Part I

Cellular Physiology

Chapter 1

Biochemical Reactions

Cells can do lots of wonderful things. Individually they can move, contract, excrete, reproduce, signal or respond to signals, and carry out the energy transactions necessary for this activity. Collectively they perform all of the numerous functions of any living organism necessary to sustain life. Yet all of what cells do can be described in terms of a few basic natural laws. The fascination with cells is that although the rules of behavior are relatively simple, they are applied to an enormously complex network of interacting chemicals and substrates. The effort of many lifetimes has been consumed in unraveling just a few of these reaction schemes, and there are many more mysteries yet to be uncovered.

1.1 The Law of Mass Action

The fundamental "law" of a chemical reaction is the law of mass action. This "law" describes the rate at which chemicals, whether large macromolecules or simple ions, collide and interact to form different chemical combinations. Suppose that two chemicals, say A and B, react upon collision with each other to form product C,

$$\mathrm{A} + \mathrm{B} \xrightarrow{k} \mathrm{C}. \tag{1.1}$$

The rate of this reaction is the rate of accumulation of product, $\frac{d[\mathrm{C}]}{dt}$. This rate is the product of the number of collisions per unit time between the two reactants and the probability that a collision is sufficiently energetic to overcome the free energy of activation of the reaction. The number of collisions per unit time is taken to be proportional to the product of the concentrations of A and B with a factor of proportionality that depends on the geometrical shapes and sizes of the reactant molecules and on the temperature of the mixture. Combining these factors we have

$$\frac{d[\mathrm{C}]}{dt} = k[\mathrm{A}][\mathrm{B}]. \tag{1.2}$$

The identification of (1.2) with the reaction (1.1) is called the *law of mass action*, and the constant k is called the *rate constant* for the reaction. However, the law

of mass action is not a law in the sense that it is inviolable, but rather it is a useful model, much like Ohm's law or Newton's law of cooling. As a model, there are situations where it is not valid. For example, at high concentrations, doubling the concentration of one reactant need not double the overall reaction rate, and at extremely low concentrations, it may not be appropriate to represent concentration as a continuous variable.

For thermodynamic reasons all reactions proceed in both directions. Thus, the reaction scheme for A, B, and C should have been written as

$$\mathrm{A} + \mathrm{B} \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \mathrm{C}, \tag{1.3}$$

with k_+ and k_- denoting, respectively, the forward and reverse rate constants of reaction. If the reverse reaction is slow compared to the forward reaction, it is often ignored, and only the primary direction is displayed. Since the quantity A is consumed by the forward reaction and produced by the reverse reaction, the rate of change of [A] for this bidirectional reaction is

$$\frac{d[\mathrm{A}]}{dt} = k_-[\mathrm{C}] - k_+[\mathrm{A}][\mathrm{B}]. \tag{1.4}$$

At equilibrium, concentrations are not changing, so that

$$\frac{k_-}{k_+} \equiv K_{\mathrm{eq}} = \frac{[\mathrm{A}]_{\mathrm{eq}}[\mathrm{B}]_{\mathrm{eq}}}{[\mathrm{C}]_{\mathrm{eq}}}. \tag{1.5}$$

The ratio k_-/k_+, denoted by K_{eq}, is called the *equilibrium constant* of the reaction. It describes the relative preference for the chemicals to be in the combined state C compared to the disassociated state. If K_{eq} is small, then at steady state most of A and B are combined to give C. It should be no surprise to learn that there is a close relationship between the equilibrium constant and the free energy of the reaction, and this is discussed more in the following section.

If there are no other reactions involving A and C, then $[\mathrm{A}]+[\mathrm{C}] = \mathrm{A}_0$ is constant, and

$$[\mathrm{C}]_{\mathrm{eq}} = A_0 \frac{[\mathrm{B}]_{\mathrm{eq}}}{K_{\mathrm{eq}} + [\mathrm{B}]_{\mathrm{eq}}}. \tag{1.6}$$

Thus, when $[\mathrm{B}]_{\mathrm{eq}} = K_{\mathrm{eq}}$, half of A is in the bound state at equilibrium.

There are several other features of the law of mass action that need to be mentioned. Suppose that the reaction involves the dimerization of two monomers of the same species A to produce species C,

$$\mathrm{A} + \mathrm{A} \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \mathrm{C}. \tag{1.7}$$

For every C that is made, two of A are used, and every time C degrades two copies of A are produced. As a result, the rate of reaction for A is

$$\frac{d[\mathrm{A}]}{dt} = 2k_-[\mathrm{C}] - 2k_+[\mathrm{A}]^2. \tag{1.8}$$

However, the rate of production of C is half that of A,

$$\frac{d[\mathrm{C}]}{dt} = -\frac{1}{2}\frac{d[\mathrm{A}]}{dt}, \tag{1.9}$$

and the quantity [A]+2[C] is conserved (provided there are no other reactions).

In a similar way, with a trimolecular reaction, the rate at which the reaction takes place is proportional to the product of three concentrations, and three molecules are consumed in the process, or released in the degradation of product. In real life, there are probably no truly trimolecular reactions, however, there are nonetheless some situations where a reaction might be effectively modeled as trimolecular. (See Exercise 2).

Unfortunately, the law of mass action cannot be used in all situations because not all chemical reaction mechanisms are known with sufficient detail. In fact, a vast number of chemical reactions cannot be described by mass action kinetics. Those reactions that follow mass action kinetics are called *elementary reactions* because presumably, they proceed directly from collision of the reactants. Reactions that do not follow mass action kinetics usually proceed by a complex mechanism consisting of several elementary reaction steps. It is often the case with biochemical reactions that the elementary reaction steps are not known or are very complicated to write down.

1.2 Thermodynamics and Rate Constants

There is a close relationship between the rate constants of a reaction and thermodynamics. The fundamental concept is that of *chemical potential*, which is the Gibbs free energy, G, per mole of a substance.

For a mixture of ideal gases, X_i, the chemical potential of gas i is a function of temperature, pressure and concentration,

$$G_i = G_i^0(T, P) + RT\ln(x_i), \tag{1.10}$$

where x_i is the mole fraction of X_i, R is the universal gas constant, T is the absolute temperature and P is the pressure of the gas (in atmospheres); values of these constants, and their units, are given in the Appendix. The quantity $G_i^0(T, P)$ is the standard free energy per mole of the pure ideal gas, i.e., when the mole fraction of the gas is 1. (Often μ is used to denote the Gibbs free energy per mole, but, to avoid notational confusion, we shall not do so here.) Note how, since $x_i \leq 1$, the free energy of an ideal gas in a mixture is always less than that of the pure ideal gas. The total Gibbs free energy of the mixture is

$$G = \sum_i n_i G_i, \tag{1.11}$$

where n_i is the number of moles of gas i.

Although we do not derive this expression for G_i, its logarithmic dependence on x_i has far-reaching implications, and appears in different guises all throughout this book. For instance, it appears in the logarithmic relationship of the Nernst

potential in Chapter 4 and thence in all discussions of the membrane potential and ionic current.

The theory of Gibbs free energy in ideal gases can be extended to ideal dilute solutions. By redefining the standard Gibbs free energy to be the free energy at a concentration of 1 M (i.e., 1 mole per liter) we obtain

$$G = G^0 + RT \ln(c), \tag{1.12}$$

where the concentration, c, must be written in units of moles per liter. In other words, c here is a pure number (as it must be, in order for us to take its logarithm); when multiplied by 1 M, c gives the concentration of the solute in M. The standard free energy, G^0, is obtained by measuring the free energy for a dilute solution and then extrapolating to $c = 1$. For biochemical applications, the dependence of free energy on pressure is ignored and the pressure is assumed to be 1 atm, while the temperature is taken to be 25° C. Derivations of these formulas can be found in physical chemistry textbooks such as Levine (1978) or Castellan (1971).

To be strictly correct, non-ideal solutions such as are found in cells should use the activity of the solute rather than its concentration. The relationship between activity and concentration is nontrivial. However, for dilute concentrations $a = c$, where c is the concentration of the chemical species in moles per liter.

Since the free energy is a potential, it denotes the preference of one state compared to another. Consider, for example, the simple reaction

$$\mathrm{A} \longrightarrow \mathrm{B}. \tag{1.13}$$

The change in chemical potential ΔG is defined as the difference between the chemical potential for state B (the product), denoted G_B, and the chemical potential for state A (the reactant), denoted G_A,

$$\begin{aligned} \Delta G &= G_B - G_A \\ &= G_B^0 - G_A^0 + RT \ln([\mathrm{B}]) - RT \ln([\mathrm{A}]) \\ &= \Delta G^0 + RT \ln([\mathrm{B}]/[\mathrm{A}]). \end{aligned} \tag{1.14}$$

The sign of ΔG is important, which is why it is defined with only one reaction direction shown even though we know that back reaction also occurs. In fact, there is a wonderful opportunity for confusion here, since there is no obvious way to decide which is the forward and which is the backward direction for a given reaction. If $\Delta G < 0$ then state B is preferred to state A and the reaction will tend to convert A into B, whereas, if $\Delta G > 0$, then state A is preferred to state B and the reaction will tend to convert B into A. Equilibrium occurs when neither state is preferred, so that $\Delta G = 0$ in which case

$$\frac{[\mathrm{B}]_{\mathrm{eq}}}{[\mathrm{A}]_{\mathrm{eq}}} = e^{\frac{-\Delta G^0}{RT}}. \tag{1.15}$$

Expressing this reaction in terms of forward and backward reaction rates,

$$\mathrm{A} \underset{k_-}{\overset{k_+}{\rightleftarrows}} \mathrm{B}, \tag{1.16}$$

we find that in steady state, $k_+[\mathrm{A}]_{\mathrm{eq}} = k_-[\mathrm{B}]_{\mathrm{eq}}$ so that

$$\frac{[\mathrm{A}]_{\mathrm{eq}}}{[\mathrm{B}]_{\mathrm{eq}}} = \frac{k_-}{k_+} = K_{\mathrm{eq}}. \tag{1.17}$$

Combining this with (1.15), we observe that

$$\frac{k_-}{k_+} = e^{\frac{\Delta G^0}{RT}}. \tag{1.18}$$

In other words, the more negative the difference in standard free energy, the greater the propensity for the reaction to proceed from left to right, and the smaller K_{eq}. Notice, however, that this gives us only the ratio of rate constants, and not their individual amplitudes. We learn nothing about whether a reaction is fast or slow from the standard free energy.

Similar relationships hold when there are multiple components in the reaction. Consider, for example, the more complex reaction

$$\alpha\mathrm{A} + \beta\mathrm{B} \longrightarrow \gamma\mathrm{C} + \delta\mathrm{D}. \tag{1.19}$$

The change of free energy for this reaction is defined as

$$\begin{aligned}
\Delta G &= \gamma G_C + \delta G_D - \alpha G_A - \beta G_B \\
&= \gamma G_C^0 + \delta G_D^0 - \alpha G_A^0 - \beta G_B^0 + RT \ln\left(\frac{[\mathrm{C}]^\gamma[\mathrm{D}]^\delta}{[\mathrm{A}]^\alpha[\mathrm{B}]^\beta}\right) \\
&= \Delta G^0 + RT \ln\left(\frac{[\mathrm{C}]^\gamma[\mathrm{D}]^\delta}{[\mathrm{A}]^\alpha[\mathrm{B}]^\beta}\right),
\end{aligned} \tag{1.20}$$

and at equilibrium,

$$\begin{aligned}
\Delta G^0 &= RT \ln\left(\frac{[\mathrm{A}]_{\mathrm{eq}}^\alpha[\mathrm{B}]_{\mathrm{eq}}^\beta}{[\mathrm{C}]_{\mathrm{eq}}^\gamma[\mathrm{D}]_{\mathrm{eq}}^\delta}\right) \\
&= RT \ln(K_{\mathrm{eq}}).
\end{aligned} \tag{1.21}$$

An important example of such a reaction is the hydrolysis of ATP to ADP and inorganic phosphate P_i, represented by the reaction

$$\mathrm{ATP} \longrightarrow \mathrm{ADP} + \mathrm{P_i}. \tag{1.22}$$

The standard free energy change for this reaction is

$$\Delta G^0 = G^0_{\mathrm{ADP}} + G^0_{\mathrm{P_i}} - G^0_{\mathrm{ATP}} = -31.0\mathrm{kJ\ mol}^{-1}, \tag{1.23}$$

and from this we could calculate the equilibrium constant for this reaction. However, the primary significance of this is not the size of the equilibrium constant, but rather the fact that ATP has free energy that can be used to drive other less favorable reactions. For example, in all living cells ATP is used to pump ions against their concentration gradient, a process called free energy transduction. In fact, if the equilibrium constant of this reaction is achieved then one can confidently assert that the system is dead. In living systems, the ratio of [ATP] to [ADP][P_i] is held well above the equilibrium value.

