$, $v = 1$. In this limit the concentration profile has a constant slope, the potential profile is nonlinear, and the linear I–V curve is obtained.

I–V curve obeys neither. Given the above derivation of the linear I–V curve, this is not surprising. A linear I–V curve is obtained when either the channel is very long or the ionic concentrations on either side of the channel are very high. In either case, one should not expect the movement of an ion through the channel to be independent of other ions, and so that the independence principle is likely to fail. Conversely, the GHK equation is obtained in the limit of low ionic concentrations or short channels, in which case the independent movement of ions is not surprising.

3.4 Barrier Models

The second type of model that has been widely used to describe ion channels is based on the assumption that the movement of an ion through the channel can be modeled as the jumping of an ion over a discrete number of free-energy barriers (Eyring et al., 1949; Woodbury, 1971; Läuger, 1973). It is assumed that the potential energy of an ion passing through a channel is described by a potential energy profile of the general form shown in Fig. 3.6. The peaks of the potential energy profile correspond to barriers that impede the ion flow, while the local minima correspond to binding sites within the channel.

To traverse the channel the ion must hop from one binding site to another. According to the theory of chemical reaction rates, the rate at which an ion jumps from one binding site to the next is an exponential function of the height of the potential energy

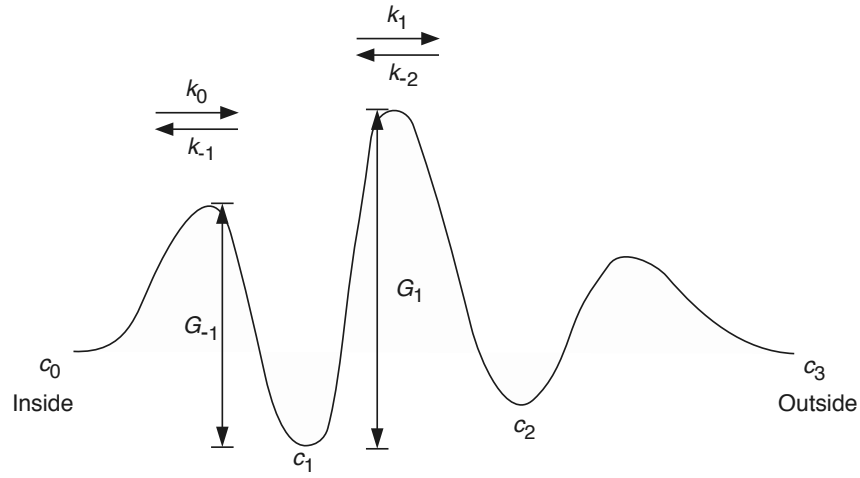


Figure 3.6 General potential energy profile for barrier models. The local minima correspond to binding sites within the channel, and the local maxima are barriers that impede the ion flow. An ion progresses through the channel by hopping over the barriers from one binding site to another.

barrier that it must cross. Thus, in the notation of the diagram,

$$k_j = \kappa \exp\left(\frac{-\Delta G_j}{RT}\right), \tag{3.43}$$

for some factor κ with units of 1/time. One of the most difficult questions in the use of this expression is deciding the precise form of the factor κ. According to Eyring rate theory (as used in this context by Hille (2001), for example), $\kappa = kT/h$, where k is Boltzmann's constant, T is the temperature, and h is Planck's constant. The derivation of this expression for κ relies on the quantization of the energy levels of the ion in some transition state as it binds to the channel binding sites. However, it is not clear that at biologically relevant temperatures energy quantization has an important effect on ionic flows. Using methods from nonequilibrium statistical thermodynamics, an alternative form of the factor has been derived by Kramers (1940), and discussions of this, and other, alternatives may be found in McQuarrie (1967) and Laidler (1969). In the appendix to this chapter, we give a derivation of Kramers' formula, but we do not enter into the debate of which answer is best. Instead, in what follows, we assume that κ is known, and independent of ΔG_j, even though, for Kramers' formula, such is not the case.

For simplicity, we assume that each local maximum occurs halfway between the local minima on each side. Barriers with this property are called *symmetrical*. An electric field in the channel also affects the rate constants. If the potential difference across the cell membrane is positive (so that the inside is more positive than the outside), it is easier for positive ions to cross the barriers in the outward direction but more difficult for positive ions to enter the cell. Thus, the heights of the barriers in the outward direction are reduced, while the heights in the inward direction are increased. If there is a potential difference of ΔV_j over the jth barrier, then

$$k_j = \kappa \exp\left[\frac{1}{RT}(-\Delta G_j + zF\Delta V_{j+1}/2)\right], \tag{3.44}$$

$$k_{-j} = \kappa \exp\left[\frac{1}{RT}(-\Delta G_{-j} - zF\Delta V_j/2)\right]. \tag{3.45}$$

The factor 2 appears because the barriers are assumed to be symmetrical, so that the maxima are lowered by $zF\Delta V_j/2$. A simple illustration of this is given in Fig. 3.7A and B and is discussed in detail in the next section.

In addition to symmetry, the barriers are assumed to have another important property, namely, that in the absence of an electric field the ends of the energy profile are at the same height, and thus

$$\sum_{j=0}^{n-1} \Delta G_j - \sum_{j=1}^{n} \Delta G_{-j} = 0. \tag{3.46}$$

If this were not so, then in the absence of an electric field and with equal concentrations on either side of the membrane, there would be a nonzero flux through the membrane, a situation that is clearly unphysiological.

A number of different models have been constructed along these general lines. First, we consider the simplest type of barrier model, in which the ionic concentration in the channel can become arbitrarily large, i.e., the channel does not saturate. This is similar to the continuous models discussed above and can be thought of as a discrete approximation to the constant field model. Because of this, nonsaturating models give the GHK I–V curve in the limit of a homogeneous membrane. We then discuss saturating barrier models and multi-ion models. Before we do so, however, it is important to note that although barrier models can provide good quantitative descriptions of some experimental data, they are phenomenological. In other words, apart from the agreement between theory and experiment, there is often no reason to suppose that the potential energy barrier used to describe the channel corresponds in any way to physical properties of the channel. Thus, although their relative simplicity has led to their widespread use, mechanistic interpretations of the models should be made only with considerable caution. Of course, this does not imply that barrier models are inferior to continuous models such as the constant field model or the Poisson–Nernst–Planck equations, which suffer from their own disadvantages (Dani and Levitt, 1990).

3.4.1 Nonsaturating Barrier Models

In the simplest barrier model (Eyring et al., 1949; Woodbury, 1971), the potential energy barrier has the general form shown in Fig. 3.7A, and it is assumed that the movement of an ion S over a barrier is independent of the ionic concentrations at the neighboring barriers. This is equivalent to assuming that the concentration of S at any particular binding site can be arbitrarily large.

The internal concentration of S is denoted by c_0, while the external concentration is denoted by c_n. There are $n-1$ binding sites (and thus n barriers) in the membrane, and the concentration of S at the jth binding site is denoted by c_j. Note the slight change in notation from above. Instead of using c_e and c_i to denote the external and internal concentrations of S, we use c_n and c_0. This allows the labeling of the concentrations on either side of the membrane to be consistent with the labeling of the concentrations at the binding sites. There is an equal voltage drop across each barrier, and thus the electrical distance between each binding site, denoted by λ, is the same. For convenience, we assume the stronger condition, that the physical distance between the binding sites is the same also, which is equivalent to assuming a constant electric field in the membrane. In the absence of an electric field, we assume that the heights of the energy barriers decrease linearly through the membrane, as in Fig. 3.7, with

$$\Delta G_j = \Delta G_0 - j\delta G, \tag{3.47}$$

for some constant increment δG. Finally, it is assumed that the flux from left to right, say, across the jth barrier, is proportional to c_{j-1}, and similarly for the flux in the opposite direction. Thus, the flux over the jth barrier, J, is given by

$$J = \lambda(k_{j-1}c_{j-1} - k_{-j}c_j). \tag{3.48}$$

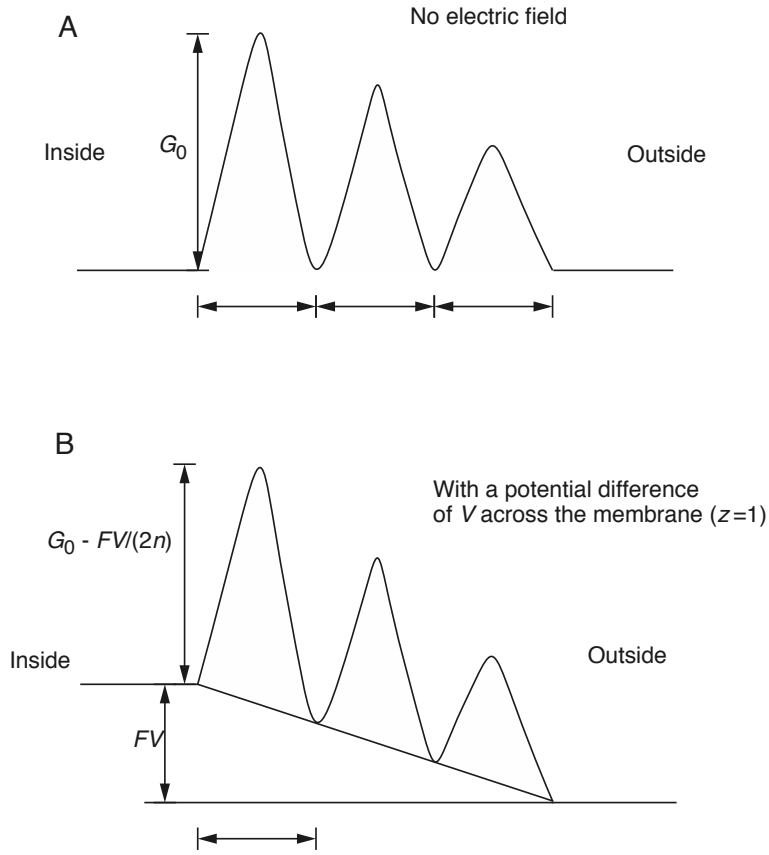


Figure 3.7 The potential energy diagram used in the nonsaturating model of Woodbury (1971). There is an equal distance between the binding sites, and the barriers are symmetrical. A. In the absence of an electric field the barrier height decreases linearly through the membrane. B. The presence of a constant electric field skews the energy profile, bringing the outside down relative to the inside. This increases the rate at which positive ions traverse the channel from inside to out and decreases their rate of entry.

Note that the units of J are concentration $\times$ distance/time, or moles per unit area per time, so J is a flux density. As usual, a flux from inside to outside (i.e., left to right) is defined as a positive flux.

At steady state the flux over each barrier must be the same, in which case we obtain a system of linear equations,

$$k_0c_0 - k_{-1}c_1 = k_1c_1 - k_{-2}c_2 = \cdots = k_{n-1}c_{n-1} - k_{-n}c_n = M, \tag{3.49}$$

where $M = J/\lambda$ is a constant. Hence

$$k_0 c_0 = (k_1 + k_{-1})c_1 - k_{-2}c_2, \tag{3.50}$$
$$k_1 c_1 = (k_2 + k_{-2})c_2 - k_{-3}c_3, \tag{3.51}$$
$$k_2 c_2 = (k_3 + k_{-3})c_3 - k_{-4}c_4, \tag{3.52}$$
$$\vdots$$

We need to determine J in terms of the concentrations on either side of the membrane, c_0 and c_n. Solving (3.51) for c_1 and substituting into (3.50) gives

$$k_0 c_0 = c_2 k_2 \phi_2 - c_3 k_{-3}\phi_1, \tag{3.53}$$

where

$$\phi_j = \sum_{i=0}^{j} \pi_i, \tag{3.54}$$
$$\pi_j = \frac{k_{-1}\cdots k_{-j}}{k_1 \cdots k_j}, \qquad \pi_0 = 1. \tag{3.55}$$

Then solving (3.52) for c_2 and substituting into (3.53) gives

$$k_0 c_0 = c_3 k_3 \phi_3 - c_4 k_{-4}\phi_2. \tag{3.56}$$

Repeating this process of sequential substitutions, we find that

$$k_0 c_0 = k_{n-1} c_{n-1}\phi_{n-1} - c_n k_{-n}\phi_{n-2}. \tag{3.57}$$

Since

$$c_{n-1} = \frac{M + k_{-n}c_n}{k_{n-1}}, \tag{3.58}$$

it follows that

$$k_0 c_0 = \phi_{n-1}(M + k_{-n}c_n) - c_n k_{-n}\phi_{n-2}, \tag{3.59}$$

and hence

$$J = \lambda M = \frac{\lambda k_0 \left(c_0 - c_n \pi_n \frac{k_n}{k_0}\right)}{\phi_{n-1}}. \tag{3.60}$$

It remains to express the rate constants in terms of the membrane potential. If there is a potential difference V across the membrane (as shown in Fig. 3.7B), the constant electric field adds $FzV/(2n)$ to the barrier when moving from right to left, and $-FzV/(2n)$ when moving in the opposite direction. Hence

$$\Delta G_j = \Delta G_0 - j\delta G - \frac{FzV}{2n}, \tag{3.61}$$
$$\Delta G_{-j} = \Delta G_0 - (j-1)\delta G + \frac{FzV}{2n}. \tag{3.62}$$

Now we use (3.43) to get

$$\frac{k_{-j}}{k_{j-1}} = \exp(-v/n), \qquad \frac{k_{-j}}{k_j} = \exp(-g - v/n), \tag{3.63}$$

where $g = \delta G/(RT)$ and $v = FzV/(RT)$. Hence

$$\pi_j = \exp(-j(g + v/n)), \tag{3.64}$$

and

$$\phi_{n-1} = \sum_{j=0}^{n-1} \exp(-j(g + v/n)) = \frac{e^{-n(g+v/n)} - 1}{e^{-(g+v/n)} - 1}, \tag{3.65}$$

so that

$$J = k_0\lambda(c_0 - c_n e^{-v})\frac{e^{-(g+v/n)} - 1}{e^{-n(g+v/n)} - 1}. \tag{3.66}$$

As expected, (3.66) satisfies both the independence principle and the Ussing flux ratio. Also, the flux is zero when v is the Nernst potential of the ion.