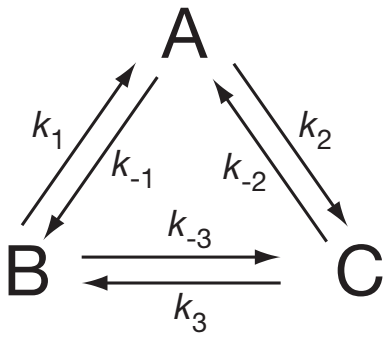

Figure 1.1: Schematic diagram of a reaction loop

1.3 Detailed Balance

Suppose that a set of reactions forms a loop, as shown in Fig. 1.1. By applying the law of mass action and setting the derivatives to zero we can find the steady-state concentrations of A, B and C. However, for the system to be in thermodynamic equilibrium a stronger condition must hold. Thermodynamic equilibrium requires that the free energy of each state be the same so that each individual reaction is in equilibrium. In other words, at equilibrium there is not only, say, no net change in [B], there is also no net conversion of B to C or B to A. This condition means that, at equilibrium, $k_1[\mathrm{B}] = k_{-1}[\mathrm{A}]$, $k_2[\mathrm{A}] = k_{-2}[\mathrm{C}]$ and $k_3[\mathrm{C}] = k_{-3}[\mathrm{B}]$. It follows that we must have

$$k_1 k_2 k_3 = k_{-1} k_{-2} k_{-3}, \tag{1.24}$$

or

$$K_1 K_2 K_3 = 1, \tag{1.25}$$

where $K_i = k_{-i}/k_i$. Since this condition does not depend on the concentrations of A, B or C, it must hold in general, not only at equilibrium.

For a more general reaction loop, the principle of detailed balance says that the product of rates in one direction around the loop must equal the product of rates in the other direction. If any of the rates are dependent on the concentrations of other chemicals, those concentrations must also be included. However, notice that since the principle of detailed balance is derived using (1.12) and the law of mass action, both of which are approximations, the principle of detailed balance is also an approximation.

1.4 Enzyme Kinetics

To see where some of the more complicated reaction schemes come from, we consider a reaction that is catalyzed by an enzyme. Enzymes are catalysts (generally proteins) that help convert other molecules called *substrates* into products, but they themselves are not changed by the reaction. Their most important features are

catalytic power, specificity, and regulation. Enzymes accelerate the conversion of substrate into product by lowering the free energy of activation of the reaction. For example, enzymes may aid in overcoming charge repulsions and allowing reacting molecules to come into contact for the formation of new chemical bonds. Or, if the reaction requires breaking of an existing bond, the enzyme may exert a stress on a substrate molecule, rendering a particular bond more easily broken. Enzymes are particularly efficient at speeding up biological reactions, giving increases in speed of up to 10 million times or more. They are also highly specific, usually catalyzing the reaction of only one particular substrate or closely related substrates. Finally, they are typically regulated by an enormously complicated set of positive and negative feedback systems, thus allowing precise control over the rate of reaction. A detailed presentation of enzyme kinetics, including many different kinds of models, can be found in Dixon and Webb (1979) or Segel (??). Here, we present some of the simplest models.

One of the first things one learns about enzyme reactions is that they do not follow the law of mass action directly. For, as the concentration of substrate (S) is increased, the rate of the reaction increases only to a certain extent, reaching a maximal reaction velocity at high substrate concentrations. This is in contrast to the law of mass action, which, when applied directly to the reaction of S with the enzyme E

$$\mathrm{S} + \mathrm{E} \longrightarrow \mathrm{P} + \mathrm{E}.$$

predicts that the reaction velocity increases linearly as [S] increases.

A model to explain the deviation from the law of mass action was first proposed by Michaelis and Menten (1913). In their reaction scheme, the enzyme E converts the substrate S into the product P through a two-step process. First E combines with S to form a complex C which then breaks down into the product P releasing E in the process. The reaction scheme is represented schematically by

$$\mathrm{S} + \mathrm{E} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \mathrm{C} \xrightarrow{k_2} \mathrm{P} + \mathrm{E}.$$

It is important to note that, although this appears to imply that P cannot combine with E to form the complex, this is not the case. In fact, all enzymes increase the speed of the reaction in both directions. Typically, however, reaction rates are measured under conditions where P is continually removed, which effectively prevents the reverse reaction from occurring. Thus, to determine the kinetic parameters from experimental data it suffices to assume that no reverse reaction occurs. Reversible enzyme kinetics are considered in Section 1.4.5.

There are two similar, but not identical, ways to analyze this equation; the equilibrium approximation, and the quasi-steady-state approximation. Because these methods give similar results it is easy to confuse these two approaches, so it is worthwhile to understand their differences.

We begin by defining $s = [\mathrm{S}], c = [\mathrm{C}], e = [\mathrm{E}]$, and $p = [\mathrm{P}]$. The law of mass action applied to this reaction mechanism yields four differential equations for the rates of change of s, c, e, and p as

$$\frac{ds}{dt} = k_{-1}c - k_1 se, \tag{1.26}$$

$$\frac{de}{dt} = (k_{-1} + k_2)c - k_1 se, \tag{1.27}$$

$$\frac{dc}{dt} = k_1 se - (k_2 + k_{-1})c, \tag{1.28}$$

$$\frac{dp}{dt} = k_2 c. \tag{1.29}$$

Notice that p can be found by direct integration, and there is a conserved quantity since $\frac{de}{dt} + \frac{dc}{dt} = 0$, so that $e + c = e_0$, where e_0 is the total amount of available enzyme.

1.4.1 The Equilibrium Approximation

In their original analysis, Michaelis and Menten assumed that the substrate is in instantaneous equilibrium with the complex, and thus

$$k_1 se = k_{-1} c. \tag{1.30}$$

Since $e + c = e_0$, we then find that

$$c = \frac{e_0 s}{K_1 + s}, \tag{1.31}$$

where $K_1 = k_{-1}/k_1$. Hence, the velocity, V, of the reaction, i.e., the rate at which the product is formed, is given by

$$V = \frac{dp}{dt} = k_2 c = \frac{k_2 e_0 s}{K_1 + s} = \frac{V_{\max} s}{K_1 + s}, \tag{1.32}$$

where $V_{\max} = k_2 e_0$ is the maximum reaction velocity, attained when all the enzyme is complexed with the substrate.

At small substrate concentrations, the reaction rate is linear, at a rate proportional to the amount of available enzyme e_0. At large concentrations, however, the reaction rate saturates to $V_{\max}$, so that the maximum rate of the reaction is limited by the amount of enzyme present and the dissociation rate constant k_2. For this reason, the dissociation reaction $\mathrm{C} \xrightarrow{k_2} \mathrm{P} + \mathrm{E}$ is said to be *rate limiting* for this reaction. At $s = K_1$, the reaction rate is half that of the maximum.

It is important to note that (1.30) cannot be exactly correct at all times; if it were, then according to (1.26) substrate would not be used up, and product would not be formed. This points out the fact that (1.30) is an approximation. It also illustrates the need for a systematic way to make approximate statements, so that one has an idea of the magnitude and nature of the errors introduced in making such an approximation.

It is a common mistake with the equilibrium approximation to conclude that since (1.30) holds, it must be that $\frac{ds}{dt} = 0$, which if this is true, implies that no substrate is being used up, nor product produced. Furthermore, if (1.30) holds, then it must be (from (1.28)) that $\frac{dc}{dt} = -k_2 c$. Where is the error here?

While you should work this out for yourself (see Exercise 11), the answer lies with the fact that the equilibrium approximation is actually equivalent to the assumption

that the reaction (1.26) is a very fast reaction, faster than others, or more precisely, that $k_{-1} \gg k_2$. Adding together (1.26) and (1.28) we find that

$$\frac{ds}{dt} + \frac{dc}{dt} = -k_2 c, \tag{1.33}$$

expressing the fact that the total quantity $s + c$ is changing on a slower time scale. Now when we use that $c = \frac{e_0 s}{K_s + s}$, we learn that

$$\frac{d}{dt}\left(s + \frac{e_0 s}{K_1 + s}\right) = -k_2 \frac{e_0 s}{K_1 + s}, \tag{1.34}$$

and thus,

$$\frac{ds}{dt}\left(1 + \frac{e_0 K_1}{(K_1 + s)^2}\right) = -k_2 \frac{e_0 s}{K_1 + s}, \tag{1.35}$$

which specifies the rate at which s is being consumed.

1.4.2 The Quasi-Steady-State Approximation

An alternative analysis of an enzymatic reaction was proposed by Briggs and Haldane (1925), and their analysis is now the basis for most present-day descriptions of enzyme reactions. Briggs and Haldane assumed that the rates of formation and breakdown of the complex were essentially equal at all times (except perhaps at the beginning of the reaction, as the complex is "filling up"). Thus, dc/dt should be approximately zero. With this approximation, it is relatively simple to determine the velocity of the reaction.

To give this approximation a systematic mathematical basis, it is useful to introduce dimensionless variables

$$\sigma = \frac{s}{s_0}, \quad x = \frac{c}{e_0}, \quad \tau = k_1 e_0 t, \quad \kappa = \frac{k_{-1} + k_2}{k_1 s_0}, \quad \epsilon = \frac{e_0}{s_0}, \quad \alpha = \frac{k_{-1}}{k_1 s_0}, \tag{1.36}$$

in terms of which we obtain the system of two differential equations

$$\frac{d\sigma}{d\tau} = -\sigma + x(\sigma + \alpha), \tag{1.37}$$

$$\epsilon \frac{dx}{d\tau} = \sigma - x(\sigma + \kappa). \tag{1.38}$$

There are usually a number of ways that a system of differential equations can be nondimensionalized. This nonuniqueness is often a source of great confusion, as it is often not obvious which choice of dimensionless variables and parameters is "best." In Section 1.6 we discuss this difficult problem briefly.

The remarkable effectiveness of enzymes as catalysts of biochemical reactions is reflected by their small concentrations needed compared to the concentrations of the substrates. For this model, this means that ϵ is small, typically in the range of 10^{-2} to 10^{-7}. Therefore, the reaction (1.38) is fast, equilibrates rapidly and remains in near-equilibrium even as the variable σ changes. Thus, we take the *quasi-steady-state approximation* $\epsilon \frac{dx}{d\tau} = 0$. Notice that this is *not* the same as taking $\frac{dx}{d\tau} = 0$.

However, because of the different scaling of x and c, it is equivalent to taking $\frac{dc}{dt} = 0$ as suggested in the introductory paragraph. The quasi-steady-state approximation means that the variable x is changing while restricted to some manifold described by setting the right-hand side of (1.38) to zero. This assumption is valid, provided that ϵ is small and $\frac{dx}{d\tau}$ is of order 1.

Notice also that the reaction of x is an exponential process with time constant at least as large as $\frac{\kappa}{\epsilon}$. To see this we write (1.38) as

$$\epsilon \frac{dx}{d\tau} + \kappa x = \sigma(1 - x). \tag{1.39}$$

Thus, the variable x "tracks" the steady state with a short delay.

It follows from the quasi-steady-state approximation that

$$x = \frac{\sigma}{\sigma + \kappa}, \tag{1.40}$$

$$\frac{d\sigma}{d\tau} = -\frac{q\sigma}{\sigma + \kappa}, \tag{1.41}$$

where $q = \kappa - \alpha = \frac{k_2}{k_1 s_0}$. Equation (1.41) describes the rate of uptake of the substrate and is called a *Michaelis–Menten law*. In terms of the original variables, this law is

$$V = \frac{dp}{dt} = -\frac{ds}{dt} = \frac{k_2 e_0 s}{s + K_m} = \frac{V_{\max} s}{s + K_m}, \tag{1.42}$$

where $K_m = \frac{k_{-1} + k_2}{k_1}$. In quasi-steady state, the concentration of the complex satisfies

$$c = \frac{e_0 s}{s + K_m}. \tag{1.43}$$

Note the similarity between (1.32) and (1.42), the only difference being that the equilibrium approximation uses K_s, while the quasi-steady-state approximation uses K_m. Despite this similarity of form, it is important to keep in mind that the two results are based on different approximations. The equilibrium approximation assumes that $k_{-1} \gg k_2$ whereas the quasi-steady state approximation assumes that $\epsilon \ll 1$. Notice, that if $k_{-1} \gg k_2$, then $K_m \approx K_s$, so that the two approximations are the same.

As with the law of mass action, the Michaelis–Menten law (1.42) is not universally applicable but is a useful approximation. It may be applicable even if $\epsilon = e_0/s_0$ is not small (see, for example, Exercise 11), and in model building it is often invoked without regard to the underlying assumptions.

While the individual rate constants are difficult to measure experimentally, the ratio K_m is relatively easy to measure because of the simple observation that (1.42) can be written in the form

$$\frac{1}{V} = \frac{1}{V_{\max}} + \frac{K_m}{V_{\max}} \frac{1}{s}. \tag{1.44}$$

In other words, $1/V$ is a linear function of $1/s$. Plots of this double reciprocal curve are called *Lineweaver–Burk plots*, and from such (experimentally determined) plots, $V_{\max}$ and K_m can be estimated.

Although a Lineweaver–Burk plot makes it easy to determine $V_{\max}$ and K_m from reaction rate measurements, it is not a simple matter to determine the reaction rate as a function of substrate concentration during the course of a single experiment. Substrate concentrations usually cannot be measured with sufficient accuracy or time resolution to permit the calculation of a reliable derivative. In practice, since it is more easily measured, the initial reaction rate is determined for a range of different initial substrate concentrations.

An alternative method to determine K_m and $V_{\max}$ from experimental data is the direct linear plot (Eisenthal and Cornish-Bowden, 1974; Cornish-Bowden and Eisenthal, 1974). First we write (1.42) in the form

$$V_{\max} = V + \frac{V}{s} K_m, \tag{1.45}$$

and then treat $V_{\max}$ and K_m as variables for each experimental measurement of V and s. (Recall that typically only the initial substrate concentration and initial velocity are used.) Then a plot of the straight line of $V_{\max}$ against K_m can be made. Repeating this for a number of different initial substrate concentrations and velocities gives a family of straight lines, which, in an ideal world free from experimental error, intersect at the single point $V_{\max}$ and K_m for that reaction. Of course, in reality, experimental error precludes an exact intersection, but $V_{\max}$ and K_m can be estimated from the median of the pairwise intersections.

1.4.3 Enzyme Inhibition

An enzyme inhibitor is a substance that inhibits the catalytic action of the enzyme. Enzyme inhibition is a common feature of enzyme reactions, and is an important means by which the activity of enzymes is controlled. Inhibitors come in many different types. For example, *irreversible inhibitors*, or *catalytic poisons*, decrease the activity of the enzyme to zero. This is the method of action of cyanide and many nerve gases. For this discussion, we restrict our attention to *competitive* inhibitors and *allosteric* inhibitors.

To understand the distinction between competitive and allosteric inhibition, it is useful to keep in mind that an enzyme molecule is usually a large protein, considerably larger than the substrate molecule whose reaction is catalyzed. Embedded in the large enzyme protein are one or more *active sites*, to which the substrate can bind to form the complex. In general, an enzyme catalyzes a single reaction or substrates with similar structures. This is believed to be a steric property of the enzyme that results from the three-dimensional shape of the enzyme allowing it to fit in a "lock-and-key" fashion with a corresponding substrate molecule.

If another molecule has a shape similar enough to that of the substrate molecule, it may also bind to the active site, preventing the binding of a substrate molecule, thus inhibiting the reaction. Because the inhibitor competes with the substrate molecule for the active site, it is called a competitive inhibitor.

However, because the enzyme molecule is large, it often has other binding sites, distinct from the active site, the binding of which affects the activity of the enzyme at the active site. These binding sites are called *allosteric* sites (from the Greek for "another solid") to emphasize that they are structurally different from the catalytic

active sites. They are also called *regulatory sites* to emphasize that the catalytic activity of the protein is regulated by binding at this allosteric site. The ligand (any molecule that binds to a specific site on a protein, from Latin *ligare*, to bind) that binds at the allosteric site is called an *effector* or *modifier*, which, if it increases the activity of the enzyme, is called an allosteric activator, while if it decreases the activity of the enzyme, it is called an allosteric inhibitor. The allosteric effect is presumed to arise because of a conformational change of the enzyme, that is, a change in the folding of the polypeptide chain, called an *allosteric transition*.

Competitive Inhibition

In the simplest example of a competitive inhibitor, the reaction is stopped when the inhibitor is bound to the active site of the enzyme. Thus,

$$\mathrm{S} + \mathrm{E} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \mathrm{C}_1 \xrightarrow{k_2} \mathrm{E} + \mathrm{P},$$

$$\mathrm{E} + \mathrm{I} \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} \mathrm{C}_2.$$

From the law of mass action we get

$$\frac{ds}{dt} = -k_1 se + k_{-1} c_1, \tag{1.46}$$

$$\frac{di}{dt} = -k_3 ie + k_{-3} c_2, \tag{1.47}$$

$$\frac{dc_1}{dt} = k_1 se - (k_{-1} + k_2) c_1, \tag{1.48}$$

$$\frac{dc_2}{dt} = k_3 ie - k_{-3} c_2. \tag{1.49}$$

where $s = [S]$, $c_1 = [C_1]$, and $c_2 = [C_2]$. We know that $e + c_1 + c_2 = e_0$, so an equation for the dynamics of e is superfluous. As before, it is not necessary to write an equation for the accumulation of the product. To be systematic, the next step is to introduce dimensionless variables, and identify those reactions that are rapid and equilibrate rapidly to their quasi-steady states. However, from our previous experience (or from a calculation on a piece of scratch paper), we know, assuming the enzyme-to-substrate ratios are small, that the fast equations are those for c_1 and c_2. Hence, the quasi-steady states are found by (formally) setting $dc_1/dt = dc_2/dt = 0$ and solving for c_1 and c_2. Recall that this does *not* mean that c_1 and c_2 are unchanging, rather that they are changing in quasi-steady-state fashion, keeping the right-hand sides of these equations nearly zero. This gives

$$c_1 = \frac{K_i e_0 s}{K_m i + K_i s + K_m K_i}, \tag{1.50}$$

$$c_2 = \frac{K_m e_0 i}{K_m i + K_i s + K_m K_i}, \tag{1.51}$$

where $K_m = \frac{k_{-1}+k_2}{k_1}$, $K_i = k_{-3}/k_3$. Thus, the velocity of the reaction is

$$V = k_2 c_1 = \frac{k_2 e_0 s K_i}{K_m i + K_i s + K_m K_i} = \frac{V_{\max} s}{s + K_m(1 + i/K_i)}. \tag{1.52}$$

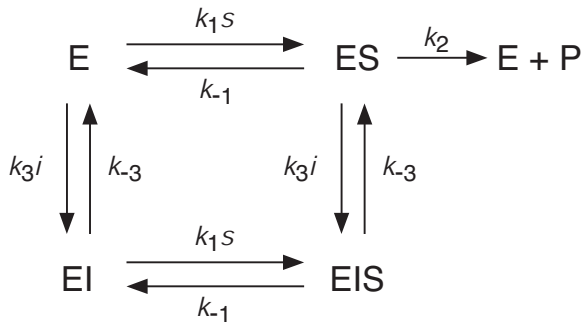

Figure 1.2: Diagram of the possible states of an enzyme with one allosteric and one catalytic binding site.

Notice that the effect of the inhibitor is to increase the effective equilibrium constant of the enzyme by the factor $1 + i/K_i$, from K_m to $K_m(1 + i/K_i)$, thus decreasing the velocity of reaction, while leaving the maximum velocity unchanged.

Allosteric Inhibitors

If the inhibitor can bind at an allosteric site, we have the possibility that the enzyme could bind both the inhibitor and the substrate simultaneously. In this case, there are four possible binding states for the enzyme, and transitions between them, as demonstrated graphically in Fig. 1.2.

The simplest analysis of this reaction scheme is the equilibrium analysis. (The more complicated quasi-steady-state analysis is posed as Exercise 5.) We define $K_s = k_{-1}/k_1$, $K_i = k_{-3}/k_3$, and let x, y, and z denote, respectively, the concentrations of ES, EI and EIS. Then, it follows from the law of mass action that at equilibrium (take each of the 4 transitions to be at equilibrium),

$$\begin{aligned}(e_0 - x - y - z)s - K_s x &= 0, &(1.53)\\ (e_0 - x - y - z)i - K_i y &= 0, &(1.54)\\ ys - K_s z &= 0, &(1.55)\\ xi - K_i z &= 0, &(1.56)\end{aligned}$$

where $e_0 = e + x + y + z$ is the total amount of enzyme. Notice that this is a linear system of equations for x, y, and z. Although there are four equations, one is a linear combination of the other three (the system is of rank three), so that we can determine x, y, and z as functions of i and s, finding

$$x = \frac{e_0 K_i}{K_i + i}\frac{s}{K_s + s}. \tag{1.57}$$

It follows that the reaction rate, $V = k_2 x$, is given by

$$V = \frac{V_{\max}}{1 + i/K_i}\frac{s}{K_s + s}, \tag{1.58}$$

where $V_{\max} = k_2 e_0$. Thus, in contrast to the competitive inhibitor, the allosteric inhibitor decreases the maximum velocity of the reaction, while leaving K_s unchanged. (Of course, the situation is more complicated if the quasi-steady-state approximation is used, and no such simple conclusion follows.)

1.4.4 Cooperativity

For many enzymes, the reaction velocity is not a simple hyperbolic curve, as predicted by the Michaelis–Menten model, but often has a sigmoidal character. This can result from cooperative effects, in which the enzyme can bind more than one substrate molecule but the binding of one substrate molecule affects the binding of subsequent ones.

Much of the original theoretical work on cooperative behavior was stimulated by the properties of hemoglobin, and this is often the context in which cooperativity is discussed. A detailed discussion of hemoglobin and oxygen binding is given in Chapter **??**, while here cooperativity is discussed in more general terms.

Suppose that an enzyme can bind two substrate molecules, so it can exist in one of three states, namely as a free molecule E, as a complex with one occupied center C_1, and as a complex with two occupied centers C_2. The reaction mechanism is represented by

$$\mathrm{S} + \mathrm{E} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \mathrm{C}_1 \xrightarrow{k_2} \mathrm{E} + \mathrm{P}, \tag{1.59}$$

$$\mathrm{S} + \mathrm{C}_1 \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} \mathrm{C}_2 \xrightarrow{k_4} \mathrm{C}_1 + \mathrm{P}. \tag{1.60}$$

Using the law of mass action, one can write down the rate equations for the 5 concentrations [S], [E], [C_1], [C_2], and [P]. However, because the amount of product [P] can be determined by quadrature, and because the total amount of enzyme molecule is conserved, we only need three equations for the three quantities [S], [C_1], and [C_2]. These are

$$\frac{ds}{dt} = -k_1 s e + k_{-1} c_1 - k_3 s c_1 + k_{-3} c_2, \tag{1.61}$$

$$\frac{dc_1}{dt} = k_1 s e - (k_{-1} + k_2) c_1 - k_3 s c_1 + (k_4 + k_{-3}) c_2, \tag{1.62}$$

$$\frac{dc_2}{dt} = k_3 s c_1 - (k_4 + k_{-3}) c_2, \tag{1.63}$$

where $s = [\mathrm{S}], c_1 = [\mathrm{C}_1], c_2 = [\mathrm{C}_2]$, and $e + c_1 + c_2 = e_0$.

Proceeding as before, we invoke the quasi-steady-state assumption that $dc_1/dt = dc_2/dt = 0$, and solve for c_1 and c_2 to get

$$c_1 = \frac{K_2 e_0 s}{K_1 K_2 + K_2 s + s^2}, \tag{1.64}$$

$$c_2 = \frac{e_0 s^2}{K_1 K_2 + K_2 s + s^2}, \tag{1.65}$$

where $K_1 = \frac{k_{-1}+k_2}{k_1}$ and $K_2 = \frac{k_4+k_{-3}}{k_3}$. The reaction velocity is thus given by

$$V = k_2 c_1 + k_4 c_2 = \frac{(k_2 K_2 + k_4 s) e_0 s}{K_1 K_2 + K_2 s + s^2}. \tag{1.66}$$

It is instructive to examine two extreme cases. First, if the active sites act independently and identically, then $k_1 = 2k_3 = 2k_+$, $2k_{-1} = k_{-3} = 2k_-$ and $2k_2 = k_4$, where k_+ and k_- are the forward and backward reaction rates for the individual binding sites. The factors of 2 occur because two identical binding sites are involved in the reaction, doubling the amount of the reactant. In this case,

$$V = \frac{2k_2 e_0 (K + s) s}{K^2 + 2Ks + s^2} = 2\frac{k_2 e_0 s}{K + s}, \tag{1.67}$$

where $K = \frac{k_- + k_2}{k_+}$ is the equilibrium constant for the individual binding site. As expected, the rate of reaction is exactly twice that for the individual binding site.