The Homogeneous Membrane Simplification

One useful simplification of the nonsaturating barrier model is obtained if it is assumed that the membrane is homogeneous. We model a homogeneous membrane by setting $g = \delta G/(RT) = 0$ and letting $n \to \infty$. Thus, there is no increase in barrier height through the membrane, and the number of barriers approaches infinity. In this limit, keeping $n\lambda = L$ fixed,

$$J = \frac{k_{00}\lambda^2}{L} v \frac{c_0 - c_n e^{-v}}{1 - e^{-v}}, \tag{3.67}$$

where k_{00} is the value of k_0 at $V = 0$, L is the width of the membrane, and $k_{00}\lambda^2$ is the diffusion coefficient of the ion over the first barrier in the absence of an electric field. Notice that for this to make sense it must be that k_{00} scales like λ^{-2} for small λ. In Section 3.7.3 we show that this is indeed the case.

It follows that in the homogeneous membrane case,

$$\begin{aligned} J &= \frac{D_S}{L} v \frac{c_0 - c_n e^{-v}}{1 - e^{-v}}, \\ &= P_S v \frac{c_0 - c_n e^{-v}}{1 - e^{-v}}, \end{aligned} \tag{3.68}$$

which is exactly the GHK current equation (2.122) derived previously.

3.4.2 Saturating Barrier Models: One-Ion Pores

If an ion channel satisfies the independence principle, the flux of S is proportional to [S], even when [S] gets large. However, this is not usually found to be true experimentally.

It is more common for the flux to saturate as [S] increases, reaching some maximum value as [S] gets large. This has motivated the development of models in which the flux is not proportional to [S] but is a nonlinear, saturating, function of [S]. As one might expect, equations for such models are similar to those of enzyme kinetics.

The basic assumptions behind saturating barrier models are that to pass through the channel, ions must bind to binding sites in the channel, but that each binding site can hold only a single ion (Läuger, 1973; Hille, 2001). Hence, if all the binding sites are full, an increase in ionic concentration does not increase the ionic flux—the channel is saturated. Saturating barrier models can be further subdivided into one-ion pore models, in which each channel can bind only a single ion at any one time, and multi-ion pore models, in which each channel can bind multiple ions simultaneously. The theory of one-ion pores is considerably simpler than that of multi-ion pores, and so we discuss those models first.

The Simplest One-Ion Saturating Model

We begin by considering the simplest one-ion pore model, with a single binding site. If we let S_e denote the ion outside, S_i the ion inside, and X the binding site, the passage of an ion through the channel can be described by the kinetic scheme

$$X + S_i \underset{k_{-1}}{\overset{k_0}{\rightleftarrows}} XS \underset{k_{-2}}{\overset{k_1}{\rightleftarrows}} X + S_e. \tag{3.69}$$

Essentially, the binding site acts like an enzyme that transfers the ion from one side of the membrane to the other, such as was encountered in Chapter 2 for the transport of glucose across a membrane. Following the notation of the previous section, we let c_0 denote $[S_i]$ and c_2 denote $[S_e]$. However, instead of using c_1 to denote the concentration of S at the binding site, it is more convenient to let c_1 denote the probability that the binding site is occupied. (In a population of channels, c_1 denotes the proportion of channels that have an occupied binding site.) Then, at steady state,

$$k_0 c_0 x - k_{-1} c_1 = k_1 c_1 - k_{-2} c_2 x, \tag{3.70}$$

where x denotes the probability that the binding site is empty. Note that (3.70) is similar to the corresponding equation for the nonsaturating pore, (3.49), with the only difference that x appears in the saturating model. In addition, we have a conservation equation for x,

$$x + c_1 = 1. \tag{3.71}$$

Solution of (3.70) and (3.71) gives the flux J as

$$J = k_0 c_0 x - k_{-1} c_1 = \frac{k_0 k_1 c_0 - k_{-1} k_{-2} c_2}{k_0 c_0 + k_{-2} c_2 + k_{-1} + k_1}. \tag{3.72}$$

It is important to note that J, as defined by (3.72), does not have the same units (concentration $\times$ distance/time) as in the previous model, but here has units of number of ions crossing the membrane per unit time. The corresponding transmembrane current,

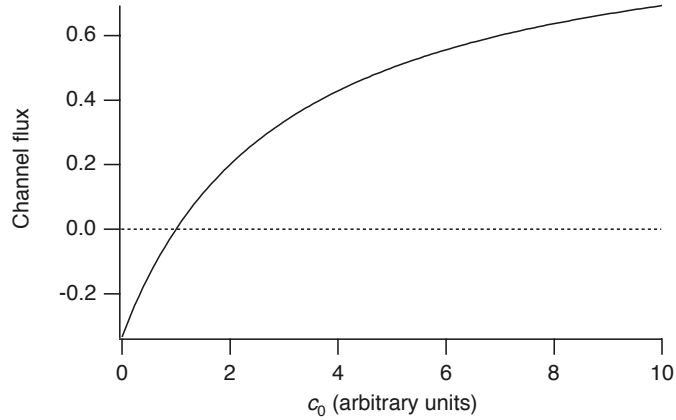


Figure 3.8 Plot of J against c_0 for the simplest saturating model with one binding site. When c_0 is small, the flux is approximately a linear function of c_0, but as c_0 increases, the flux saturates to a maximum value.

I, is given by $I = zqJ$, where q is the unit charge, and has the usual units of number of charges crossing the membrane per unit time. A plot of J as a function of c_0 is shown in Fig. 3.8. When c_0 is small, J is approximately a linear function of c_0, but as c_0 increases, J saturates at the maximum value k_1.

We now use (3.43) to express the rate constants in terms of the membrane potential. As before, we assume that the local maxima of the energy profile occur midway between the local minima; i.e., we assume that the barriers are symmetrical. However, we no longer assume that the barriers are equally spaced through the channel. If the local minimum occurs at an electrical distance δ from the left-hand side, it follows that

$$k_0 = \kappa \exp\left[\frac{1}{RT}\left(-\Delta G_0 + \delta zFV/2\right)\right], \tag{3.73}$$

$$k_1 = \kappa \exp\left[\frac{1}{RT}\left(-\Delta G_1 + (1-\delta)zFV/2\right)\right], \tag{3.74}$$

$$k_{-1} = \kappa \exp\left[\frac{1}{RT}\left(-\Delta G_{-1} - \delta zFV/2\right)\right], \tag{3.75}$$

$$k_{-2} = \kappa \exp\left[\frac{1}{RT}\left(-\Delta G_{-2} - (1-\delta)zFV/2\right)\right]. \tag{3.76}$$

Because δ denotes an electrical, not a physical, distance, it is not necessary to assume that the electric field in the membrane is constant, only that there is a drop of δV over the first barrier and $(1-\delta)V$ over the second. In general, the energy profile of any particular channel is unknown. However, the number and positions of the binding sites and the values of the local maxima and minima can, in principle at least, be determined by fitting to experimental data. We consider an example of this procedure (for a slightly more complicated model) below.

The Ussing Flux Ratio

Earlier in this chapter we stated that it is possible for a model to obey the Ussing flux ratio but not the independence principle. Single-ion saturating models provide a simple example of this. First, note that they cannot obey the independence principle, since the flux is not linearly proportional to the ionic concentration. This nonlinear saturation effect is illustrated in Fig. 3.8.

To see that the model obeys the Ussing flux ratio, it is necessary to set up the model in a slightly different form. Suppose we have two isotopes, S and $\bar{\text{S}}$, similar enough so that they have identical energy profiles in the channel. Then, we suppose that a channel has only S on the left-hand side and only $\bar{\text{S}}$ on the right. We let c denote [S] and $\bar{c}$ denote [$\bar{\text{S}}$]. Since S and $\bar{\text{S}}$ have identical energy profiles in the channel, the rate constants for the passage of $\bar{\text{S}}$ through the channel are the same as those for S. From the kinetic schemes for S and $\bar{\text{S}}$ we obtain

$$k_0 c_0 x - k_{-1} c_1 = k_1 c_1 - k_{-2} c_2 x = J_{\text{S}}, \tag{3.77}$$

$$k_0 \bar{c}_0 x - k_{-1} \bar{c}_1 = k_1 \bar{c}_1 - k_{-2} \bar{c}_2 x = J_{\bar{\text{S}}}, \tag{3.78}$$

but here the conservation equation for x is

$$x + \bar{c}_1 + c_1 = 1. \tag{3.79}$$

To calculate the individual fluxes of S and $\bar{\text{S}}$ it is necessary to eliminate x from (3.77) and (3.78) using the conservation equation (3.79). However, to calculate the flux ratio this is not necessary. Solving (3.77) for J_{S} in terms of x, c_0, and c_2, we find

$$J_{\text{S}} = x \left(\frac{k_0 c_0 - \dfrac{k_{-1} k_{-2}}{k_1} c_2}{1 + k_{-1}/k_1} \right), \tag{3.80}$$

and similarly,

$$J_{\bar{\text{S}}} = x \left(\frac{k_0 \bar{c}_0 - \dfrac{k_{-1} k_{-2}}{k_1} \bar{c}_2}{1 + k_{-1}/k_1} \right). \tag{3.81}$$

If S is present only on the left-hand side and $\bar{\text{S}}$ only on the right, we then have $c_2 = 0$ and $\bar{c}_0 = 0$, in which case

$$\frac{J_{\text{S}}}{J_{\bar{\text{S}}}} = -\frac{k_0 k_1}{k_{-1} k_{-2}} \frac{c_0}{\bar{c}_2}. \tag{3.82}$$

The minus sign on the right-hand side appears because the fluxes are in different directions. Now we substitute for the rate constants, (3.73) to (3.76), and use the fact that the

ends of the energy profile are at the same height (and thus $\Delta G_0+\Delta G_1-\Delta G_{-1}-\Delta G_{-2}=0$) to find

$$\left|\frac{J_S}{J_{\bar{S}}}\right| = \exp\left(\frac{zVF}{RT}\right)\frac{c_0}{\bar{c}_2}, \tag{3.83}$$

which is the Ussing flux ratio, as proposed.

Multiple Binding Sites

When there are multiple binding sites within the channel, the analysis is essentially the same as the simpler case discussed above, but the details are more complicated. When there are n barriers in the membrane (and thus $n-1$ binding sites), the steady-state equations are

$$k_0c_0x - k_{-1}c_1 = k_1c_1 - k_{-2}c_2 = \cdots = k_{n-1}c_{n-1} - k_{-n}c_nx = J, \tag{3.84}$$

where x is the probability that all of the binding sites are empty and c_j is the probability that the ion is bound to the jth binding site. Because the channel must be in either state x or one of the states $c_1,\ldots,c_{n-1}$ (since there is only one ion in the channel at a time), it follows that

$$x = 1 - \sum_{i=1}^{n-1} c_i. \tag{3.85}$$

It is left as an exercise to show that

$$J = \frac{k_0c_0 - k_{-n}c_n\pi_{n-1}}{\phi_{n-1} + \beta k_0c_0 + k_{-n}c_n(\alpha\phi_{n-1} - \beta\phi_{n-2})}, \tag{3.86}$$

where

$$\alpha = \sum_{j=1}^{n-1}\frac{\phi_{n-2}-\phi_{j-1}}{k_j\pi_j}, \tag{3.87}$$

$$\beta = \sum_{j=1}^{n-1}\frac{\phi_{n-1}-\phi_{j-1}}{k_j\pi_j}, \tag{3.88}$$

where ϕ_j and π_j are defined in (3.54) and (3.55).

Equation (3.86) does not satisfy the independence principle, but it does satisfy the Ussing flux ratio. However, the details are left as an exercise (Exercise 5).

3.4.3 Saturating Barrier Models: Multi-Ion Pores

We showed above that single-ion models obey the Ussing flux ratio, even though they do not obey the independence principle. This means that to model the type of channel described in Fig. 3.2 it is necessary to use models that show flux coupling as predicted by Hodgkin and Keynes (1955). Such flux coupling arises in models in which more than one ion can be in the channel at any one time. Although the equations for such

multi-ion models are essentially the same as the equations for the single-ion models described in the previous section, the analysis is complicated considerably by the fact that there are many more possible channel states. Hence, numerical techniques are the most efficient for studying such models. A great deal has been written about multi-ion models (e.g., Hille and Schwartz, 1978; Begenisich and Cahalan, 1980; Schumaker and MacKinnon, 1990; Urban and Hladky, 1979; Kohler and Heckmann, 1979). Space does not allow for a detailed discussion of the properties of these models, but so we present only a brief discussion of the simplest model. Hille and Schwartz (1978) and Hille (2001) give more detailed discussions.