In the opposite extreme, suppose that the binding of the first substrate molecule is slow, but that with one site bound, binding of the second is fast (this is large positive cooperativity). This can be modeled by letting $k_3 \to \infty$ and $k_1 \to 0$, while keeping $k_1 k_3$ constant, in which case $K_2 \to 0$ and $K_1 \to \infty$ while $K_1 K_2$ is constant. In this limit, the velocity of the reaction is

$$V = \frac{k_4 e_0 s^2}{K_m^2 + s^2} = \frac{V_{\max} s^2}{K_m^2 + s^2}, \tag{1.68}$$

where $K_m^2 = K_1 K_2$, and $V_{\max} = k_4 e_0$.

In general, if n substrate molecules can bind to the enzyme, there are n equilibrium constants, K_1 through K_n. In the limit as $K_n \to 0$ and $K_1 \to \infty$ while keeping $K_1 K_n$ fixed, the rate of reaction is

$$V = \frac{V_{\max} s^n}{K_m^n + s^n}, \tag{1.69}$$

where $K_m^n = \Pi_{i=1}^n K_i$. This rate equation is known as the *Hill equation*. Typically, the Hill equation is used for reactions whose detailed intermediate steps are not known but for which cooperative behavior is suspected. The exponent n and the parameters $V_{\max}$ and K_m are usually determined from experimental data. Observe that

$$n \ln s = n \ln K_m + \ln \left(\frac{V}{V_{\max} - V} \right), \tag{1.70}$$

so that a plot of $\ln(\frac{V}{V_{\max}-V})$ against $\ln s$ (called a *Hill plot*) should be a straight line of slope n. Although the exponent n suggests an n-step process (with n binding sites), in practice it is not unusual for the best fit for n to be noninteger.

An enzyme can also exhibit negative cooperativity (Koshland and Hamadani, 2002), in which the binding of the first substrate molecule *decreases* the rate of subsequent binding. This can be modeled by decreasing k_3. In Fig. 1.3 we plot the reaction velocity against the substrate concentration for the cases of independent binding sites (no cooperativity), extreme positive cooperativity (the Hill equation), and negative cooperativity. From this figure it can be seen that with positive cooperativity, the reaction velocity is a sigmoidal function of the substrate concentration, while negative cooperativity primarily decreases the overall reaction velocity.

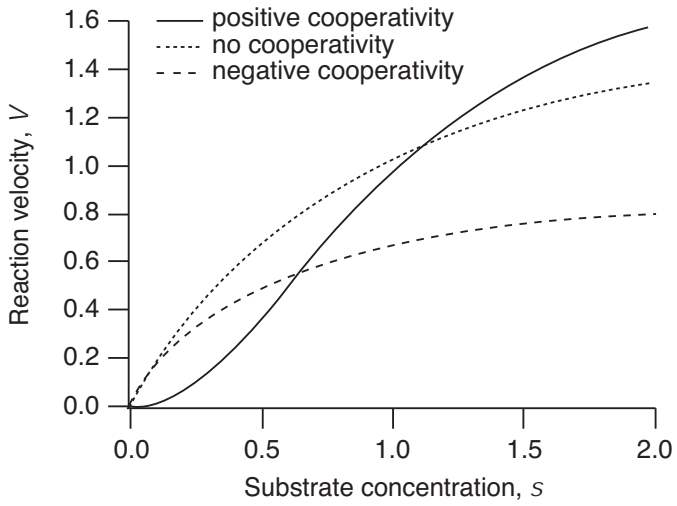

Figure 1.3: Reaction velocity plotted against substrate concentration, for three different cases. Positive cooperativity, $K_1 = 1000$, $K_2 = 0.001$; independent binding sites, $K_1 = 0.5$, $K_2 = 2$; and negative cooperativity, $K_1 = 0.5$, $K_2 = 100$. The other parameters were chosen as $e_0 = 1$, $k_2 = 1$, $k_4 = 2$. Concentration and time units are arbitrary.

The Monod–Wyman–Changeux Model

Cooperative effects occur when the binding of one substrate molecule alters the rate of binding of subsequent ones. However, the above models give no explanation of how such alterations in the binding rate occur. The earliest model proposed to account for cooperative effects in terms of the enzyme's conformation was that of Monod, Wyman, and Changeux (1965). Their model is based on the following assumptions about the structure and behavior of enzymes.

1. Cooperative proteins are composed of several identical reacting units, called *protomers*, or subunits, each containing one binding site, that occupy equivalent positions within the protein.

2. The protein has two conformational states, usually denoted by R and T, which differ in their ability to bind ligands.

3. If the binding of a ligand to one protomer induces a conformational change in that protomer, an identical conformational change is induced in all protomers. Because of this assumption, Monod-Wyman-Changeux (MWC) models are often called *concerted* models, as each subunit acts in concert with the others.

To illustrate how these assumptions can be quantified, we consider a protein with only two binding sites. Thus, the protein can exist in one of six states: $\mathrm{R}_i, i = 0, 1, 2$, or $\mathrm{T}_i, i = 0, 1, 2$, where the subscript i is the number of bound ligands. (In the

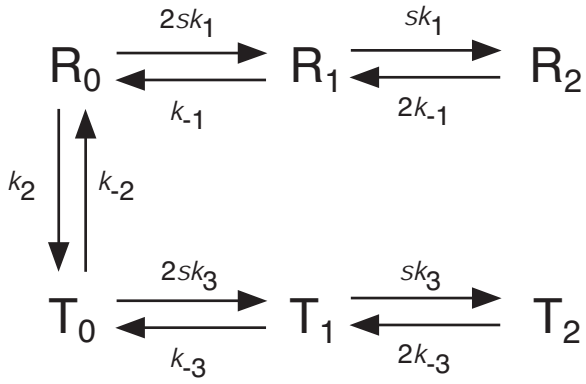

Figure 1.4: Diagram of the states of the protein, and the possible transitions, in a six-state Monod–Wyman–Changeux model.

original model of Monod, Wyman and Changeux, R denoted a *relaxed* state, while T denoted a *tense* state.) For simplicity, we also assume that R_1 cannot convert directly to T_1, or vice versa, and similarly for R_2 and T_2. The general case is left for Exercise 6. The states of the protein and the allowable transitions are illustrated in Fig. 1.4. As with other enzyme models, we assume that the production rate of product is proportional to the amount of substrate that is bound to the enzyme.

We now assume that all the reactions are in equilibrium. We let a lowercase letter denote a concentration, and thus r_i and t_i denote the concentrations of chemical species R_i and T_i respectively. Also, as before, we let s denote the concentration of the substrate. Then, the fraction Y of occupied sites (also called the *saturation function*) is

$$Y = \frac{r_1 + 2r_2 + t_1 + 2t_2}{2(r_0 + r_1 + r_2 + t_0 + t_1 + t_2)}. \tag{1.71}$$

(This is also proportional to the production rate of product.) Furthermore, with $K_i = k_{-i}/k_i$, for $i = 1, 2, 3$, we find that

$$r_1 = 2sK_1^{-1}r_0, \qquad r_2 = s^2K_1^{-2}r_0, \tag{1.72}$$

$$t_1 = 2sK_3^{-1}t_0, \qquad t_2 = s^2K_3^{-2}t_0. \tag{1.73}$$

Substituting these into (1.71) gives

$$Y = \frac{sK_1^{-1}(1 + sK_1^{-1}) + K_2^{-1}[sK_3^{-1}(1 + sK_3^{-1})]}{(1 + sK_1^{-1})^2 + K_2^{-1}(1 + sK_3^{-1})^2}, \tag{1.74}$$

where we have used that $r_0/t_0 = K_2$. More generally, if there are n binding sites, then

$$Y = \frac{sK_1^{-1}(1 + sK_1^{-1})^{n-1} + K_2^{-1}[sK_3^{-1}(1 + sK_3^{-1})^{n-1}]}{(1 + sK_1^{-1})^n + K_2^{-1}(1 + sK_3^{-1})^n}. \tag{1.75}$$

In general, Y is a sigmoidal function of s.

It is not immediately apparent how cooperative binding kinetics arises from this model. After all, each binding site in the R conformation is identical, as is each binding site in the T conformation. In order to get cooperativity it is necessary that the binding affinity of the R conformation be different from that of the T conformation. In the special case that the binding affinities of the R and T conformations are equal (i.e., $K_1 = K_3 = K$, say) the binding curve (1.75) reduces to

$$Y = \frac{s}{K + s}, \tag{1.76}$$

which is just non-cooperative Michaelis-Menten kinetics.

Suppose that one conformation, T say, binds the substrate with a higher affinity. Then, when the substrate concentration increases, T_0 will be pushed through to T_1 faster than R_0 will be pushed to R_1, resulting in an increase in the amount of substrate bound to the T state, and thus increased overall binding of substrate. Hence the cooperative behavior of the model.

If $K_2 = \infty$, so that only one conformation exists, then once again the saturation curve reduces to the Michaelis-Menten equation, $Y = s/(s + K_1)$. Hence each conformation, by itself, has non-cooperative Michaelis-Menten binding kinetics. It is only when the overall substrate binding can be biased to one conformation or the other that cooperativity appears.

Interestingly, MWC models cannot exhibit negative cooperativity. No matter whether$K_1 > K_3$ or *vice versa*, the binding curve always exhibits positive cooperativity.

The Koshland-Nemethy-Filmer model

One major alternative to the MWC model is that proposed by Koshland, Nemethy and Filmer in 1966 (the KNF model). Instead of requiring that all subunit transitions occur in concert, as in the MWC model, the KNF model assumes that substrate binding to one subunit causes a conformational change in that subunit only, and that this conformational change causes a change in the binding affinity of the neighboring subunits. Thus, in the KNF model, each subunit can be in a different conformational state, and transitions from one state to the other occur sequentially as more substrate is bound. For this reason KNF models are often called *sequential* models. The increased generality of the KNF model allows for the possibility of negative cooperativity, as the binding to one subunit can *decrease* the binding affinity of its neighbors.

When binding shows positive cooperativity, it has proven difficult to distinguish between the MWC and KNF models on the basis of experimental data. In one of the most intensely studied cooperative mechanisms, that of oxygen binding to hemoglobin, there is experimental evidence for both models, and the actual mechanism is probably a combination of both.

There are many other models of enzyme cooperativity, and the interested reader is referred to Dixon and Webb (1979) for a comprehensive discussion and comparison of other models in the literature.

1.4.5 Reversible Enzyme Reactions

Since all enzyme reactions are reversible, a general understanding of enzyme kinetics must take this reversibility into account. In this case, the reaction scheme is

$$\mathrm{S} + \mathrm{E} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \mathrm{C} \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} \mathrm{P} + \mathrm{E}.$$

Proceeding as usual, we let $e + c = e_0$ and make the quasi-steady state assumption

$$0 = \frac{dc}{dt} = k_1 s(e_0 - c) - (k_{-1} + k_2)c + k_{-2}p(e_0 - c), \tag{1.77}$$

from which it follows that

$$c = \frac{e_0(k_1 s + k_{-2}p)}{k_1 s + k_{-2}p + k_{-1} + k_2}. \tag{1.78}$$

The reaction velocity, $V = \frac{dP}{dt} = k_2 c - k_{-2}pe$, can then be calculated to be

$$V = e_0 \cdot \frac{k_1 k_2 s - k_{-1}k_{-2}p}{k_1 s + k_{-2}p + k_{-1} + k_2}. \tag{1.79}$$

When p is small (i.e., if product is being continually removed), the reverse reaction is negligible and we get the previous answer (1.42).

Unlike the irreversible case, the equilibrium and quasi-steady-state assumptions for reversible enzyme kinetics give qualitatively different answers. If we assume that S, E and C are in fast equilibrium (instead of assuming that C is at quasi-steady state) we get

$$k_1 s(e_0 - c) = k_{-1}c, \tag{1.80}$$

from which it follows that

$$V = k_2 c - k_{-2}p(e_0 - c) = e_0 \frac{k_1 k_2 s - k_{-1}k_{-2}p}{k_1 s + k_{-1}}. \tag{1.81}$$

Comparing this to (1.79) we see that the quasi-steady-state assumption gives additional terms in the denominator involving the product, p. These differences result from the assumption underlying the fast equilibrium assumption, that k_{-1} and k_1 are both substantially larger than k_{-2} and k_2, respectively. Which of these approximations is best depends, of course, on the details of the reaction.

Calculation of the equations for a reversible enzyme reaction in which the enzyme has multiple binding sites is left for the exercises (Exercise 9).

1.5 Glycolysis and Glycolytic Oscillations

Metabolism is the process of extracting useful energy from chemical bonds. A metabolic pathway is the sequence of enzymatic reactions that take place in order to transfer chemical energy from one form to another. The common carrier of energy in the cell is the chemical *adenosine triphosphate* (ATP). ATP is formed

by the addition of an inorganic phosphate group (HPO_4^{2-}) to *adenosine diphosphate* (ADP), or by the addition of two inorganic phosphate groups to *adenosine monophosphate* (AMP). The process of adding an inorganic phosphate group to a molecule is called *phosphorylation*. Since the three phosphate groups on ATP carry negative charges, considerable energy is required to overcome the natural repulsion of like-charged phosphates as additional groups are added to AMP. Thus, the hydrolysis (the cleavage of a bond by water) of ATP to ADP releases large amounts of energy.

Energy to perform chemical work is made available to the cell by the oxidation of glucose to carbon dioxide and water, with a net release of energy. The overall chemical reaction for the oxidation of glucose can be written as

$$C_6H_{12}O_6 + 6O_2 \longrightarrow 6CO_2 + 6H_2O + \text{energy}, \tag{1.82}$$

but of course, this is not an elementary reaction. Instead, this reaction takes place in a series of enzymatic reactions, with three major reaction stages, *glycolysis*, the *Krebs cycle*, and the *electron transport* (or *cytochrome) system*.

The oxidation of glucose is associated with a large negative free energy, $\Delta G^0 = -2878.41\text{kJ/mol}$, some of which is dissipated as heat. However, in living cells much of this free energy in stored in ATP, with one molecule of glucose resulting in 38 molecules of ATP.

Glycolysis involves 11 elementary reaction steps, each of which is an enzymatic reaction. Here we consider a simplified model of the initial steps. (To understand more of the labyrinthine complexity of glycolysis, interested readers are encouraged to consult a specialized book on biochemistry, such as Stryer, 1988.) The first three steps of glycolysis are (Fig. 1.5)

1. the phosphorylation of glucose to glucose 6-phosphate;
2. the isomerization of glucose 6-phosphate to fructose 6-phosphate; and
3. the phosphorylation of fructose 6-phosphate to fructose 1,6-bisphosphate.

The direct reaction of glucose with phosphate to form glucose 6-phosphate has a relatively large positive standard free energy change ($\Delta G^0 = 14.3\text{kJ/mol}$) and so will not occur significantly under physiological conditions. However, the first step of metabolism is coupled with the hydrolysis of ATP to ADP (catalyzed by the enzyme hexokinase), giving this step a net negative standard free energy change and making the reaction strongly spontaneous. This feature turns out to be important for the efficient operation of glucose membrane transporters, which we discuss in the next Chapter.

The second step of glycolysis has a relatively small positive standard free energy change ($\Delta G^0 = 1.7$ kJ/mol), with an equilibrium constant of 0.5. This means that significant amounts of product are formed under normal conditions.

The third step is, like the first step, energetically unfavorable, were it not coupled with the hydrolysis of ATP. However, the net standard free energy change ($\Delta G^0 = -14.2$ kJ/mol) means that not only is this reaction strongly favored, but also that it augments the reaction in the second step by depleting the product of the second step.

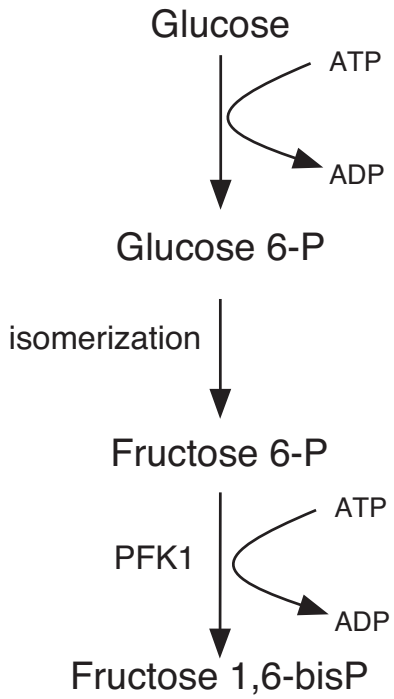

Figure 1.5: The first three reactions in the glycolytic pathway.

This third reaction is catalyzed by the enzyme phosphofructokinase (PFK1). PFK1 is an example of an allosteric enzyme as it is allosterically inhibited by ATP. Note that ATP is both a substrate of PFK1, binding at a catalytic site, and an allosteric inhibitor, binding at a regulatory site. The inhibition due to ATP is removed by AMP, and thus the activity of PFK1 increases as the ratio of ATP to AMP decreases. This feedback enables PFK1 to regulate the rate of glycolysis based on the availability of ATP. If ATP levels fall, PFK1 activity increases thereby increasing the rate of production of ATP, whereas, if ATP levels become high, PFK1 activity drops shutting down the production of ATP.

As PFK1 phosphorylates fructose 6-P, ATP is converted to ADP. ADP, in turn, is converted back to ATP and AMP by the reaction

$$2\text{ADP} \rightleftharpoons \text{ATP} + \text{AMP},$$

which is catalyzed by the enzyme adenylate kinase. Since there is normally little AMP in cells, the conversion of ADP to ATP and AMP serves to significantly decrease the ATP/AMP ratio, thus activating PFK1. This is an example of a positive feedback loop; the greater the activity of PFK1, the lower the ATP/AMP ratio, thus further increasing PFK1 activity.

It was discovered in 1980 that in some cell types, another important allosteric activator of PFK1 is fructose 2,6-bisphosphate (Stryer, 1988), which is formed from fructose 6-phosphate in a reaction catalyzed by phosphofructokinase 2 (PFK2), a different enzyme from phosphofructokinase (PFK1) (you were given fair warning about the labyrinthine nature of this process!). Of particular significance is that an abundance of fructose 6-phosphate leads to a corresponding abundance of fructose 2,6-bisphosphate, and thus a corresponding increase in the activity of PFK1. This is an example of a negative feedback loop, where an increase in the substrate concentration leads to a greater rate of substrate reaction and consumption. Clearly, PFK1 activity is controlled by an intricate system of reactions, the collective behavior of which is not obvious a priori.

Under certain conditions the rate of glycolysis is known to be oscillatory, or even chaotic (Nielsen et al., 1997). This biochemical oscillator has been known and studied experimentally for some time. For example, Hess and Boiteux (1973) devised a flow reactor containing yeast cells into which a controlled amount of substrate (either glucose or fructose) was continuously added. They measured the pH and fluorescence of the reactants, thereby monitoring the glycolytic activity, and they found ranges of continuous input under which glycolysis was periodic.

A mathematical model describing this oscillation was proposed by Sel'kov (1968) and later modified by Goldbeter and Lefever (1972). It is meant to capture only the positive feedback of ADP on PFK1 activity, and does not take into account the negative feedback process that was discovered more recently. (An interesting exercise would be to construct a more detailed model, including both positive and negative feedback processes, to see what difference this makes to the conclusions.) In the Sel'kov model, PFK1 is inactive in its unbound state but is activated by binding with several ADP molecules. Note that, for simplicity, the model does not take into account the conversion of ADP to AMP and ATP, but assumes that ADP activates PFK1 directly, since the overall effect is similar. In the active state, the

enzyme catalyzes the production of ADP from ATP as fructose-6-P is phosphorylated. Sel'kov's reaction scheme for this process is as follows: PFK1 (denoted by E) is activated or deactivated by binding or unbinding with γ molecules of ADP (denoted by S_2)

$$\gamma S_2 + E \underset{k_{-3}}{\overset{k_3}{\rightleftarrows}} ES_2^\gamma,$$

and ATP (denoted S_1) can bind with the activated form of enzyme to produce a product molecule of ADP. In addition, there is assumed to be a steady supply rate of S_1, while product S_2 is irreversibly removed. Thus,

$$\xrightarrow{v_1} \quad S_1, \tag{1.83}$$

$$S_1 + ES_2^\gamma \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} S_1ES_2^\gamma \xrightarrow{k_2} ES_2^\gamma + S_2, \tag{1.84}$$

$$S_2 \xrightarrow{v_2}. \tag{1.85}$$

Applying the law of mass action to the Sel'kov kinetic scheme, we find five differential equations for the production of the five species $s_1 = [S_1], s_2 = [S_2], e = [E], x_1 = [ES_2^\gamma], x_2 = [S_1ES_2^\gamma]$:

$$\frac{ds_1}{dt} = v_1 - k_1 s_1 x_1 + k_{-1} x_2, \tag{1.86}$$

$$\frac{ds_2}{dt} = k_2 x_2 - k_3 s_2^\gamma e + k_{-3} x_1 - v_2 s_2, \tag{1.87}$$

$$\frac{dx_1}{dt} = -k_1 s_1 x_1 + (k_{-1} + k_2) x_2 + k_3 s_2^\gamma e - k_{-3} x_1, \tag{1.88}$$

$$\frac{dx_2}{dt} = k_1 s_1 x_1 - (k_{-1} + k_2) x_2. \tag{1.89}$$

The fifth differential equation is not necessary, because the total available enzyme is conserved, $e + x_1 + x_2 = e_0$. Now we introduce dimensionless variables $\sigma_1 = \frac{k_1 s_1}{k_2 + k_{-1}}, \sigma_2 = (\frac{k_3}{k_{-3}})^{1/\gamma} s_2, u_1 = x_1/e_0, u_2 = x_2/e_0, t = \frac{k_2 + k_{-1}}{e_0 k_1 k_2}\tau$ and find

$$\frac{d\sigma_1}{d\tau} = \nu - \frac{k_2 + k_{-1}}{k_2} u_1 \sigma_1 + \frac{k_{-1}}{k_2} u_2, \tag{1.90}$$

$$\frac{d\sigma_2}{d\tau} = \alpha \left[u_2 - \frac{k_{-3}}{k_2} \sigma_2^\gamma (1 - u_1 - u_2) + \frac{k_{-3}}{k_2} u_1 \right] - \eta \sigma_2, \tag{1.91}$$

$$\epsilon \frac{du_1}{d\tau} = u_2 - \sigma_1 u_1 + \frac{k_{-3}}{k_2 + k_{-1}} \left[\sigma_2^\gamma (1 - u_1 - u_2) - u_1 \right], \tag{1.92}$$

$$\epsilon \frac{du_2}{d\tau} = \sigma_1 u_1 - u_2, \tag{1.93}$$

where $\epsilon = \frac{e_0 k_1 k_2}{(k_2 + k_{-1})^2}, \nu = \frac{v_1}{k_2 e_0}, \eta = \frac{v_2 (k_2 + k_{-1})}{k_1 k_2 e_0}, \alpha = \frac{k_2 + k_{-1}}{k_1} (\frac{k_3}{k_{-3}})^{1/\gamma}$. If we assume that ϵ is a small number, then both u_1 and u_2 are "fast" variables and can be set to their quasi-steady values,

$$u_1 = \frac{\sigma_2^\gamma}{\sigma_2^\gamma \sigma_1 + \sigma_2^\gamma + 1}, \tag{1.94}$$

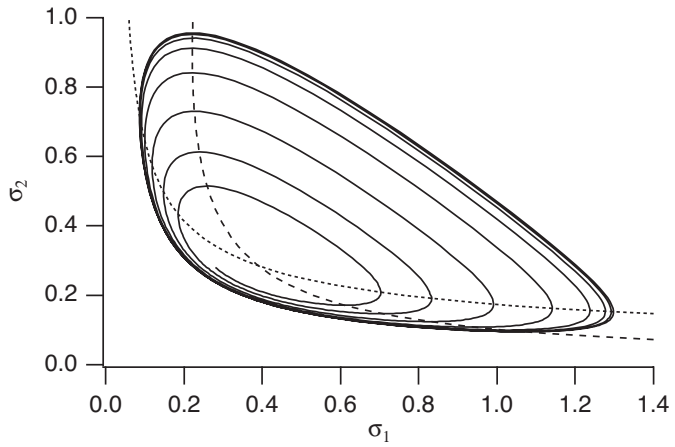

Figure 1.6: Phase portrait of the Sel'kov glycolysis system with $\nu = 0.0285, \eta = 0.1, \alpha = 1.0$, and $\gamma = 2$. Dotted curve: $\frac{d\sigma_1}{d\tau} = 0$. Dashed curve: $\frac{d\sigma_2}{d\tau} = 0$.

$$u_2 = \frac{\sigma_1 \sigma_2^\gamma}{\sigma_2^\gamma \sigma_1 + \sigma_2^\gamma + 1} = f(\sigma_1, \sigma_2), \tag{1.95}$$

and with these quasi-steady values, the evolution of σ_1 and σ_2 is governed by

$$\frac{d\sigma_1}{d\tau} = \nu - f(\sigma_1, \sigma_2), \tag{1.96}$$

$$\frac{d\sigma_2}{d\tau} = \alpha f(\sigma_1, \sigma_2) - \eta\sigma_2. \tag{1.97}$$

The goal of the following analysis is to demonstrate that this system of equations has oscillatory solutions for some range of the supply rate ν. First observe that because of saturation, the function $f(\sigma_1, \sigma_2)$ is bounded by 1. Thus, if $\nu > 1$, the solutions of the differential equations are not bounded. For this reason we consider only $0 < \nu < 1$. The nullclines of the flow are given by the equations

$$\sigma_1 = \frac{\nu}{1-\nu}\frac{1+\sigma_2^\gamma}{\sigma_2^\gamma} \qquad \left(\frac{d\sigma_1}{d\tau} = 0\right), \tag{1.98}$$

$$\sigma_1 = \frac{1+\sigma_2^\gamma}{\sigma_2^{\gamma-1}(p-\sigma_2)} \qquad \left(\frac{d\sigma_2}{d\tau} = 0\right), \tag{1.99}$$

where $p = \alpha/\eta$. These two nullclines are shown plotted as dotted and dashed curves respectively in Fig. 1.6.