Multi-ion models are based on assumptions similar to one-ion models. It is assumed that the passage of an ion through the channel can be described as the jumping of an ion over energy barriers, from one binding site to another. In one-ion models each binding site can either have an ion bound or not, and thus a channel with n binding sites can be in one of n independent states (i.e., the ion can be bound to any one of the binding sites). Hence, the steady-state ion distribution is found by solving a system of n linear equations, treating the concentrations on either side of the membrane as known. If more than one ion can be present simultaneously in the channel, the situation is more complicated. Each binding site can be in one of two states: binding an ion or empty. Therefore, a channel with n binding sites can be in any of 2^n states (at least; more states are possible if there is more than one ion type passing through the channel), and the steady-state probability distribution must be found by solving a large system of linear equations.

The simplest possible multi-ion model has three barriers and two binding sites, and so the channel can be in one of 4 possible states (Fig. 3.9). Arbitrary movements from one state to another are not possible. For example, the state OO (where both binding

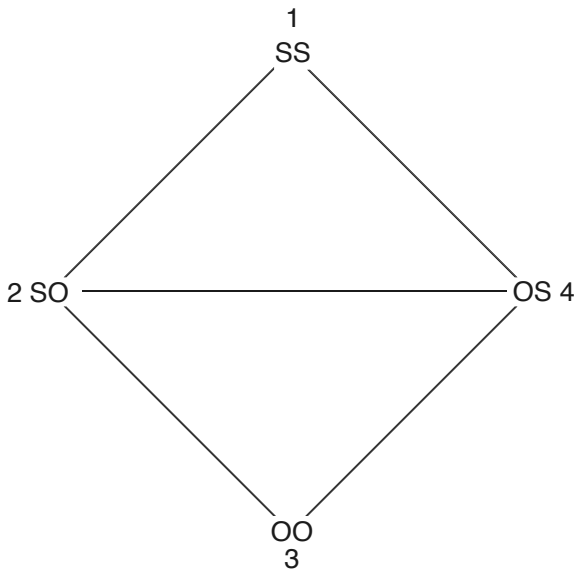


Figure 3.9 State diagram for a multi-ion barrier model with two binding sites and a single ion.

sites are empty) can change to OS or SO but cannot change to SS in a single step, as this would require two ions entering the channel simultaneously. We number the states as in Fig. 3.9 and let k_{ij} denote the rate of conversion of state i to state j. Also, let P_j denote the probability that the channel is in the jth state, and let c_e and c_i denote the external and internal ion concentrations, respectively. Then, the equations for the probabilities follow from the law of mass action; they are

$$\frac{dP_1}{dt} = -(k_{12} + k_{14})P_1 + k_{21}c_eP_2 + k_{41}c_iP_4, \tag{3.89}$$

$$\frac{dP_2}{dt} = -(k_{21}c_e + k_{23} + k_{24})P_2 + k_{12}P_1 + c_ik_{32}P_3 + k_{42}P_4, \tag{3.90}$$

$$\frac{dP_3}{dt} = -(c_ik_{32} + c_ek_{34})P_3 + k_{43}P_4 + k_{23}P_2, \tag{3.91}$$

$$\frac{dP_4}{dt} = -(k_{41}c_i + k_{42} + k_{43})P_4 + k_{14}P_1 + k_{24}P_2 + c_ek_{34}P_3. \tag{3.92}$$

The probabilities must also satisfy the conservation equation

$$\sum_{i=1}^{4} P_i = 1. \tag{3.93}$$

Using the conservation equation in place of the equation for P_4, the steady-state probability distribution is given by the linear system

$$\begin{pmatrix} -k_{12} - k_{14} & k_{21} & 0 & k_{41} \\ k_{12} & -k_{21} - k_{23} - k_{24} & c_ek_{32} & k_{42} \\ 0 & k_{23} & -c_ek_{32} - c_ik_{34} & k_{43} \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{3.94}$$

Since each rate constant is determined as a function of the voltage in the same way as one-ion models (as in, for example, (3.73)–(3.76)), solution of (3.94) gives each P_i as a function of voltage and the ionic concentrations on each side of the membrane. Finally, the membrane fluxes are calculated as the net rate of ions crossing any one barrier, and so, choosing the middle barrier arbitrarily, we have

$$J = P_2k_{24} - P_4k_{42}. \tag{3.95}$$

Although it is possible to solve such linear systems exactly (particularly with the help of symbolic manipulators such as Maple or Mathematica), it is often as useful to solve the equations numerically for a given energy profile. It is left as an exercise to show that the Ussing flux ratio is not obeyed by a multi-ion model with two binding sites and to compare the I–V curves of multi-ion and single-ion models.

3.4.4 Electrogenic Pumps and Exchangers

Recall from Chapter 2 that detailed balance required that rate constants in models of electrogenic exchangers and pumps be dependent on the membrane potential. See,

for example, (2.78) or (2.96). However, although the arguments from chemical equilibrium show that voltage-dependency must exist, they do not specify exactly which rate constants depend on the voltage, or what the functional dependency is. As we have come to expect, it is much more difficult to answer these questions. Just as there are many ways to model ionic current flow, so there are many ways to model how rate constants depend on the membrane potential. In addition, depending on the exact assumptions, any of the steps in the model could depend on membrane potential. In other words, not only are there a number of ways to model the voltage-dependence when it occurs, there are also many places where it could occur. It is, in general, a very difficult task to determine the precise place and nature of the voltage-dependence.

One simple approach is to assume that the conformational change of the carrier protein is the step that moves the charge across the membrane, and thus requires the crossing of a free energy barrier. Consider the diagram shown in Fig. 2.9. If we assume that the transition from state X_2 to Y_2 involves the movement of 2 positive ions across an energy barrier and a potential difference V, then we can model the rate constants as

$$k_2 = \kappa \exp\left[\frac{1}{RT}(-\Delta G_+ + 2FV/2)\right] \tag{3.96}$$

$$= \bar{k}_2 \exp\left(\frac{FV}{RT}\right), \tag{3.97}$$

$$k_{-2} = \kappa \exp\left[\frac{1}{RT}(-\Delta G_- - 2FV/2)\right] \tag{3.98}$$

$$= \bar{k}_{-2} \exp\left(\frac{-FV}{RT}\right), \tag{3.99}$$

where $\bar{k}_2 = \kappa \exp[-\Delta G_+/(RT)]$ and similarly for $\bar{k}_{-2}$. In (3.96) and (3.98) $2FV$ is divided by 2 as we assume, for simplicity, that the energy barrier occurs halfway through the membrane.

If we make similar assumptions for k_4 and k_{-4}, i.e., that these transitions involve the reverse movement of 3 positive charges across an energy barrier and a potential difference, we obtain similar equations for those rate constants. Then

$$\frac{k_{-2}}{k_2}\frac{k_{-4}}{k_4} = \frac{\bar{k}_{-2}}{\bar{k}_2}\frac{\bar{k}_{-4}}{\bar{k}_4}\exp\left(\frac{-2FV}{RT}\right)\exp\left(\frac{3FV}{RT}\right)$$

$$= \frac{\bar{k}_{-2}}{\bar{k}_2}\frac{\bar{k}_{-4}}{\bar{k}_4}\exp\left(\frac{FV}{RT}\right), \tag{3.100}$$

from which it follows that $K_1K_2K_3K_4 = \exp\left(\frac{FV}{RT}\right)$ (cf. (2.78)), which is the necessary equilibrium condition.

3.5 Channel Gating

So far in this chapter we have discussed how the current through a single open channel depends on the membrane potential and the ionic concentrations on either side of the membrane. However, it is of equal interest to determine how ionic channels open and close in response to voltage. As described in Chapter 5, the opening and closing of ionic channels in response to changes in the membrane potential is the basis for electrical excitability and is thus of fundamental significance in neurophysiology.

Recall that there is an important difference between the instantaneous and steady-state I–V curves. In general, the current through a population of channels is the product of three terms,

$$I = Ng(V,t)\phi(V), \tag{3.101}$$

where $\phi(V)$ is the I–V curve of a single open channel, and $g(V,t)$ is the proportion of open channels in the population of N channels. In the previous sections we discussed electrodiffusion and barrier models of $\phi(V)$; in this section we discuss models of the dependence of g on voltage and time.

Consider, for example, the curves in Fig. 3.10, which show typical responses of populations of Na^+ and K^+ channels. When the voltage is stepped from -65 mV to -9 mV, and held fixed at the new level, the K^+ conductance (g_K) slowly increases to a new level, while the Na^+ conductance (g_{Na}) first increases and then decreases.

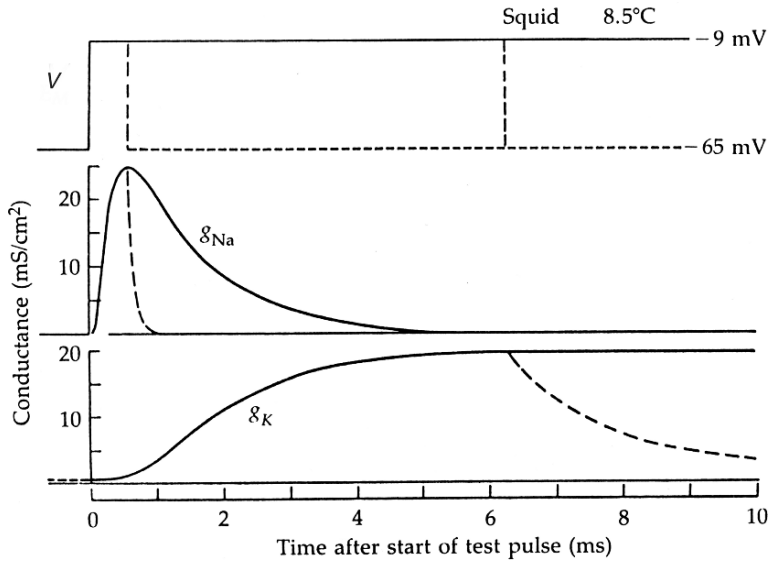


Figure 3.10 Na^+ and K^+ conductances as a function of time after a step change in voltage from -65 mV to -9 mV. The dashed line shows that after repolarization g_{Na} recovers quickly, and g_K recovers more slowly. (Hille, 2001, Fig. 2.11, p. 41.)

From these data we can draw several conclusions. First, as the voltage increases, the proportion of open K^+ channels increases. Second, although the proportion of open Na^+ channels initially increases, a second process is significant at longer times, as the Na^+ channel is inactivated. Thus, Na^+ channels first activate and then inactivate.

3.5.1 A Two-State K^+ Channel

The simplest model of the K^+ channel assumes that the channel can exist in either a closed state, C, or an open state, O, and that the rate of conversion from one state to another is dependent on the voltage. Thus,

$$\mathrm{C} \underset{\beta(V)}{\overset{\alpha(V)}{\rightleftharpoons}} \mathrm{O}. \tag{3.102}$$

If g denotes the proportion of channels in the open state (so $1 - g$ is the proportion of closed channels), the differential equation for the rate of change of g is

$$\frac{dg}{dt} = \alpha(V)(1 - g) - \beta(V)g. \tag{3.103}$$

Under voltage-clamp conditions (i.e., where the voltage is piecewise constant, as in Fig. 3.10), α and β are constants, and thus one can readily solve for g as a function of time. Equation (3.103) is often written as

$$\tau_g(V)\frac{dg}{dt} = g_\infty(V) - g, \tag{3.104}$$

where $g_\infty(V) = \alpha/(\alpha + \beta)$ is the steady-state value of g, and $\tau_g(V) = 1/(\alpha + \beta)$ is the time constant of approach to the steady state. From experimental data, such as those shown in Fig. 3.10, one can obtain values for g_∞ and τ_g, and thus α and β can be unambiguously determined.

The form of $g_\infty(V)$ can be determined from free energy arguments. The reason for voltage dependence must be that the subunit is charged and that to change from one conformation to another charges must move in the potential field. This movement of charge is a current, called the gating current. Now the difference in free energy between the two conformations is of the form

$$\Delta G = \Delta G^0 + aFV, \tag{3.105}$$

where ΔG^0 is the free energy difference between the two states in the absence of a potential, and a is a constant related to the number of charges that move and the relative distance they move during a change of conformation. It follows that the equilibrium constant for the subunit must be of the form

$$\frac{\beta}{\alpha} = k_0 \exp\left(\frac{aFV}{RT}\right), \tag{3.106}$$

in which case

$$g_\infty(V) = \frac{\alpha}{\alpha + \beta} = \frac{1}{1 + k_0 \exp(\frac{aFV}{RT})}, \tag{3.107}$$

which can also be expressed in the form

$$g_\infty(V) = \frac{1}{2} + \frac{1}{2}\tanh(b(V - V_0)). \tag{3.108}$$

3.5.2 Multiple Subunits

An important generalization of the two-state model occurs when the channel is assumed to consist of multiple identical subunits, each of which can be in either the closed or open state. For example, suppose that the channel consists of two identical subunits, each of which can be closed or open. Then, the channel can take any of four possible states, S_{00}, S_{10}, S_{01}, or S_{11}, where the subscripts denote the different subunits, with 1 and 0 denoting open and closed subunits, respectively. A general model of this channel involves three differential equations (although there is a differential equation for each of the four variables, one equation is superfluous because of the conservation equation $S_{00} + S_{10} + S_{01} + S_{11} = 1$), but we can simplify the model by grouping the channel states with the same number of closed and open subunits. Because the subunits are identical, there should be no difference between S_{10} and S_{01}, and thus they are amalgamated into a single variable.

So, we let S_i denote the group of channels with exactly i open subunits. Then, conversions between channel groups are governed by the reaction scheme

$$S_0 \underset{\beta}{\overset{2\alpha}{\rightleftarrows}} S_1 \underset{2\beta}{\overset{\alpha}{\rightleftarrows}} S_2. \tag{3.109}$$

The corresponding differential equations are

$$\frac{dx_0}{dt} = \beta x_1 - 2\alpha x_0, \tag{3.110}$$

$$\frac{dx_2}{dt} = \alpha x_1 - 2\beta x_2, \tag{3.111}$$

where x_i denotes the proportion of channels in state S_i, and $x_0 + x_1 + x_2 = 1$. We now make the change of variables $x_2 = n^2$, where n satisfies the differential equation

$$\frac{dn}{dt} = \alpha(1 - n) - \beta n. \tag{3.112}$$

A simple substitution shows that (3.110) and (3.111) are satisfied by $x_0 = (1 - n)^2$ and $x_1 = 2n(1 - n)$.

In fact, we can derive a stronger result. We let

$$x_0 = (1 - n)^2 + y_0, \tag{3.113}$$

$$x_2 = n^2 + y_2, \tag{3.114}$$

so that, of necessity, $x_1 = 2n(1-n) - y_0 - y_2$. It follows that

$$\frac{dy_0}{dt} = -2\alpha y_0 - \beta(y_0 + y_2), \tag{3.115}$$

$$\frac{dy_2}{dt} = -\alpha(y_0 + y_2) - 2\beta y_2. \tag{3.116}$$

This is a linear system of equations with eigenvalues $-(\alpha+\beta)$, $-2(\alpha+\beta)$, and so y_0, y_2 go exponentially to zero. This means that $x_0 = (1-n)^2$, $x_2 = n^2$ is an invariant stable manifold for the original system of equations; the solutions cannot leave this manifold, and with arbitrary initial data, the flow approaches this manifold exponentially. Notice that this is a stable invariant manifold even if α and β are functions of time (so they can depend on voltage or other concentrations).

This argument generalizes to the case of k identical independent binding sites where the invariant manifold for the flow is the binomial distribution with probability n satisfying (3.112) (see Exercise 14). Thus, the channel conductance is proportional to n^k, where n satisfies the simple equation (3.112). This multiple subunit model of channel gating provides the basis for the model of excitability that is examined in Chapter 5.

3.5.3 The Sodium Channel

A more complex model is needed to explain the behavior of the Na^+ channel, which both activates and inactivates. The simplest approach is to extend the above analysis to the case of multiple subunits of two different types, m and h, say, where each subunit can be either closed or open, and the channel is open, or conducting, only when all subunits are open. To illustrate, we assume that the channel has one h subunit and two m subunits. The reaction diagram of such a channel is shown in Fig. 3.11. We let S_{ij} denote the channel with i open m subunits and j open h subunits, and we let x_{ij} denote the fraction of channels in state S_{ij}. The dynamics of x_{ij} are described by a system of six differential equations. However, as above, direct substitution shows that this system of equations has an invariant manifold $x_{00} = (1-m)^2(1-h)$, $x_{10} = 2m(1-m)(1-h)$,

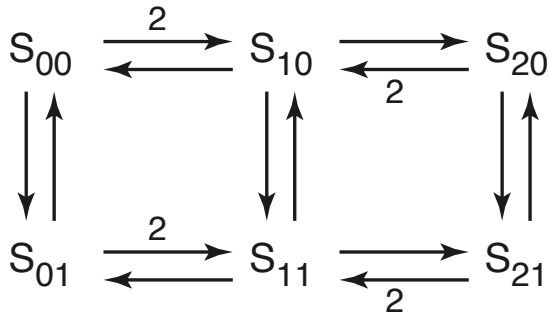


Figure 3.11 Diagram of the possible states in a model of the Na^+ channel.

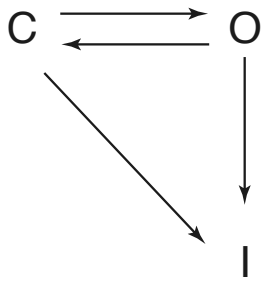


Figure 3.12 A: Schematic diagram of the states of the Na^+ channel. C, O, and I denote the closed, open, and inactivated states, respectively.

$x_{20} = m^2(1-h)$, $x_{01} = (1-m)^2 h$, $x_{11} = 2m(1-m)h$, and $x_{21} = m^2 h$, provided

$$\frac{dm}{dt} = \alpha(1-m) - \beta m, \tag{3.117}$$

$$\frac{dh}{dt} = \gamma(1-h) - \delta h. \tag{3.118}$$

Furthermore, the invariant manifold is stable. A model of this type was used by Hodgkin and Huxley in their model of the nerve axon, which is discussed in detail in Chapter 5.

In an alternate model of the Na^+ channel (Aldrich et al., 1983; Peskin, 1991), it is assumed that the Na^+ channel can exist in three states, closed (C), open (O), or inactivated (I), and that once the channel is inactivated, it cannot return to either the closed or the open state (Fig. 3.12). Thus, the state I is absorbing. While this is clearly not true in general, it is a reasonable approximation at high voltages.

As before, we let g denote the proportion of open channels and let c denote the proportion of closed channels. Then,

$$\frac{dc}{dt} = -(\alpha+\delta)c + \beta g, \tag{3.119}$$

$$\frac{dg}{dt} = \alpha c - (\beta+\gamma)g, \tag{3.120}$$

where the proportion of channels in the inactivated state is $i = 1 - c - g$. Initial conditions are $c(0) = 1$, $g(0) = 0$, i.e., all the channels are initially in the closed state. This system of first-order differential equations can be solved directly to give

$$g(t) = a(e^{\lambda_1 t} - e^{\lambda_2 t}), \tag{3.121}$$

where $\lambda_2 < \lambda_1 < 0$ are the roots of the characteristic polynomial

$$\lambda^2 + (\alpha+\beta+\gamma+\delta)\lambda + (\alpha+\delta)(\beta+\gamma) - \alpha\beta = 0, \tag{3.122}$$

and where

$$g'(0) = \alpha = a(\lambda_1 - \lambda_2) > 0. \tag{3.123}$$

As in the simple two-state model, the function g can be fit to data to determine the parameters a, λ_1, and λ_2. However, unlike the two-state model, the rate constants

cannot be determined uniquely from these parameters. For, since λ_1 and λ_2 are the roots of (3.122), it follows that

$$\alpha + \beta + \gamma + \delta = -\lambda_1 - \lambda_2, \tag{3.124}$$

$$(\alpha + \delta)(\beta + \gamma) - \alpha\beta = \lambda_1\lambda_2. \tag{3.125}$$

Thus, there are only three equations for the four unknowns, α, β, γ, and δ, so the system is underdetermined (see Exercise 16). This problem cannot be resolved using the macroscopic data that have been discussed so far, but requires data collected from a single channel, as described in Section 3.6.

3.5.4 Agonist-Controlled Ion Channels

Many ion channels are controlled by agonists, rather than by voltage. For example, the opening of ion channels in the postsynaptic membrane of the neuromuscular junction (Chapter 8) is controlled by the neurotransmitter acetylcholine, while in the central nervous system a host of neurotransmitters such as glutamate, dopamine, γ-aminobutyric acid (GABA), and serotonin have a similar role. The inositol trisphosphate receptor and ryanodine receptor are other important agonist-controlled ion channels (Chapter 7).

Early theories of agonist-controlled ion channels (Clark, 1933) assumed that the channel was opened simply by the binding of the agonist. Thus,

$$\mathrm{A} + \mathrm{T} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{AT}, \tag{3.126}$$

where the state AT is open. However, this simple theory is unable to account for a number of experimental observations. For example, it can happen that only a fraction of channels are open at any given time, even at high agonist concentrations, a result that cannot be explained by this simple model.

In 1957, del Castillo and Katz proposed a model that explicitly separated the agonist-binding step from the gating step:

$$\begin{array}{ccccccc} \mathrm{A} & + & \mathrm{T} & \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} & \mathrm{AT} & \underset{\alpha}{\overset{\beta}{\rightleftharpoons}} & \mathrm{AR}. \\ \text{agonist} & & \begin{array}{c}\text{state 3}\\ \text{closed}\\ \text{unoccupied}\end{array} & & \begin{array}{c}\text{state 2}\\ \text{closed}\\ \text{occupied}\end{array} & & \begin{array}{c}\text{state 1}\\ \text{open}\\ \text{occupied}\end{array} \end{array} \tag{3.127}$$

Note that the only open state is AR (state 1; the slightly unusual numbering of the states follows Colquhoun and Hawkes, 1981). Thus, in this model, binding of the agonist places the channel into an occupied state that allows, but does not require, opening. The agonist-binding step is controlled by the *affinity* of the channel for the agonist, while the gating is determined by the *efficacy* of the agonist. This separation of affinity and efficacy has proven to be an extremely powerful way of understanding agonist-controlled channels, and is at the heart of practically all modern approaches (Colquhoun, 2006).

The conductance of a population of agonist-controlled channels is determined as the solution of the system of differential equations

$$\frac{d\Phi_1}{dt} = \beta\Phi_2 - \alpha\Phi_1, \tag{3.128}$$

$$\frac{d\Phi_2}{dt} = \alpha\Phi_1 + k_1 a(1 - \Phi_2 - \Phi_1) - (\beta + k_{-1})\Phi_2, \tag{3.129}$$

where Φ_1, Φ_2, and $1 - \Phi_1 - \Phi_2$ represent the percentage of channels in states 1, 2, and 3, respectively, and a is the concentration of agonist A. The solution of this system of differential equations is easy to determine, provided a is constant. However, the practical usefulness of this exact solution is extremely limited, since in any realistic situation a is changing in time.

The steady-state solution is also readily found to be

$$\Phi_1 = \frac{1}{1 + \frac{\beta}{\alpha} + \frac{\beta k_{-1}}{\alpha k_1 a}}, \tag{3.130}$$

and this can be fit to data to find the equilibrium constants $\frac{\beta}{\alpha}$ and $\frac{k_{-1}}{k_1}$. However, complete determination of the four kinetic parameters is much more challenging. One could imagine a "concentration clamp" experiment, in which the concentration of a is suddenly switched from one level to another and the conductance of the channels monitored. From these data one could then determine the two eigenvalues of the system (3.128)–(3.129). However, usually such experiments are very difficult to perform. In Section 3.6 we show that there is more information contained in single-channel recordings and how this additional information can be used to determine the kinetic parameters of channel models.

3.5.5 Drugs and Toxins

Many drugs act by blocking a specific ion channel. There are numerous specific channel blockers, such as Na^+ channel blockers, K^+ channel blockers, Ca^{2+} channel blockers, and so on. In fact, the discovery of site-specific and channel-specific blockers has been of tremendous benefit to the experimental study of ion channels. Examples of important channel blockers include verapamil (Ca^{2+}-channel blocker), quinidine, sotolol, nicotine, DDT, various barbiturates (K^+-channel blockers), tetrodotoxin (TTX, the primary ingredient of puffer fish toxin), and scorpion toxins (Na^+-channel blockers).

To include the effects of a drug or toxin like TTX in a model of a Na^+ channel is a relatively simple matter. We assume that a population P of Na^+ channels is available for ionic conduction and that a population B is blocked because they are bound by the toxin. Thus,

$$\mathrm{P} + \mathrm{D} \underset{k_-}{\overset{k_+}{\rightleftarrows}} \mathrm{B}, \tag{3.131}$$

where D represents the concentration of the drug. Clearly, $P + B = P_0$, so that

$$\frac{dP}{dt} = k_-(P_0 - P) - k_+ DP, \tag{3.132}$$

and the original channel conductance must be modified by multiplying by the percentage of unbound channels, P/P_0.

In steady state, we have

$$\frac{P}{P_0} = \frac{K_d}{K_d + D}. \tag{3.133}$$

The remarkable potency of TTX is reflected by its small equilibrium constant K_d, as $K_d \approx 1$–5 nM for Na^+ channels in nerve cells, and $K_d \approx 1$–10 μM for Na^+ channels in cardiac cells. By contrast, verapamil has $K_d \approx$140–940 μM.

Other important drugs, such as lidocaine, flecainide, and encainide are so-called *use-dependent* Na^+-channel blockers, in that they interfere with the Na^+ channel only when it is open. Thus, the more the channel is used, the more likely that it is blocked. Lidocaine is an important drug used in the treatment of cardiac arrhythmias. The folklore explanation of why it is useful is that because it is use-dependent, it helps prevent high-frequency firing of cardiac cells, which is commonly associated with cardiac arrhythmias. In fact, lidocaine, flecainide, and encainide are officially classified as antiarrhythmic drugs, even though it is now known that flecainide and encainide are proarrhythmic in certain postinfarction (after a heart attack) patients. A full explanation of this behavior is not known.

To keep track of the effect of a use-dependent drug on a two-state channel, we suppose that there are four classes of channels, those that are closed but unbound by the drug (C), those that are open and unbound by the drug (O), those that are closed and bound by the drug (CB), and those that are open and bound by the drug (OB) (but unable to pass a current). For this four-state model a reasonable reaction mechanism is

$$\begin{aligned}
\mathrm{C} &\underset{\beta}{\overset{\alpha}{\rightleftarrows}} \mathrm{O}, \\
\mathrm{CB} &\underset{\beta}{\overset{\alpha}{\rightleftarrows}} \mathrm{OB}, \\
\mathrm{CB} &\xrightarrow{k_+} \mathrm{C} + \mathrm{D}, \\
\mathrm{O} + \mathrm{D} &\underset{k_-}{\overset{k_+}{\rightleftarrows}} \mathrm{OB}.
\end{aligned}$$

Notice that we have assumed that the drug does not interfere with the process of opening and closing, only with the flow of ionic current, and that the drug can bind the channel only when it is open. It is now a straightforward matter to find the differential equations governing these four states, and we leave this as an exercise.