The steady-state solution is unique and satisfies

$$\sigma_2 = p\nu, \tag{1.100}$$

$$\sigma_1 = \frac{\nu(1+\sigma_2^\gamma)}{(1-\nu)\sigma_2^\gamma}. \tag{1.101}$$

The stability of the steady solution is found by linearizing the differential equations about the steady-state solution and examining the eigenvalues of the linearized system. The linearized system has the form

$$\frac{d\tilde{\sigma}_1}{d\tau} = -f_1\tilde{\sigma}_1 - f_2\tilde{\sigma}_2, \tag{1.102}$$

$$\frac{d\tilde{\sigma}_2}{d\tau} = \alpha f_1\tilde{\sigma}_1 + (\alpha f_2 - \eta)\tilde{\sigma}_2, \tag{1.103}$$

where $f_j = \frac{\partial f}{\partial \sigma_j}, j = 1, 2$, evaluated at the steady-state solution, and where $\tilde{\sigma}_i$ denotes the deviation from the steady-state value of σ_i. The characteristic equation for the eigenvalues λ of the linear system (1.102)–(1.103) is

$$\lambda^2 - (\alpha f_2 - \eta - f_1)\lambda + f_1\eta = 0. \tag{1.104}$$

Since f_1 is always positive, the stability of the linear system is determined by the sign of $H = \alpha f_2 - \eta - f_1$, being stable if $H < 0$ and unstable if $H > 0$. Changes of stability, if they exist, occur at $H = 0$, and are Hopf bifurcations to periodic solutions with approximate frequency $\omega = \sqrt{f_1\eta}$.

The function $H(\nu)$ is given by

$$H(\nu) = \frac{(1-\nu)}{(1+y)}(\eta\gamma + (\nu-1)y) - \eta, \tag{1.105}$$

$$y = (p\nu)^\gamma. \tag{1.106}$$

Clearly, $H(0) = \eta(\gamma - 1), H(1) = -\eta$, so for $\gamma > 1$, there must be at least one Hopf bifurcation point, below which the steady solution is unstable. Additional computations show that this Hopf bifurcation is supercritical, so that for ν slightly below the bifurcation point, there is a stable periodic orbit.

An example of this periodic orbit is shown in Fig. 1.6 with coefficients $\nu = 0.0285, \eta = 0.1, \alpha = 1.0$, and $\gamma = 2$. The evolution of σ_1 and σ_2 are shown plotted as functions of time in Fig. 1.7. This periodic orbit exists only in very small regions of parameter space, rapidly expanding until it contacts the $S_2 = 0$ axis, into which it collapses.

While the Sel'kov model has certain features that are qualitatively correct, it fails to agree with the experimental results at a number of points. Hess and Boiteux (1973) report that for high and low substrate injection rates, there is a stable steady-state solution. There are two Hopf bifurcation points, one at the flow rate of 20 mM/hr and another at 160 mM/hr. The period of oscillation at the low flow rate is about 8 minutes and decreases as a function of flow rate to about 3 minutes at the upper Hopf bifurcation point. In contrast, the Sel'kov model has but one Hopf bifurcation point.

To reproduce these additional experimental features we consider a more detailed model of the reaction. In 1972, Goldbeter and Lefever proposed a model of Monod–Wyman–Changeux type that provided a more accurate description of the oscillations. More recently, by fitting a simpler model to experimental data on PFK1 kinetics in skeletal muscle, Smolen (1995) has shown that this level of complexity is not necessary; his model assumes that PFK1 consists of four independent, identical

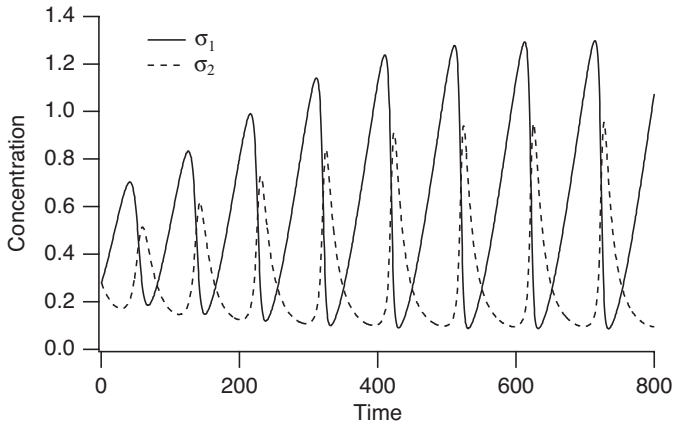

Figure 1.7: Evolution of σ_1 and σ_2 for the Sel'kov glycolysis system toward a periodic solution. Parameters are the same as in Fig. 1.6.

subunits, and reproduces the observed oscillations well. Despite this, we consider only the Goldbeter–Lefever model in detail, as it provides an excellent example of the use of Monod–Wyman–Changeux models.

In the Goldbeter–Lefever model of the phosphorylation of fructose-6-P, the enzyme PFK1 is assumed to be a dimer that exists in two states, an active state R and an inactive state T. The substrate, S_1, can bind to both forms, but the product, S_2, which is an activator, or positive effector, of the enzyme, binds only to the active form. The enzymatic forms of R carrying substrate decompose irreversibly to yield the product ADP. In addition, substrate is supplied to the system at a constant rate, while product is removed at a rate proportional to its concentration. The reaction scheme for this is as follows: let T_j represent the inactive T form of the enzyme bound to j molecules of substrate and let R_{ij} represent the active form R of the enzyme bound to i substrate molecules and j product molecules. Then

$$\mathrm{R}_{00} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \mathrm{T}_0,$$

$$\mathrm{S}_1 + \mathrm{R}_{0j} \underset{k_{-2}}{\overset{2k_2}{\rightleftarrows}} \mathrm{R}_{1j} \xrightarrow{k} \mathrm{R}_{0j} + \mathrm{S}_2, \qquad \mathrm{S}_1 + \mathrm{R}_{1j} \underset{2k_{-2}}{\overset{k_2}{\rightleftarrows}} \mathrm{R}_{2j} \xrightarrow{2k} \mathrm{R}_{1j} + \mathrm{S}_2, \qquad j = 0, 1, 2,$$

$$\mathrm{S}_2 + \mathrm{R}_{00} \underset{k_{-2}}{\overset{2k_2}{\rightleftarrows}} \mathrm{R}_{01}, \qquad \mathrm{S}_2 + \mathrm{R}_{01} \underset{2k_{-2}}{\overset{k_2}{\rightleftarrows}} \mathrm{R}_{02},$$

$$\mathrm{S}_1 + \mathrm{T}_0 \underset{k_{-3}}{\overset{2k_3}{\rightleftarrows}} \mathrm{T}_1, \qquad \mathrm{S}_1 + \mathrm{T}_1 \underset{2k_{-3}}{\overset{k_3}{\rightleftarrows}} \mathrm{T}_2.$$

The possible receptor states are illustrated graphically in Fig. 1.8. In this system, the substrate S_1 holds the enzyme in the inactive state by binding with T_0 to produce T_1 and T_2, while product S_2 holds enzyme in the active state by binding with R_{00} to produce R_{01} and binding with R_{01} to produce R_{02}. There is a factor two in the rates of reaction because a dimer with two available binding sites reacts like twice the same amount of monomer.

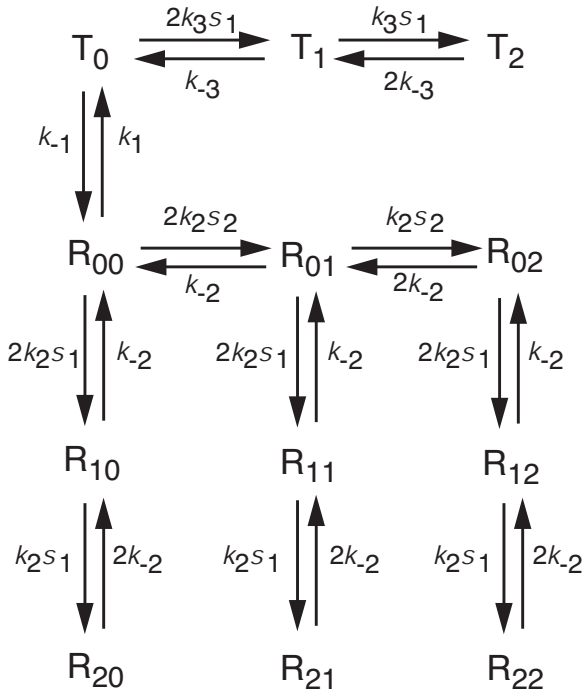

Figure 1.8: Possible receptor states of the Goldbeter–Lefever model for glycolytic oscillations.

The analysis of this reaction scheme is substantially more complicated than that of the Sel'kov scheme, although the idea is the same. We use the law of mass action to write differential equations for the fourteen chemical species. For example, the equation for $s_1 = [\mathrm{S}_1]$ is

$$\frac{ds_1}{dt} = v_1 - F, \tag{1.107}$$

where

$$\begin{aligned} F = \quad & k_{-2}(r_{10} + r_{11} + r_{12}) + 2k_{-2}(r_{20} + r_{21} + r_{22}) \\ & -2k_2 s_1(r_{00} + r_{01} + r_{02}) - k_2 s_1(r_{10} + r_{11} + r_{12}) \\ & -2k_3 s_1 t_0 - k_3 s_1 t_1 + k_{-3} t_1 + 2k_{-3} t_2, \end{aligned} \tag{1.108}$$

and the equation for $r_{00} = [\mathrm{R}_{00}]$ is

$$\frac{dr_{00}}{dt} = -(k_1 + 2k_2 s_1 + 2k_2 s_2) r_{00} + (k_{-2} + k) r_{10} + k_{-2} r_{01} + k_{-1} t_0. \tag{1.109}$$

We then assume that all twelve of the intermediates are in quasi-steady state. This leads to a 12 by 12 linear system of equations, which, if we take the total amount of enzyme to be e_0, can be solved. We substitute this solution into the differential equations for s_1 and s_2 with the result that

$$\frac{ds_1}{dt} = v_1 - F(s_1, s_2), \tag{1.110}$$

$$\frac{ds_2}{dt} = F(s_1, s_2) - v_2 s_2, \tag{1.111}$$

where

$$F(s_1, s_2) = \left(\frac{2k_2 k_{-1} k e_0}{k + k_{-2}} \right) \left(\frac{s_1 (1 + \frac{k_2}{k + k_{-2}} s_1)(k_2 s_2 + k_{-2})^2}{k_{-2}^2 k_1 (\frac{k_3}{k_{-3}} s_1 + 1)^2 + k_{-1}(1 + \frac{k_2}{k + k_{-2}} s_1)^2 (k_{-2} + k_2 s_2)^2} \right). \tag{1.112}$$