This is not the only way that drugs might interfere with a channel. For example, for a channel with multiple subunits, the drug may bind only when certain of the subunits are

in specific states. Indeed, the binding of drugs with channels can occur in many ways, and there are numerous unresolved questions concerning this complicated process.

3.6 Single-Channel Analysis

Since the late 1970s, the development of patch-clamp recording techniques has allowed the measurement of ionic current through a small piece of cell membrane, containing only a few, or even a single, ionic channel (Hamill et al., 1981; Sakmann and Neher, 1995; Neher and Sakmann received the 1991 Nobel Prize in Physiology or Medicine for their development of the patch-clamp technique).

Much of the mathematical theory of how to analyze single-channel recordings was worked out in a series of papers by Colquhoun and Hawkes (1977, 1981, 1982). As is true of most things written by Colquhoun and Hawkes, these are eminently readable. However, newcomers to the field should first read the two chapters in the *Plymouth Workshop Handbook on Microelectrode Techniques* (Colquhoun, 1994; Colquhoun and Hawkes, 1994), since these are an excellent introduction. The chapter by Colquhoun and Hawkes in the book by Sakmann and Neher (1995) is also a valuable reference.

An example of an experimental record for Na^+ channels is given in Fig. 3.13. The current through an individual channel is stochastic (panel A) and cannot be described by a deterministic process. Nevertheless, the ensemble average over many experiments (panel B) is deterministic and reproduces the same properties that are seen in the macroscopic measurements of Fig. 3.10. However, the single-channel recordings contain more information than does the ensemble average.

What information is available from single-channel recordings that is not available from ensemble averages? First of all, one can measure how long a channel is open on average, or more generally, the distribution of open times. Similarly, one can measure the distribution of times the channel is in the closed state. If there are additional dynamical processes underlying the opening and closing of channels, as there are with Na^+ channels, one can measure (for example) how many times a channel opens (and closes) before it is inactivated (permanently closed) or how many channels fail to open even once.

The most common models of ion channels are discrete-space continuous-time Markov processes, the basic theory of which was described in Section 2.9.2. Since it is this theory that lies at the heart of the analysis of single-channel data, the reader is encouraged to review the relevant sections of Chapter 2 before continuing.

3.6.1 Single-Channel Analysis of a Sodium Channel

Consider the Na^+ channel model shown in Fig. 3.12. For this model there are two obvious waiting-time problems. The probability that the amount of time spent in the closed state before opening is less than t is governed by the differential equation

$$\frac{dP}{dt} = \alpha(1 - P), \tag{3.134}$$

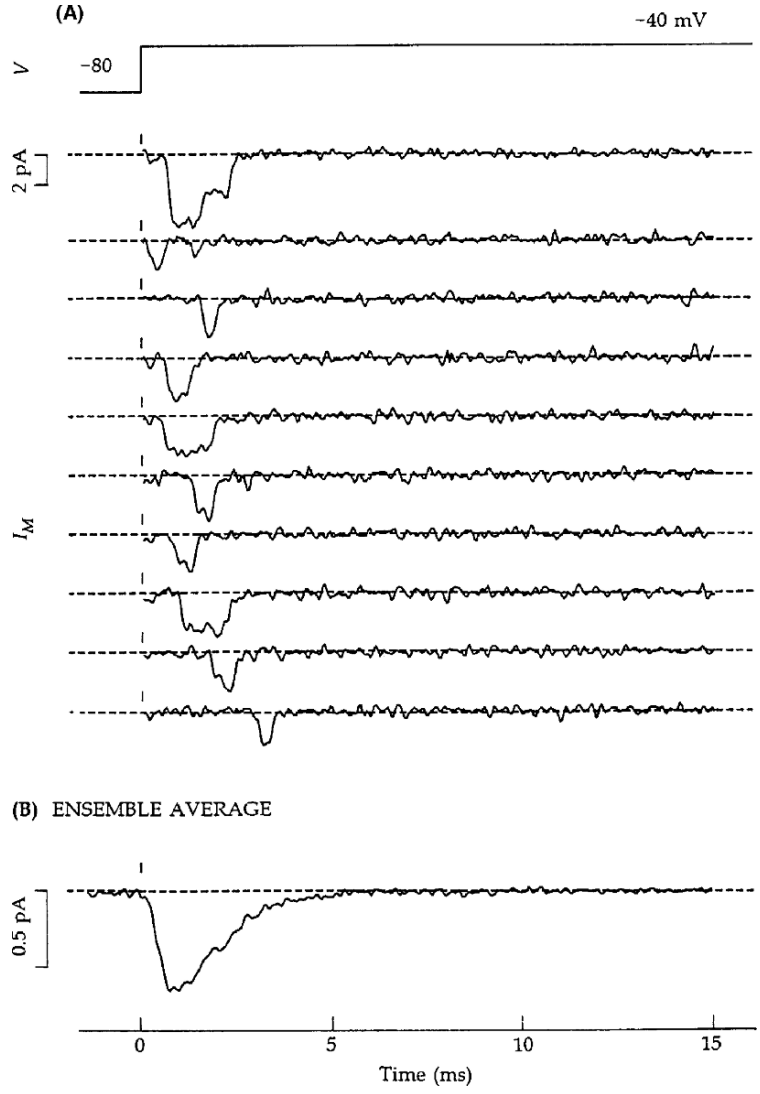


Figure 3.13 A: Na^+ currents from a single channel (or possibly two in the first trace) following a voltage step from −80 mV to −40 mV. B: Average open probability of the Na^+ channel, obtained by averaging over many traces of the type shown in A. (Hille, 2001, Fig. 3.16, p. 90.)

with $P(0) = 1$, and therefore is

$$P(t) = 1 - \exp(-\alpha t), \tag{3.135}$$

so that the closed time distribution is $\alpha \exp(-\alpha t)$.

Similarly, the probability that the amount of time spent in the open state is less than t is $1 - \exp(-(\beta + \gamma)t)$, so that the open time distribution is $(\beta + \gamma)\exp(-(\beta + \gamma)t)$.

According to this model, the channel can inactivate from the closed state at rate δ or from the open state at rate γ. At the beginning of the experiment the channel is in the closed state, and from there it can either open or inactivate. The probability that the first transition is to the open state is $A = \frac{\alpha}{\alpha+\delta}$, and the probability that the first transition is to the inactivated state is $1 - A$. Thus, $1 - A$ can be estimated by the proportion of experimental records in which no current is observed, even after the depolarizing stimulus was maintained for a long time.

A channel may open and close several times before it finally inactivates. To understand this, we let N be the number of times the channel opens before it finally inactivates and calculate the probability distribution for N. Clearly, $P[N = 0] = 1 - A$. Furthermore,

$$\begin{aligned} P[N = k] &= P[N = k \text{ and channel enters I from O}] \\ &\quad + P[N = k \text{ and channel enters I from C}] \\ &= A^k B^{k-1}(1 - B) + A^k B^k (1 - A) \\ &= (AB)^k \left(\frac{1 - AB}{B} \right), \end{aligned} \tag{3.136}$$

where $B = \frac{\beta}{\beta+\gamma}$.

We now have enough information to estimate the four channel rate constants. Since A can be determined from the proportion of channels that never open, B can be determined from a plot of the experimental data for $P[N = k]$ vs. k. Then, $\beta + \gamma$ can be determined from the open time distribution of the channel and α can be determined from the closed time distribution.

Since the work of Hodgkin and Huxley (described in Chapter 5), the traditional view of a Na^+ channel has been that it activates quickly and inactivates slowly. According to this view, the decreasing portion of the g_{Na} curve in Fig. 3.10 is due entirely to inactivation of the channel. However, single-channel analysis has shown that this interpretation of macroscopic data is not always correct. It turns out that the rate of inactivation of some mammalian Na^+ channels is faster than the rate of activation. For example, Aldrich et al. (1983) found $\alpha = 1$/ms, $\beta = 0.4$/ms, $\gamma = 1.6$/ms, and $\delta = 1$/ms at $V = 0$ for channels in a neuroblastoma cell line and a pituitary cell line. Although this reversal of activation and inactivation rates is not correct for all Na^+ channels in all species, the result does overturn some traditional ideas of how Na^+ channels work.

More modern models of the Na^+ channel are based on a wide range of experimental data, including single-channel recordings and macroscopic ionic and gating currents. It is a very difficult matter to decide, on the basis of these data, which is the best model of the channel. One of the most rigorous attempts is that of Vandenberg and Bezanilla (1991), who concluded that a sequential Markov model with three closed states, one open state, and one inactivated state was the best at reproducing the widest array of data. However, because of the ill-posed nature of this inverse problem, it is impossible to rule out the existence of multiple other states.

3.6.2 Single-Channel Analysis of an Agonist-Controlled Ion Channel

The single-channel analysis of the agonist-controlled ion channel (3.127) is more subtle than that of the Na^+ channel. This is because there are states that cannot be directly observed, but can only be inferred from the data. If the state AR is the only open state, then a typical single-channel recording might look (at least, in an ideal situation) like that shown in Fig. 3.14. The openings occur in bursts as the channel flicks between the AR and AT states, with longer interburst periods occurring when the channel escapes from AT into the closed state, T, because of agonist unbinding. However, because the binding and unbinding transitions are not directly observable, this process is called a *hidden Markov process*.

There are two distributions that are readily determined from the data. These are the open time and closed time distributions. Since the open state, AR, can close only by a single pathway to AT, the open time distribution is the exponential distribution $\alpha e^{-\alpha t}$, with mean $1/\alpha$.

The Closed Time Distribution

Every period during which the channel is closed must begin with the channel in state AT and end with the channel in state AR. However, during this time the channel can be in either state AT or state T. Thus, the closed time distribution is the transition time from state AT to state AR (see Section 2.9.2).

To calculate the closed time probability, we set state AR to be an absorbing state (i.e., set $\alpha = 0$), and impose the initial condition that the receptor starts in state AT. Hence,

$$\frac{d\Phi_1}{dt} = \beta\Phi_2, \tag{3.137}$$

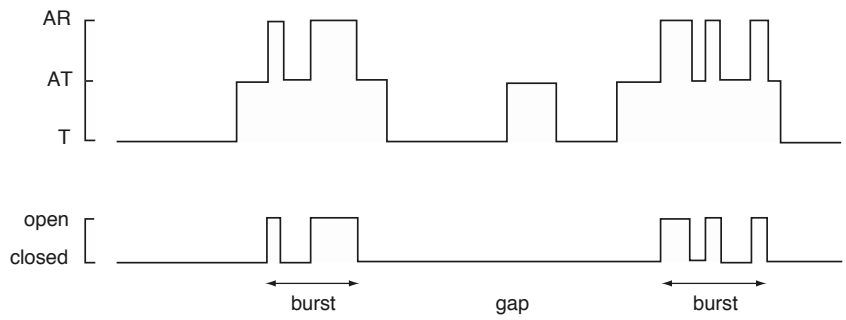


Figure 3.14 Schematic diagram of a possible single-channel recording in the model described by (3.127). The openings occur in bursts as the channel flickers between states AR and AT. However, the transitions between states T and AT cannot be observed. (Adapted from Colquhoun and Hawkes (1981), Fig. 1.)

$$\frac{d\Phi_2}{dt} = -(\beta + k_{-1})\Phi_2 + ak_1\Phi_3, \tag{3.138}$$

$$\frac{d\Phi_3}{dt} = k_{-1}\Phi_2 - ak_1\Phi_3, \tag{3.139}$$

with initial data $\Phi_1(0) = 0$, $\Phi_2(0) = 1$, $\Phi_3(0) = 0$.

We can readily solve this system of differential equations to determine that the transition time from state 2 to state 1 has the probability density

$$\phi_{21}(t) = \frac{d\Phi_1}{dt} = \beta x_2 = \frac{\beta}{\lambda_1 - \lambda_2}[(\lambda_1 + ak_1)e^{\lambda_1 t} - (\lambda_2 + ak_1)e^{\lambda_2 t}], \tag{3.140}$$

where the eigenvalues λ_1 and λ_2 are the roots (both negative) of $\lambda^2 + \lambda(\beta + k_{-1} + k_1 a) + ak_1\beta = 0$.

Since the closed time distribution is the sum of two exponentials, the open and closed time distributions along with the steady-state open probability (3.130) theoretically provide enough information to determine uniquely the four kinetic parameters of the model.

Other Distributions

There are other distributions that can be calculated, but obtaining the data for these is somewhat subjective. These are the distribution of closed times during a burst, the number of openings in a burst, and the distribution of gap closed times.

The distribution of closed times during a burst is the easiest to calculate, being simply $\beta \exp(-\beta t)$.

Each time the channel is in state AT a choice is made to go to state AR (with probability $\beta/(\beta + k_{-1})$) or to go to state T (with probability $k_{-1}/(\beta + k_{-1})$). Thus, for there to be N openings in a burst, the channel must reopen by going from state AT to state AR $N - 1$ times, and then end the burst by going from state AT to state T. Hence,

$$P(N \text{ openings}) = \left(\frac{\beta}{\beta + k_{-1}}\right)^{N-1}\left(\frac{k_{-1}}{\beta + k_{-1}}\right), \tag{3.141}$$

where $N \geq 1$, which has mean $1 + \beta/k_{-1}$.

To determine the closed time distribution for gaps, we observe that a gap begins in state AT and then moves back and forth between states T and AT before finally exiting into state AR. The waiting-time distribution for leaving state AT into state T is

$$\phi_{23}(t) = k_{-1}e^{-k_{-1}t}. \tag{3.142}$$

The transition time from state T to state AR is $\phi_{31} = \frac{d\Phi_1}{dt}$, determined as the solution of the system (3.137)–(3.139), subject to initial conditions $\Phi_1(0) = 0$, $\Phi_2(0) = 0$, and $\Phi_3(0) = 1$. One readily determines that

$$\phi_{31} = \beta\frac{(\lambda_2 + ak_1)(\lambda_1 + ak_1)}{k_{-1}(\lambda_2 - \lambda_1)}(e^{\lambda_1 t} - e^{\lambda_2 t}). \tag{3.143}$$

Now, to calculate the gap time distribution, we observe that the time in a gap is the sum of two times, namely the time in state AT before going to state T and the time to go from state T to state AR. Here we invoke a standard result from probability theory regarding the distribution for the sum of two random variables. That is, if $p_1(t_1)$ and $p_2(t_2)$ are the probability densities for random variables t_1 and t_2, then the probability density for the sum of these is the convolution

$$p_{\text{sum}}(t) = \int_0^t p_1(s)p_2(t-s)\,ds. \tag{3.144}$$

Thus, in this problem, the probability density for the gap time is

$$\phi_{\text{gap}}(t) = \int_0^t \phi_{23}(s)\phi_{31}(t-s)\,ds. \tag{3.145}$$

It is again straightforward to determine (use Laplace transforms and the convolution theorem) that

$$\phi_{\text{gap}}(t) = \beta\frac{(\lambda_2 + ak_1)(\lambda_1 + ak_1)}{(\lambda_2 - \lambda_1)}\left(\frac{e^{\lambda_1 t}}{\lambda_1 + k_{-1}} - \frac{e^{\lambda_2 t}}{\lambda_2 + k_{-1}} + \frac{(\lambda_1 - \lambda_2)e^{-k_{-1}t}}{(\lambda_1 + k_{-1})(\lambda_2 + k_{-1})}\right),$$

a sum of three exponentials.

Making use of these distributions is tricky, because it is not clear how to distinguish between a short gap and a long closed interval during a burst. In fact, if a is large enough and k_{-1} is not too small, then the mean gap length is shorter than the mean burst closed time, so that errors of classification are likely.

3.6.3 Comparing to Experimental Data

Experimental data typically come in lists of open and closed times. They are then displayed in a histogram, where the area under each histogram bar corresponds to the number of events in that interval.

However, it can be very difficult to determine from a histogram how many exponential components are in the distributions; an exponential distribution with three exponentials can look very similar to one with two exponentials, even when the time constants are widely separated.

This problem is sometimes avoided by first taking the log of the times, and plotting a histogram of the log(time) distributions. Since the log function is monotone increasing, we know that

$$P[t < t_1] = P[\log(t) < \log(t_1)], \tag{3.146}$$

and thus the cumulative distributions are the same, whether functions of the log or the linear times. However, the functions that are of interest to us are the probability density functions, which are the derivatives of the cumulative distribution functions. Suppose $\Phi(t)$ is a cumulative distribution function with corresponding probability density function $\phi(t) = \frac{d\Phi}{dt}$. For any monotone increasing function $g(x)$, $\Phi(g(x))$ is also a cumulative

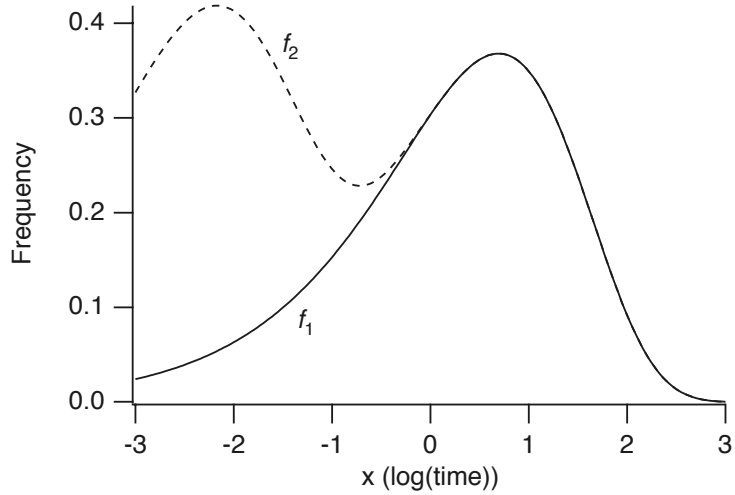


Figure 3.15 Two exponential distributions, transformed according to (3.147), and plotted against log(time). Solid line: $f_1(t) = 0.5e^{-0.5t}$. Dotted line: $f_2(t) = 0.5e^{-0.5t} + 10e^{-10t}$.

distribution function in x. However, the corresponding probability density function is

$$\frac{d\Phi(g(x))}{dx} = g'(x)\phi(g(x)). \tag{3.147}$$

Thus, to use a log(time) transformation, $x = \ln(t)$, for a given probability density function, $\phi(t)$, we plot $e^x\phi(e^x)$ and fit this to the histogram of the log(time).

There are significant advantages to this scaling of time, especially for exponential distributions, illustrated in Fig. 3.15. The solid line in Fig. 3.15 corresponds to the distribution $\phi_1(t) = 0.5e^{-0.5t}$, transformed according to (3.147). In other words, this is the plot of the function $0.5e^xe^{-0.5e^x}$ against x. Notice that the maximum of the curve occurs at the mean of the distribution, $t = 2$ ($x = \log 2 = 0.69$) (see Exercise 21). The dotted line is the distribution $f_2(t) = 0.5e^{-0.5t} + 10e^{-10t}$, again transformed according to (3.147). (This is not a true probability density function since the area under the curve is 2, rather than 1.) The two peaks occur at the means of the individual component exponential distributions, i.e., at $t = 2$ and $t = 0.1$ ($x = 0.69$ and $x = -2.3$ respectively).

Modern methods of fitting models to single-channel data are considerably more sophisticated than merely fitting histograms, as described above. Fitting directly to the set of open and closed times using the log likelihood is a common approach, but, more recently, methods to fit the model directly to the single-channel time course raw data (not simply to a list of open and closed times), using Markov chain Monte Carlo and Bayesian inference, have been developed (Fredkin and Rice, 1992; Ball et al., 1999; Hodgson and Green, 1999).

3.7 Appendix: Reaction Rates

In Section 3.4 we made extensive use of the formula (3.43), i.e.,

$$k_i = \kappa \exp\left(\frac{-\Delta G_i}{RT}\right), \tag{3.148}$$

which states that the rate, k_i, at which a molecule leaves a binding site is proportional to the exponential of the height of the energy barrier ΔG_i that must be crossed to exit. This is called the *Arrhenius* equation, after Svante Arrhenius, who first discovered it experimentally in the late 1800s (Arrhenius received the 1903 Nobel Prize in Chemistry). Arrhenius determined, not the dependence of k_i on ΔG_i, but its dependence on temperature T. He showed experimentally that the rate of reaction is proportional to $\exp(-B/T)$, for some positive constant B. He then used the Boltzmann distribution to argue that $B = \Delta G_i/R$, as discussed below.

As was described in Section 1.2, the equilibrium constant, K_{eq}, for a reaction is related to the change in free energy, ΔG^0, by

$$K_{\text{eq}} = e^{\frac{\Delta G^0}{RT}}. \tag{3.149}$$

Note that if κ is independent of ΔG_i, then (3.148) is consistent with (3.149). Given the potential energy profile in Fig. 3.16, it is clear that if $k_1 = \kappa \exp\left(\frac{-\Delta G_1}{RT}\right)$ and $k_{-1} = \kappa \exp\left(\frac{-\Delta G_{-1}}{RT}\right)$, then $K_{\text{eq}} = k_{-1}/k_1 = \exp\left(\frac{\Delta G_1 - \Delta G_{-1}}{RT}\right) = \exp\left(\frac{\Delta G^0}{RT}\right)$, where $\Delta G^0 = \Delta G_1 - \Delta G_{-1}$.

However, despite this consistency, (3.148) does not follow from (3.149); the equilibrium relationship tells us nothing about how each rate constant might depend on T or ΔG_i. Although the derivation of the equilibrium condition depends only on fundamental thermodynamical principles, derivation of a rate expression is much more difficult, and the exact rate expression depends, in general, on the choice of model. There is still enormous controversy over exactly how to derive rate equations, and which is most

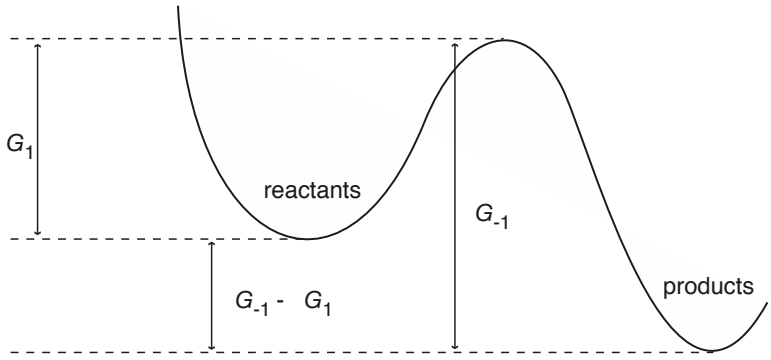


Figure 3.16 Schematic diagram of a potential energy profile of a hypothetical reaction.

suitable in any given situation. Here we give only a brief discussion of this important problem, enough to give a plausible derivation of the general form of (3.148). The exponential dependence occurs in every rate equation; it is the prefactor, κ, and its possible functional dependences, that is the source of so much discussion.

3.7.1 The Boltzmann Distribution

To show that $B = \Delta G/R$ (or $\Delta G/k$, depending on whether ΔG has units of per mole or per molecule; we have dropped the subscript i for this discussion), Arrhenius assumed that the rate of reaction was proportional to the fraction of molecules with energy greater than some minimum amount. Given this assumption, the Arrhenius equation follows from the Boltzmann distribution, which we now derive.

We begin with a brief digression. It is intuitively clear that were we to toss a fair coin 10^{20} times, the chance of obtaining any distribution of heads and tails significantly different from 50:50 is insignificant. We might, of course, get $\frac{10^{20}}{2} - 100$ heads and $\frac{10^{20}}{2} + 100$ tails, but the relative deviation from 50:50 is inconsequential.

To express this mathematically, suppose that we toss a fair coin n times to get a sequence of heads and tails. Of all the possible sequences, the total number that have h heads and $n - h$ tails (in any order) is $\frac{n!}{h!(n-h)!}$, and thus, since there are 2^n possible sequences, the probability of getting h heads and $n - h$ tails is given by

$$\text{Prob}[n \text{ heads}, n - h \text{ tails}] = \frac{n!}{2^n h!(n-h)!}. \tag{3.150}$$

As n gets very large, the graph of (3.150) becomes sharply peaked, with a maximum of 1 at $h = n/2$. (This can be shown easily using Stirling's formula, $\ln(n!) \approx n\ln(n) - n$ for n large). In other words, the probability of obtaining any sequence that does not contain an equal number of heads and tails is vanishingly small in the limit of large n.

An identical argument underlies the Boltzmann distribution. Suppose we have n particles each of which can be in one of k states, where state i has energy U_i. Let n_i denote the number of particles in state i. We assume that the total energy, U_{tot}, is fixed, so that

$$\sum_{i=1}^{k} n_i = n, \tag{3.151}$$

$$\sum_{i=1}^{k} U_i n_i = U_{\text{tot}}. \tag{3.152}$$

The number of ways, W, that these n particles can be partitioned into k states, with n_i particles in the state i, is given by the multinomial factor

$$W = \frac{n!}{\Pi_{i=1}^{k} n_i!}. \tag{3.153}$$

Now, like the function for the probability of heads and tails for a coin toss, the function W is sharply peaked when n is large, and the distribution corresponding to the peak is the one most likely to occur. Furthermore, the likelihood of distributions other than those near the peak is vanishingly small when n is large. Thus, to find this overwhelmingly most likely distribution we maximize W subject to the constraints (3.151) and (3.152). That is, we seek to maximize (using Lagrange multipliers)

$$F = \ln W - \lambda \left(\sum_{i=1}^{k} n_i - n \right) - \beta \left(\sum_{i=1}^{k} U_i n_i - U_{\text{tot}} \right). \tag{3.154}$$

(It is equivalent, and much more convenient, to use $\ln W$ rather than W.) According to (3.153),

$$\begin{aligned} \ln W &= \ln(n!) - \sum_{i=1}^{k} \ln(n_i!) \\ &\approx n \ln(n) - n + \sum_{i=1}^{k} n_i - \sum_{i=1}^{k} n_i \ln(n_i) \\ &= n \ln(n) - \sum_{i=1}^{k} n_i \ln(n_i), \end{aligned} \tag{3.155}$$

where we have used Stirling's formula, assuming all the n_i's are large. Thus,

$$\frac{\partial F}{\partial n_i} = -\ln(n_i) - 1 - \lambda - \beta U_i, \tag{3.156}$$

which is zero when

$$n_i = \alpha e^{-\beta U_i}, \tag{3.157}$$

for positive constants α and β, which are independent of i. This most likely distribution of n_i is the *Boltzmann distribution*.