Now we introduce dimensionless variables $\sigma_1 = \frac{k_2 s_1}{k_{-2}}, \sigma_2 = \frac{k_2 s_2}{k_{-2}}, t = \frac{\tau}{\tau_c}$ and parameters $\nu = \frac{k_2 v_1}{k_{-2} \tau_c}, \eta = \frac{v_2}{\tau_c}$, where $\tau_c = \frac{2k_2 k_{-1} k e_0}{k_1 (k + k_{-2})}$, and arrive at the system (1.96)–(1.97), but with a different function $f(\sigma_1, \sigma_2)$, and with $\alpha = 1$. If, in addition, we assume that

1. the substrate does not bind to the T form ($k_3 = 0$, T is completely inactive),
2. T_0 is preferred over R_{00} ($k_1 \gg k_{-1}$), and
3. if the substrate S_1 binds to the R form, then formation of product S_2 is preferred to dissociation ($k \gg k_{-2}$),

then we can simplify the equations substantially to obtain

$$f(\sigma_1, \sigma_2) = \sigma_1 (1 + \sigma_2)^2. \tag{1.113}$$

The nullclines for this system of equations are somewhat different from the Sel'kov system, being

$$\sigma_1 = \frac{\nu}{(1+\sigma_2)^2} \qquad \left(\frac{d\sigma_1}{d\tau} = 0\right), \tag{1.114}$$

$$\sigma_1 = \frac{\eta\sigma_2}{(1+\sigma_2)^2} \qquad \left(\frac{d\sigma_2}{d\tau} = 0\right), \tag{1.115}$$

and the unique steady-state solution is given by

$$\sigma_1 = \frac{\nu}{(1+\sigma_2)^2}, \tag{1.116}$$

$$\sigma_2 = \frac{\nu}{\eta}. \tag{1.117}$$

The stability of the steady-state solution is again determined by the characteristic equation (1.104), and the sign of the real part of the eigenvalues is the same as the sign of

$$H = f_2 - f_1 - \eta = 2\sigma_1(1+\sigma_2)^2 - (1+\sigma_2) - \eta, \tag{1.118}$$

evaluated at the steady state (1.114)–(1.115). Equation (1.118) can be written as the cubic polynomial

$$\frac{1}{\eta}y^3 - y + 2 = 0, \qquad y = 1 + \frac{\nu}{\eta}. \tag{1.119}$$

For η sufficiently large, the polynomial (1.119) has two roots greater than 2, say, y_1 and y_2. Recall that ν is the nondimensional flow rate of substrate ATP. To make some correspondence with the experimental data, we assume that the flow rate ν is proportional to the experimental supply rate of glucose. This is not strictly correct, although ATP is produced at about the same rate that glucose is supplied. Accepting this caveat, we see that to match experimental data, we require

$$\frac{y_2 - 1}{y_1 - 1} = \frac{\nu_2}{\nu_1} = \frac{160}{20} = 8. \tag{1.120}$$

Requiring (1.119) to hold at y_1 and y_2 and requiring (1.120) to hold as well, we find numerical values

$$y_1 = 2.08, y_2 = 9.61, \eta = 116.7, \tag{1.121}$$

corresponding to $\nu_1 = 126$ and $\nu_2 = 1005$.

At the Hopf bifurcation point, the period of oscillation is

$$T_i = \frac{2\pi}{\omega_i} = \frac{2\pi}{\sqrt{\eta}(1+\sigma_2)} = \frac{2\pi}{\sqrt{\eta}y_i}. \tag{1.122}$$

For the numbers (1.121), we obtain a ratio of periods $T_1/T_2 = 4.6$, which is acceptably close to the experimentally observed ratio $T_1/T_2 = 2.7$.

A typical phase portrait for the periodic solution that exists between the Hopf bifurcation points is shown in Fig. 1.9, and the concentrations of the two species are shown as functions of time in Fig. 1.10.

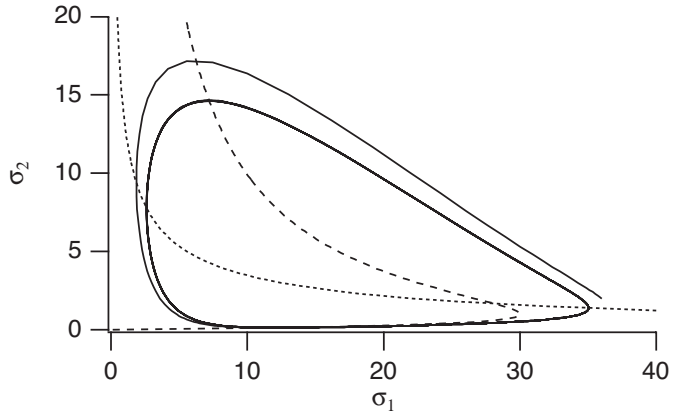

Figure 1.9: Phase portrait of the Goldbeter–Lefever model with $\nu = 200$, $\eta = 120$. Dotted curve: $\frac{d\sigma_1}{d\tau} = 0$. Dashed curve: $\frac{d\sigma_2}{d\tau} = 0$.

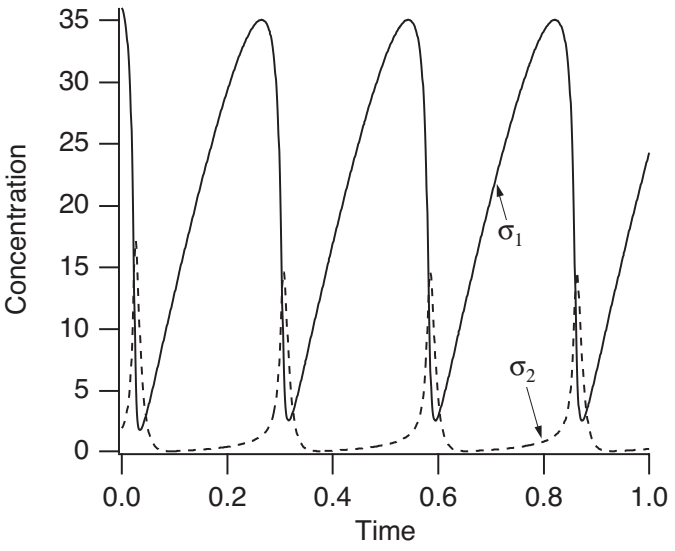

Figure 1.10: Solution of the Goldbeter–Lefever model with $\nu = 200$, $\eta = 120$.

1.6 Appendix: Math Background

It is certain that some of the mathematical concepts and tools that we routinely invoke here are not familiar to all of our readers. In this first chapter alone, we have used nondimensionalization, phase-plane analysis, linear stability analysis, bifurcation theory, and asymptotic analysis, all the while assuming that these are familiar to the reader.

The purpose of this appendix is to give a brief guide to those techniques that are a basic part of the applied mathematician's toolbox but that may not be familiar to all our readers.

1.6.1 Basic Techniques

In any problem, there are a number of parameters that are dictated by the problem. However, it often happens that not all parameters are *independent*; that is, different variations in different parameters may lead to identical changes in the behavior of the model. Second, there may be parameters whose influence on a behavior is negligible and can be safely ignored for a given context.

The way to identify independent parameters and to determine their relative magnitudes is to nondimensionalize the problem. Unfortunately, having said that, we cannot describe a unique algorithm for nondimensionalization, because one does not exist; nondimensionalization is as much art as it is science.

There are, however, rules of thumb to apply. In any system of equations, there are a number of independent variables (time, space, etc.), dependent variables (concentrations, etc.) and parameters (rates of reaction, sizes of containers, etc.). Nondimensionalization begins by rescaling the independent and dependent variables by "typical" units, rendering them thereby dimensionless. One goal may be to ensure that the dimensionless variables remain of a fixed order of magnitude, not becoming too large or negligibly small. This usually requires some a priori knowledge about the solution, as it can be difficult to choose typical scales unless something is already known about typical solutions. Time and space scales can be vastly different depending on the context.

Once this selection of scales has been made, the governing equations are written in terms of the rescaled variables and dimensionless combinations of the remaining parameters are identified. The number of remaining free dimensionless parameters is usually less than the original number of physical parameters. The primary difficulty (at least to understand and apply the process) is that there is not necessarily a single way to scale and nondimensionalize the equations. Some scalings may highlight certain features of the solution, while other scalings may emphasize others. Nonetheless, nondimensionalization often provides a good starting point for the analysis of a model system.

An excellent discussion of scaling and nondimensionalization can be found in Lin and Segel (1988, Chapter 6). A great deal of more advanced work has also been done on this subject, particularly its application to the quasi-steady-state approximation, by Segel and his collaborators (Segel, 1988; Segel and Slemrod, 1989; Segel and Perelson, 1992; Segel and Goldbeter, 1994; Borghans et al., 1996; see also Frenzen and Maini, 1988).

Phase-plane analysis and linear stability analysis are standard fare in introductory courses on differential equations. A nice introduction to these topics for the biologically inclined can be found in Edelstein-Keshet (1988, Chapter 5) or Braun (1993, Chapter 4). A large number of books discuss the qualitative theory of differential equations, for example, Boyce and Diprima (1997), or at a more advanced level, Hale and Koçak (1991), or Hirsch and Smale (1974).

Bifurcation theory is a topic that is gradually finding its way into introductory literature. The most important terms to understand are those of *steady-state* bifurcations, *Hopf* bifurcations, *homoclinic* bifurcations, and *saddle-node* bifurcations, all of which appear in this book. An excellent introduction to these concepts is found in Strogatz (1994, Chapters 3, 6, 7, 8). More advanced treatments include those in Guckenheimer and Holmes (1983), Arnold (1983), or Kuznetzov (199?).

1.6.2 Asymptotic Analysis

Applied mathematicians love small parameters, because of the hope that the solution of a problem with a small parameter might be approximated by an *asymptotic representation*. A commonplace notation has emerged in which ϵ is often the small parameter. An asymptotic representation has a precise mathematical meaning. Suppose that $G(\epsilon)$ is claimed to be an asymptotic representation of $g(\epsilon)$, expressed as

$$g(\epsilon) = G(\epsilon) + O(\phi(\epsilon)). \tag{1.123}$$

The precise meaning of this statement is that there is a constant A such that

$$\left| \frac{g(\epsilon) - G(\epsilon)}{\phi(\epsilon)} \right| \leq A \tag{1.124}$$

for all ϵ with $|\epsilon| \leq \epsilon_0$ and $\epsilon > 0$. The function $\phi(\epsilon)$ is called a *gauge function*, a typical example of which is a power of ϵ.

Perturbation Expansions

It is often the case that an asymptotic representation can be found as a development in powers of the small parameter ϵ. Such representations are called *perturbation expansions*. Usually, a few terms of this power series representation suffice to give a good approximation to the solution. It should be kept in mind that under no circumstances does this power series development imply that a complete power series (with an infinite number of terms) exists or is convergent. Terminating the series at one or two terms is deliberate.

However, there are times when a full power series could be found and would be convergent in some nontrivial ϵ domain. Such problems are called *regular perturbation problems* because their solutions are regular, or analytic, in the parameter ϵ.

There are numerous examples of regular perturbation problems, including all of those related to bifurcation theory. These problems are regular because their solutions can be developed in a convergent power series of some parameter.

There are, however, many problems with small parameters whose solutions are not regular, called *singular perturbation problems*. Singular perturbation problems are characterized by the fact that their dependence on the small parameter is not regular, but *singular*, and their convergence as a function of ϵ is not uniform.

Singular problems come in two basic varieties. Characteristic of the first type is a small region of width ϵ somewhere in the domain of interest (either space or time) in which the solution changes rapidly. For example, the solution of the boundary value problem

$$\epsilon u'' + u' + u = 0 \tag{1.125}$$

subject to boundary conditions $u(0) = u(1) = 1$ is approximated by the asymptotic representation

$$u(x;\epsilon) = (1-e)e^{-x/\epsilon} + e^{1-x} + O(\epsilon). \tag{1.126}$$

Notice the nonuniform nature of this solution, as

$$e = \lim_{x\to 0^+}(\lim_{\epsilon\to 0^+} u(x;\epsilon)) \neq \lim_{\epsilon\to 0^+}(\lim_{x\to 0^+} u(x;\epsilon)) = 1. \tag{1.127}$$

Here the term $e^{-x/\epsilon}$ is a *boundary layer correction*, as it is important only in a small region near the boundary at $x = 0$.