How does this relate to reaction rates? Suppose that we have a population of particles with two energy levels: a ground energy level U_0 and a reactive energy level $U_r > U_0$. If the particles are at statistical equilibrium, i.e., the Boltzmann distribution, then the proportion of particles in the reactive state is

$$e^{\beta(U_0 - U_r)} = e^{-\beta \Delta U}. \tag{3.158}$$

Since this is also assumed to be the rate at which the reaction takes place, we have that

$$k \propto e^{\beta(U_0 - U_r)} = e^{-\beta \Delta U}. \tag{3.159}$$

To obtain the Arrhenius rate equation it remains to show that $\beta = \frac{1}{RT}$. To do so rigorously is beyond the scope of this text, but a simple dimensional argument can at least demonstrate plausibility. Recall that it is known from experiment that

$$k \propto e^{-B/T}, \tag{3.160}$$

for some constant $B > 0$. Thus, from (3.159), β must be proportional to $1/T$, and to get the correct units, $\beta \propto 1/(RT)$ or $1/(kT)$, depending on whether the units of U are per mole or per molecule.

3.7.2 A Fokker–Planck Equation Approach

The above derivation relies on the assumption that there is a large number of particles, each of which can be in one of a number of different states. It is less obvious how such a derivation can be applied to the behavior of a small number of molecules, or a single molecule. To do this, we need to use the methods developed in Appendix 2.9, and turn to a Fokker–Planck description of molecular motion.

Suppose that a molecule moves via Brownian motion, but also experiences a force generated by some potential, $U(x)$, and is subject to friction. If $x(t)$ denotes the position of the molecule, the Langevin equation for the molecular motion (Section 2.9.5) is

$$m\frac{d^2x}{dt^2} + \nu\frac{dx}{dt} + U'(x) = \sqrt{2\nu kT}W(t), \tag{3.161}$$

where W is a Wiener process. Here, ν is the friction coefficient, and is analogous to friction acting on a mass–spring system. If inertial effects can be neglected, which they can in most physiological situations, this simplifies to

$$\nu\frac{dx}{dt} = -U'(x) + \sqrt{2\nu kT}W(t). \tag{3.162}$$

Hence, the probability distribution that the particle is at position x at time t is given by $p(x,t)$, the solution of the Fokker–Planck equation

$$\nu\frac{\partial p}{\partial t} = \frac{\partial}{\partial x}(U'(x)p) + kT\frac{\partial^2 p}{\partial x^2}. \tag{3.163}$$

At steady state, i.e., when $\frac{\partial p}{\partial t} = 0$, (3.163) can be readily solved to give

$$p(x) = \frac{1}{A}\exp\left(-\frac{U(x)}{kT}\right), \tag{3.164}$$

where $A = \int_{-\infty}^{\infty}\exp(-\frac{U(x)}{kT})\,dx$ is chosen so that $\int_{-\infty}^{\infty}p(x)\,dx = 1$. We have thus regained a continuous version of the Boltzmann distribution; if $U(x)$ is a quadratic potential well ($U(x) = Ax^2$), then $p(x)$ is a Gaussian distribution.

If $U(x)$ is a double well potential with its maximum at $x = 0$ separating the two wells, then the ratio of the probability of finding the particle on the left to the probability of finding the particle on the right is

$$K_{\text{eq}} = \frac{\int_{-\infty}^{0}p(x)\,dx}{\int_{0}^{\infty}p(x)\,dx}. \tag{3.165}$$

Since it is difficult to calculate K_{eq} for general functions $U(x)$, it is useful to discretize the state space into a finite number of states $j = 1, 2, \ldots, n$ with energies U_j. For

this, we know that the Boltzmann distribution is

$$p_j = \frac{1}{A} \exp\left(-\frac{U_j}{kT}\right), \tag{3.166}$$

where

$$A = \sum_j^n \exp\left(-\frac{U_j}{kT}\right). \tag{3.167}$$

We can make the association between the discrete case and the continuous case precise if we determine U_j by requiring

$$\exp\left(-\frac{U_j}{kT}\right) = \int_{x_{j-1}}^{x_j} \exp\left(-\frac{U(x)}{kT}\right) dx, \tag{3.168}$$

where x_j separates the $j-1$st from the jth potential well. Furthermore, if there are only two energy wells, the ratio of the probability of finding the particle in state one to the probability of finding the particle in state two is

$$K_{\text{eq}} = \frac{p_1}{p_2} = \exp\left(\frac{\Delta U}{kT}\right) = \exp\left(\frac{\Delta G^0}{RT}\right), \tag{3.169}$$

where $\Delta U = U_2 - U_1$ is the change in energy per molecule, so that ΔG^0 is the change in energy per mole. Here we have recovered (3.149) for the equilibrium distribution of a reaction in terms of the difference of standard free energy. However, one should note that with this identification, U_j is approximately, but not exactly, the value of U at the bottom of the jth potential well.

3.7.3 Reaction Rates and Kramers' Result

As noted in Chapter 1, equilibrium relationships give information only about the ratio of rate constants, not their individual values. To derive an expression for a rate constant, one must construct a model of how the reaction occurs. The consequent expression for the rate constant is only as good as the assumptions underlying the model.

One common model of a reaction rate is based on the mean first exit time of the time-dependent Fokker–Planck equation (3.163). (Mean first exit times are discussed in Section 2.9.6). This model assumes that a reactant particle can be modeled as a damped oscillator driven by a stochastic force, and that the reaction occurs once the particle reaches the peak of the energy profile between the reactant and product states. Although this model is based on a number of crude assumptions, it gives reasonably good results for a range of potential energy profiles, particularly those for which the energy wells are deep.

The mean first exit time is found from the solution of the ordinary differential equation

$$-U'(x)\frac{d\tau}{dx} + kT\frac{d^2\tau}{dx^2} = -\nu, \tag{3.170}$$

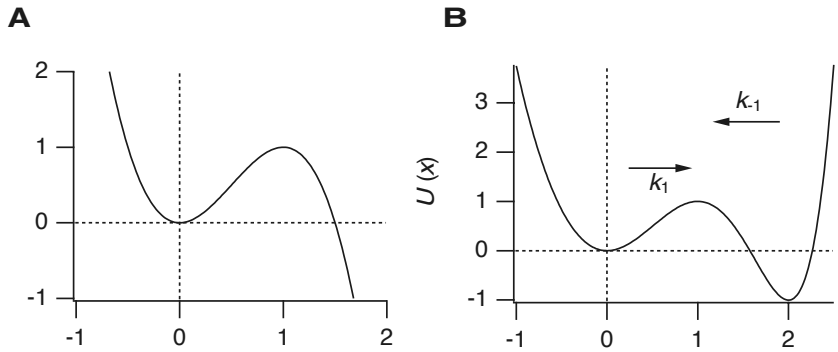


Figure 3.17 Potential energy profiles. A: a cubic profile, $U(x) = \Delta G(2x^2(3/2 - x))$, plotted for $\Delta G = 1$. B: a double well potential, $U(x) = \Delta G^0(\frac{19}{144}x^6 - \frac{1}{24}x^5 - \frac{77}{144}x^4 - \frac{4}{3}x^3 + \frac{25}{9}x^2)$, plotted for $\Delta G^0 = 1$.

subject to $\tau(x_0) = 0$ at any boundary point x_0 where escape is allowed, or $\tau'(x_1) = 0$ at any boundary point x_1 where escape is not allowed, but instead there is reflection. The off-rate, or unbinding rate, is defined as the inverse of the mean first exit time from the bottom of the potential well.

To be specific, consider a potential $U(x)$ such as shown in Fig. 3.17A. Here $U(x)$ is a cubic polynomial, with a minimum at $x = 0$ and a maximum at $x = 1$, with $U(1) = \Delta G$. We expect the particle to spend most of its time near $x = 0$. However, if the particle gets to $x = 1$ it can escape to $x = \infty$, and is assumed to have reacted. Thus, the time to react (the inverse of the reaction rate) is approximated by the mean first passage time from $x = 0$ to $x = 1$.

More generally, suppose $U(x) = \Delta G u(\frac{x}{L})$, where $u'(0) = u'(1) = u(0) = 0$ and $u(1) = 1$, so that $x = 0$ is a local minimum and $x = L$ is a local maximum, and the height of the energy barrier is ΔG. The mean first passage time is the solution of (3.170) together with the boundary conditions $\tau(-\infty) = 0$ and $\tau(L) = 0$.

To find the solution it is useful to nondimensionalize (3.170). We set $y = \frac{x}{L}$ and $\sigma = \alpha\tau$ and obtain

$$-au'(y)\frac{d\sigma}{dy} + \frac{d^2\sigma}{dy^2} = -1, \tag{3.171}$$

where $a = \frac{\Delta G}{kT}$ and $\alpha = \frac{\nu L^2}{kT}$. Using an integrating factor, it is easily shown that

$$\sigma(y) = \int_x^1 e^{au(s')} \left(\int_{-\infty}^{s'} e^{-au(s)}\, ds \right) ds', \tag{3.172}$$

and thus the time to react is $\tau(0) = \frac{\nu L^2}{kT}\sigma(0)$, where

$$\sigma(0) = \int_0^1 e^{au(s')} \left(\int_{-\infty}^{s'} e^{-au(s)}\, ds \right) ds'. \tag{3.173}$$

As we demonstrate below, this formula does not agree with the Arrhenius rate law for all parameter values. However, when the potential well at $x = 0$ is deep (i.e., when $a = \Delta G/(kT) \gg 1$), the two are in agreement. Here we provide a demonstration of this agreement.

Notice first that

$$\sigma(0) = \int_0^1 e^{au(s')} \left(\int_{-\infty}^1 e^{-au(s)}\, ds - \int_{s'}^1 e^{-au(s)}\, ds \right) ds'. \tag{3.174}$$

Clearly,

$$\begin{aligned} \int_0^1 e^{au(s')} \left(\int_{s'}^1 e^{-au(s)}\, ds \right) ds' &= \int_0^1 \left(\int_{s'}^1 e^{a(u(s')-u(s))}\, ds \right) ds' \\ &< \int_0^1 \left(\int_{s'}^1 ds \right) ds' = \frac{1}{2}. \end{aligned} \tag{3.175}$$

In fact, with a bit of work one can show that this integral approaches zero as $a \to \infty$. Thus,

$$\sigma(0) \approx \left(\int_0^1 e^{au(s')}\, ds' \right) \left(\int_{-\infty}^1 e^{-au(s)}\, ds \right). \tag{3.176}$$

We now use the fact that $y = 0$ and $y = 1$ are extremal values of $u(y)$ to approximate these integrals. When a is large, the integrands are well approximated by Gaussians, which decay to zero rapidly. Thus, near $y = 0$, $u(y) \approx \frac{1}{2}u''(0)y^2$, so that

$$\begin{aligned} \int_{-\infty}^1 e^{-au(s)}\, ds &\approx \int_{-\infty}^1 e^{-\frac{1}{2}au''(0)s^2}\, ds \\ &\approx \int_{-\infty}^{\infty} e^{-\frac{1}{2}au''(0)s^2}\, ds \\ &= \sqrt{\frac{2\pi}{au''(0)}}. \end{aligned} \tag{3.177}$$

Similarly, near $y = 1$, $u(y) \approx 1 - \frac{1}{2}|u''(1)|(y-1)^2$, so that

$$\begin{aligned} \int_0^1 e^{au(s)}\, ds &\approx e^a \int_0^1 e^{-\frac{1}{2}|u''(1)|(s-1)^2}\, ds \\ &\approx e^a \int_{-\infty}^0 e^{-\frac{1}{2}|u''(1)|s^2}\, ds \\ &= \frac{1}{2} e^a \sqrt{\frac{2\pi}{a|u''(1)|}}. \end{aligned} \tag{3.178}$$

Combining (3.176), (3.177), and (3.178) gives

$$\tau(0) \approx \frac{\pi \nu L^2}{\Delta G \sqrt{u''(0)|u''(1)|}} e^{\frac{\Delta G}{kT}}. \tag{3.179}$$

Since the reaction rate is the inverse of the mean first passage time, this gives the Arrhenius rate expression with

$$\kappa = \frac{\Delta G\sqrt{u''(0)|u''(1)|}}{\pi \nu L^2}, \tag{3.180}$$

which is independent of T, but not ΔG. This formula was first derived by Kramers (1940).

A Double Well Potential Profile

Now suppose that $U(x)$ is a double well potential, such as that shown in Fig. 3.17B. In particular, suppose that $U(x) = \Delta G^0 u(\frac{x}{L})$, where $u(x)$ has two local minima at $x = 0$ and $x = b > 1$, with a local maximum at $x = 1$. For the example in Fig. 3.17B, $\Delta G^0 = L = 1$ and $b = 2$. Note also that the potential profile is such that $\Delta G_{-1} = 2\Delta G^0$, $\Delta G_1 = \Delta G^0$.

According to Kramers' rate theory,

$$k_1 \approx \frac{\Delta G_1\sqrt{u''(0)|u''(1)|}}{\pi \nu L^2} e^{-\frac{\Delta G_1}{kT}}, \tag{3.181}$$

$$k_{-1} \approx \frac{\Delta G_{-1}\sqrt{u''(b)|u''(1)|}}{\pi \nu L^2 (b-1)^2} e^{-\frac{\Delta G_{-1}}{kT}}. \tag{3.182}$$

To compare these with the exact solutions, in Fig. 3.18A we plot k_{-1} and k_1 for the double well potential shown Fig. 3.17B, calculated by numerical integration of (3.173), and using the approximations (3.181) and (3.182). Note that the reaction rates (both exact and approximate) are not exactly exponential functions of ΔG_i, and thus the curves in Fig. 3.18A are not straight lines (on a log scale). For the approximate rate constants this is because the prefactor is proportional to ΔG_i. Interestingly, the approximate

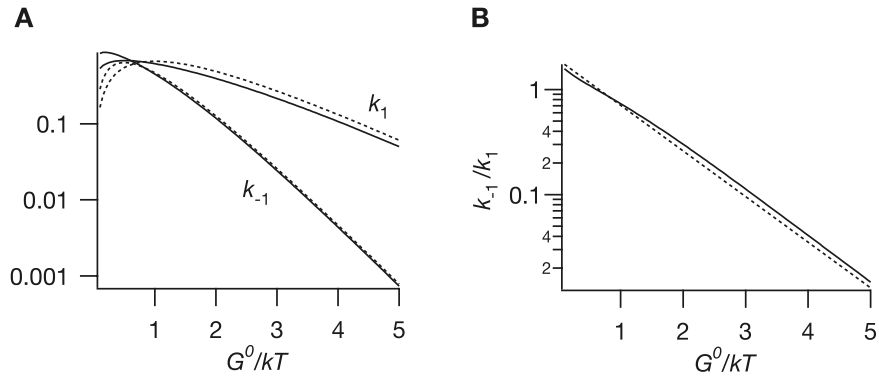


Figure 3.18 Reaction rates for the potential profile shown in Fig. 3.17B. A: exact (solid lines) and approximate (dashed lines) solutions for k_1 and k_{-1}, plotted as functions of $\Delta G^0/kT$. The exact solutions are calculated from (3.173), while the approximations are calculated from (3.179). For simplicity, we set $\nu L^2/kT = 1$. B: exact and approximate calculations of k_{-1}/k_1. As in A, the exact solution is plotted as a solid line.

solutions agree exactly with the Arrhenius rate law, when viewed as functions of T, while the exact solutions do not.

Next, we observe that, using Kramers' formula, the equilibrium constant is

$$K_{\text{eq}} = \frac{k_{-1}}{k_1} = \frac{1}{(b-1)^2}\frac{u(1)-u(b)}{u(1)-u(0)}\sqrt{\frac{u''(b)}{u''(0)}}e^{-\frac{\Delta G^0}{kT}}. \tag{3.183}$$

In Fig. 3.18B we plot $\frac{k_{-1}}{k_1}$ for the double well potential shown in Fig. 3.17B, with the exact ratio shown as a solid curve and the approximate ratio from (3.183) shown as a dashed curve. As before, the exact ratio k_{-1}/k_1 (solid line, Fig. 3.17B) is not an exact exponential function of ΔG^0, and thus does not give the correct equilibrium behavior. This results from the fact that, for small ΔG^0, the mean first exit time of the Fokker–Planck equation is not a good model of the reaction rate.

However, the ratio of the approximate expressions for the rate constants (3.183) is a true exponential function of ΔG^0, since the dependence of ΔG^0 in the prefactors cancels out in the ratio. Hence, the dashed line in Fig. 3.18B is straight. Thus, paradoxically, the approximate solution gives better agreement to the correct equilibrium behavior than does the exact solution. However, one must be somewhat cautious with this statement, since there is a factor multiplying the exponential that is not equal to one (as it should be for correct equilibrium behavior), but depends on the details of the shape of the energy function. Thus, if the shape of the potential energy function is modified by, for example, an external voltage potential, this factor is modified as well, in a voltage-dependent way.

3.8 EXERCISES

1. Show that the GHK equation (3.2) satisfies both the independence principle and the Ussing flux ratio, but that the linear I–V curve (3.1) satisfies neither.

2. Using concentrations typical of Na^+, determine whether the long channel limit or the short channel limit for (3.25) is the most appropriate approximation for Na^+ channels. (Estimate λ where $\lambda^2 = L^2 qFN_a\tilde{c}/(\epsilon RT)$, for Na^+ ions.)

3. In Section 3.3.1 the PNP equations were used to derive I–V curves when two ions with opposite valence are allowed to move through a channel. Extend this analysis by assuming that two types of ions with positive valence and one type of ion with negative valence are allowed to move through the channel. Show that in the high concentration limit, although the negative ion obeys a linear I–V curve, the two positive ions do not. Details can be found in Chen, Barcilon, and Eisenberg (1992), equations (43)–(45).

4. (a) Show that (3.66) satisfies the independence principle and the Ussing flux ratio.

   (b) Show that (3.66) can be made approximately linear by choosing g such that

   $$ng = \ln\left(\frac{c_n}{c_0}\right). \tag{3.184}$$

   Although a linear I–V curve does not satisfy the independence principle, why does this result not contradict part (a)?

5. Show that (3.86) does not satisfy the independence principle, but does obey the Ussing flux ratio.

6. Derive (3.86) by solving the steady-state equations (3.84) and (3.85). First show that

$$J = x\frac{k_0c_0 - k_{-n}c_n\pi_{n-1}}{\phi_{n-1}}. \tag{3.185}$$

Then show that

$$k_0c_0x = k_{n-1}c_{n-1}\phi_{n-1} - xk_{-n}c_n\phi_{n-2}, \tag{3.186}$$

$$k_jc_j = \frac{k_{n-1}c_{n-1}}{\pi_j}(\phi_{n-1} - \phi_{j-1}) - \frac{k_{-n}c_nx}{\pi_j}(\phi_{n-2} - \phi_{j-1}), \tag{3.187}$$

for $j = 1, \ldots, n-1$. Substitute these expressions into the conservation equation and solve for x.

7. Draw the state diagrams showing the channel states and the allowed transitions for a multi-ion model with two binding sites when the membrane is bathed with a solution containing:

   (a) Only ion S on the left and only ion S′ on the right.

   (b) Ion S on both sides and ion S′ only on the right.

   (c) Ions S and S′ on both the left and right.

   In each case write the corresponding system of linear equations that determine the steady-state ionic concentrations at the channel binding sites.

8. Using an arbitrary symmetric energy profile with two binding sites, show numerically that the Ussing flux ratio is not obeyed by a multi-ion model with two binding sites. (Note that since unidirectional fluxes must be calculated, it is necessary to treat the ions on each side of the membrane differently. Thus, an eight-state channel diagram must be used.) Hodgkin and Keynes predicted that the flux ratio is the Ussing ratio raised to the $(n+1)$st power (cf. (3.16)). How does n depend on the ionic concentrations on either side of the membrane, and on the energy profile?

9. Choose an arbitrary symmetric energy profile with two binding sites, and compare the I–V curves of the one-ion and multi-ion models. Assume that the same ionic species is present on both sides of the membrane, so that only a four-state multi-ion model is needed.

10. Suppose the Na^+ Nernst potential of a cell is 56 mV, its resting potential is −70 mV, and the extracellular Ca^{2+} concentration is 1 mM. At what intracellular Ca^{2+} concentration is the flux of a three-for-one Na^+–Ca^{2+} exchanger zero? (Use that $RT/F = 25.8$ mV at 27° C.)

11. Modify the pump–leak model of Chapter 2 to include a Ca^{2+} current and the 3-for-1 Na^+–Ca^{2+} exchanger. What effect does this modification have on the relationship between pump rate and membrane potential?

12. Because there is a net current, the Na^+–K^+ pump current must be voltage-dependent. Determine this dependence by including voltage dependence in the rates of conformational change in expression (2.100). How does voltage dependence affect the pump–leak model of Chapter 2?

13. Intestinal epithelial cells have a glucose–Na^+ symport that transports one Na^+ ion and one glucose molecule from the intestine into the cell. Model this transport process. Is the transport of glucose aided or hindered by the cell's negative membrane potential?

14. Suppose that a channel consists of k identical, independent subunits, each of which can be open or closed, and that a current can pass through the channel only if all units are open.

    (a) Let S_j denote the state in which j subunits are open. Show that the conversions between states are governed by the reaction scheme

    $$S_0 \underset{\beta}{\overset{k\alpha}{\rightleftarrows}} S_1 \ldots S_{k-1} \underset{k\beta}{\overset{\alpha}{\rightleftarrows}} S_k. \tag{3.188}$$

    (b) Derive the differential equation for x_j, the proportion of channels in state j.

    (c) By direct substitution, show that $x_j = \binom{k}{j} n^j (1-n)^{k-j}$, where $\binom{k}{j} = \frac{k!}{j!(k-j)!}$ is the *binomial coefficient*, is an invariant manifold for the system of differential equations, provided that

    $$\frac{dn}{dt} = \alpha(1-n) - \beta n. \tag{3.189}$$

15. Consider the model of the Na^+ channel shown in Fig. 3.11. Show that if α and β are large compared to γ and δ, then x_{21} is given (approximately) by

    $$x_{21} = \left(\frac{\alpha}{\alpha+\beta}\right)^2 h, \tag{3.190}$$

    $$\frac{dh}{dt} = \gamma(1-h) - \delta h, \tag{3.191}$$

    while conversely, if γ and δ are large compared to α and β, then (approximately)

    $$x_{21} = m^2 \left(\frac{\gamma}{\gamma+\delta}\right), \tag{3.192}$$

    $$\frac{dm}{dt} = \alpha(1-m) - \beta m. \tag{3.193}$$

16. Show that (3.122) has two negative real roots. Show that when $\beta = 0$ and $a \leq \frac{-\lambda_1}{\lambda_1 - \lambda_2}$, then (3.123)–(3.125) have two possible solutions, one with $\alpha + \delta = -\lambda_1$, $\gamma = -\lambda_2$, the other with $\alpha + \delta = -\lambda_2$, $\gamma = -\lambda_1$. In the first solution inactivation is faster than activation, while the reverse is true for the second solution.

17. Write a computer program to simulate the behavior of the stochastic three-state Na^+ channel shown in Fig. 3.12, assuming it starts in the closed state. Use $\alpha = 1$/ms, $\beta = 0.4$/ms, $\gamma = 1.6$/ms and $\delta = 1$/ms. Take the ensemble average of many runs to reproduce its macroscopic behavior. Using the data from simulations, reconstruct the open time distribution, the latency distribution, and the distribution of N, the number of times the channel opens. From these distributions estimate the rate constants of the simulation and compare with the known values.

18. Consider the Markov model of a Na^+ channel (Patlak, 1991) shown in Fig. 3.19. Write a computer program to simulate the behavior of this stochastic channel assuming it starts in state C_1. Take the ensemble average of many runs to reproduce its macroscopic behavior. Using the data generated by these simulations, determine the open time distribution, the latency distribution, and the distribution of N, the number of times the channel opens. Compare these with the analytically calculated distributions.

19. Construct a set of differential equations that models the interaction of a two-state channel with a use-dependent blocker.

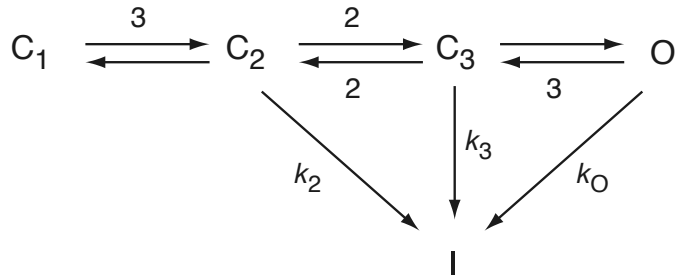


Figure 3.19 The Na^+ channel model of Exercise 18. Parameter values are $k_2 = 0.24\ \text{ms}^{-1}$, $k_3 = 0.4\ \text{ms}^{-1}$, $k_O = 1.5\ \text{ms}^{-1}$, $\alpha = 1\ \text{ms}^{-1}$, and $\beta = 0.4\ \text{ms}^{-1}$.

20. Write a computer program to simulate the behavior of the stochastic three-state agonist-binding channel of (3.127) Use $\alpha = 1/\text{ms}$, $\beta = 0.4/\text{ms}$, $k_{-1} = 0.5/\text{ms}$, and $ak_1 = 0.2/\text{ms}$. Using the data from simulations, plot the open time distribution and the closed time distribution and estimate the parameters of the model. Is the closed time distribution obviously a double exponential distribution? Repeat this experiment for several different values of ak_1.

21. Show that for an exponential distribution $\phi(t) = \alpha \exp(-\alpha t)$ the plot of the corresponding distribution function on the $\ln(t)$ scale has a maximum at the expected value of the distribution, $t = -\frac{1}{a}$.

22. Find the distribution for the length of a burst in the model of (3.127).
Hint: The apparent length of the burst is the time taken to get from AR to T, minus the length of one sojourn in AT. Use the Laplace transform and the convolution theorem.

23. Find the mean first exit time from the piecewise-linear potential

$$U(x) = \begin{cases} -\frac{\Delta G x}{L}, & -L < x < 0, \\ \frac{\Delta G x}{L}, & 0 < x < L, \end{cases} \tag{3.194}$$

with a reflecting boundary at $x = -L$ and absorbing boundary at $x = L$.

24. Find the most likely event for the binomial distribution

$$P(h) = \frac{n!}{h!(n-h)!} p^h (1-p)^{(n-h)} \tag{3.195}$$

when n is large. Show that the probability of this event approaches 1 in the limit $n \to \infty$.