Other terms that are typical in singular perturbation problems are *interior layers* or *transition layers*, typified by expressions of the form $\tan(\frac{x-x_0}{\epsilon})$, and *corner layers*, locations where the derivative changes rapidly but the solution itself changes little. Transition layers are of great significance in the study of excitable systems (Chapter 4). While corner layers show up in this book, we do not study or use them in any detail.

Singular problems of this type can often be identified by the fact that the order of the system decreases if ϵ is set to zero. An example that we have already seen is the quasi-steady-state analysis used to simplify reaction schemes in which some reactions are significantly faster than others. Setting ϵ to zero in these examples reduces the order of the system of equations, signaling a possible problem. Indeed, solutions of these equations typically have *initial layers* near time $t = 0$. We take a closer look at this example below.

The second class of singular perturbation problems is that in which there are two scales in operation everywhere in the domain of interest. Problems of this type show up throughout this book. For example, action potential propagation in cardiac tissue is through a cellular medium whose detailed structure varies rapidly compared to the length scale of the action potential wave front. Physical properties of the cochlear membrane in the inner ear vary slowly compared to the wavelength of waves that propagate along it. For problems of this type, one must make explicit the dependence on multiple scales, and so solutions are often expressed as functions of two variables, say x and x/ϵ, which are treated as independent variables. Solution techniques that exploit the multiple-scale nature of the solution are called *multiscale methods* or *averaging methods*.

Detailed discussions of these asymptotic methods may be found in Murray (1984), Kevorkian and Cole (1996), and Holmes (1995).

1.6.3 Enzyme Kinetics and Singular Perturbation Theory

In most of the examples of enzyme kinetics discussed in this chapter, extensive use was made of the quasi-steady-state approximation (1.43), according to which the concentration of the complex remains constant during the course of the reaction. Although this assumption gives the right answers (which, some might argue, is justification enough), mathematicians have sought for ways to justify this approximation rigorously. Bowen et al. (1963) and Heineken et al. (1967) were the first to show that the quasi-steady-state approximation can be derived as the lowest-order term in an asymptotic expansion of the solution. This has since become one of the standard examples of the application of singular perturbation theory to biological systems, and it is discussed in detail by Rubinow (1973), Lin and Segel (1988), and Murray (1989), among others.

First, note that the quasi-steady-state assumption cannot be correct for all times if one starts the reaction from arbitrary initial concentrations. This is apparent with the Michaelis–Menten kinetics, for example, because the single first-order differential equation (1.42) describes the rate of conversion of substrate into product, but for this to be valid the quasi-steady-state approximation is assumed to hold. There must therefore be a brief period of time at the start of the reaction during which the quasi-steady-state equilibrium does not hold. During this initial time period the enzyme is "filling up" with substrate, until the concentration of complexed enzyme reaches the value given by the quasi-steady-state approximation. Since there is little enzyme compared to the total amount of substrate, the concentration of substrate remains essentially constant during this period.

For most biochemical reactions this transition to the quasi-steady-state happens so fast that it is not physiologically important, but for mathematical reasons, it is interesting to understand these kinetics for early times as well. To see how the reaction runs for early times from arbitrary initial conditions, we make a change of time scale, $\eta = \tau/\epsilon$. This change of variables expands the time scale on which we look at the reaction and allows us to study events that happen on a fast time scale. In the new time scale, the dimensionless reaction equations (1.37)–(1.38) become

$$\frac{d\sigma}{d\eta} = \epsilon(-\sigma + x(\sigma + \alpha)), \tag{1.128}$$

$$\frac{dx}{d\eta} = \sigma - x(\sigma + \kappa). \tag{1.129}$$

For small ϵ, these equations are well approximated by the simplification

$$\frac{d\sigma}{d\eta} = 0, \tag{1.130}$$

$$\frac{dx}{d\eta} = \sigma - x(\sigma + \kappa). \tag{1.131}$$

Simply stated, this means that on this time scale the variable σ does not change, so that $\sigma = 1$. Furthermore, we can solve for x as

$$x = \frac{1}{1+\kappa} + \left(x_0 - \frac{1}{1+\kappa}\right) e^{-(1+\kappa)\eta}. \tag{1.132}$$

If [E] $= e_0$ at time $t = 0$, then $x_0 = 0$. In terms of the original time variable τ, this solution is

$$x(\tau) = \frac{1}{1+\kappa}(1 - e^{-(1+\kappa)\frac{\tau}{\epsilon}}), \tag{1.133}$$

and it is valid only for times of order ϵ. The exponential term is significant only when τ is small of order ϵ. Thus, this simple analysis shows that if the reaction is started from arbitrary initial conditions, there is first a time span during which the enzyme products rapidly equilibrate, consuming little substrate, and after this initial "layer" the reaction proceeds according to Michaelis–Menten kinetics along the quasi-steady-state curve.

In this problem the analysis of the initial layer is relatively easy and not particularly revealing. However, this type of analysis will be of much greater importance later when we discuss the behavior of excitable systems.

Exercises

1. Consider the simple chemical reaction in which two monomers of A combine to form a dimer B, according to

$$\mathrm{A} + \mathrm{A} \underset{k_-}{\overset{k_+}{\rightleftarrows}} \mathrm{B}. \tag{1.134}$$

   (a) Use the law of mass action to find the differential equations governing the rates of production of A and B.

   (b) What quantity is conserved? Use this conserved quantity to find an equation governing the rate of production of A that depends only on the concentration of A.

   (c) Non-dimensionalize this equation and show that these dynamics depend on only one non-dimensional parameter.

2. In the real world trimolecular reactions are rare, although trimerizations are not. Consider the following trimerization reaction in which three monomers of A combine to form the trimer C,

$$\mathrm{A} + \mathrm{A} \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} \mathrm{B},$$
$$\mathrm{A} + \mathrm{B} \underset{k_{-2}}{\overset{k_2}{\rightleftarrows}} \mathrm{C}.$$

   (a) Use the law of mass action to find the rate of production of the trimer C.

   (b) Suppose $k_{-1} \gg k_{-2}, k_2$A. Use the appropriate quasi-steady state approximation to find the rates of production of A and C, and show that the rate of production of C is proportional to $[\mathrm{A}]^3$. Explain in words why this is so.

3. The length of microtubules changes by a process called treadmilling, in which monomer is added to one end of the microtubule and taken off at the other end. To model this process, suppose that monomer A_1 is self-polymerizing in that it can form dimer A_2 via

$$\mathrm{A}_1 + \mathrm{A}_1 \xrightarrow{k_+} \mathrm{A}_2. \tag{1.135}$$

Furthermore, suppose A_1 can polymerize an n-polymer A_n at one end making an $n+1$-polymer A_{n+1}

$$\mathrm{A}_1 + \mathrm{A}_n \xrightarrow{k_+} \mathrm{A}_{n+1}. \tag{1.136}$$

Finally, degradation can occur one monomer at a time from the opposite end at rate k_-. Find the steady state distribution of polymer lengths after an initial amount of monomer A_0 has fully polymerized.

4. Consider an enzymatic reaction in which an enzyme can be activated or inactivated by the same chemical substance, as follows:

$$\mathrm{E} + \mathrm{X} \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{E}_1, \tag{1.137}$$

$$\mathrm{E}_1 + \mathrm{X} \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{E}_2, \tag{1.138}$$

$$\mathrm{E}_1 + \mathrm{S} \xrightarrow{k_3} \mathrm{P} + \mathrm{Q} + \mathrm{E}. \tag{1.139}$$

Suppose further that X is supplied at a constant rate and removed at a rate proportional to its concentration. Use quasi-steady-state analysis to find the nondimensional equation describing the degradation of X,

$$\frac{dx}{dt} = \gamma - x - \frac{\beta x y}{1 + x + y + \frac{\alpha}{\delta} x^2}. \tag{1.140}$$

Identify all the parameters and variables, and the conditions under which the quasi-steady state approximation is valid.

5. Using the quasi-steady-state approximation, show that the velocity of the reaction for an enzyme with an allosteric inhibitor (Section 1.4.3) is given by

$$V = \left(\frac{V_{\max} K_3}{i + K_3}\right)\left(\frac{s(k_{-1} + k_3 i + k_1 s + k_{-3})}{k_1(s + K_1)^2 + (s + K_1)(k_3 i + k_{-3} + k_2) + k_2 k_{-3}/k_1}\right), \tag{1.141}$$

Identify all parameters. Under what conditions on the rate constants is this a valid approximation? Show that this reduces to (1.58) in the case $K_1 = \kappa_1$.

6. (a) Derive the expression (1.75) for the fraction of occupied sites in a Monod–Wyman–Changeux model with n binding sites.

   (b) Modify the Monod–Wyman–Changeux model shown in Fig.1.4 to include transitions between states R_1 and T_1, and between states R_2 and T_2. Use the principle of detailed balance to derive an expression for the equilibrium constant of each of these transitions. Do these transitions change the expression for Y, the fraction of occupied sites?

Table 1.1: Data for Problem 7.

Substrate Concentration (mM)	Reaction Velocity (mM/s)
0.1	0.04
0.2	0.08
0.5	0.17
1.0	0.24
2.0	0.32
3.5	0.39
5.0	0.42

Table 1.2: Data for Problem 8.

Substrate Concentration (mM)	Reaction Velocity (mM/s)
0.2	0.01
0.5	0.06
1.0	0.27
1.5	0.50
2.0	0.67
2.5	0.78
3.5	0.89
4.0	0.92
4.5	0.94
5.0	0.95

7. An enzyme-substrate system is believed to proceed at a Michaelis- Menten rate. Data for the (initial) rate of reaction at different concentrations is shown in Table 1.1.

   (a) Plot the data V vs. s. Is there evidence that this is a Michaelis-Menten type reaction?

   (b) Plot V vs. V/s. Is this data well approximated by a straight line?

   (c) Use linear regression and (1.45) to estimate K_m and $V_{\max}$. Compare the data to the Michaelis-Menten rate function using these parameters. Does this provide a reasonable fit to the data?

8. Suppose the maximum velocity of a chemical reaction is known to be 1 mM/s, and the measured velocity V of the reaction at different concentrations s is shown in Table 1.2.

   (a) Plot the data V vs. s. Is there evidence that this is a Hill type reaction?

   (b) Plot $\ln(\frac{V}{V_{\max}-V})$ vs. $\ln(s)$. Is this approximately a straight line, and if so, what is its slope?

(c) Use linear regression and (1.70) to estimate K_m and the Hill exponent n. Compare the data to the Hill rate function with these parameters. Does this provide a reasonable fit to the data?

9. Find the velocity of reaction for an enzyme with three active sites. Under what conditions does the velocity reduce to a Hill function with exponent three? Identify all parameters.

   (a) What is the relationship between rate constants when the three sites are independent? What is the velocity when the three sites are independent?

10. Suppose that a substrate can be broken down by two different enzymes with different kinetics. (This happens, for example, in the case of cAMP or cGMP, which can be hydrolyzed by two different forms of phosphodiesterase—see Chapter **??**).

    (a) Write down the reaction scheme and differential equations, and nondimensionalize, to get the system of equations

$$\frac{d\sigma}{dt} = -\sigma + \alpha_1(\mu_1 + \sigma)x + \alpha_2(\mu_2 + \sigma)y, \tag{1.142}$$

$$\epsilon_1 \frac{dx}{dt} = \frac{1}{\lambda_1}\sigma(1-x) - x, \tag{1.143}$$

$$\epsilon_2 \frac{dy}{dt} = \frac{1}{\lambda_2}\sigma(1-y) - y. \tag{1.144}$$

    where x and y are the nondimensional concentrations of the two complexes. Identify all parameters.

    (b) Apply the quasi-steady-state approximation to find the equation governing the dynamics of substrate σ. Under what conditions is the quasi-steady state approximation valid?

    (c) Solve the differential equation governing σ.

    (d) For this system of equations, show that the solution can never leave the positive octant $\sigma, x, y \geq 0$. By showing that $\sigma + \epsilon_1 x + \epsilon_2 y$ is decreasing everywhere in the positive octant, show that the solution approaches the origin for large time.

11. For some enzyme reactions (for example, the hydrolysis of cAMP by phosphodiesterase in vertebrate retinal cones) the enzyme is present in large quantities, so that e_0/s_0 is not a small number. Fortunately, there is an alternate derivation of the Michaelis-Menten rate equation that does not require that $\epsilon = \frac{e_0}{s_0}$ be small. Instead, if one or both of k_{-1} and k_2 are much larger than $k_1 e_0$, then the formation of complex c is a rapid exponential process, and can be taken to be in quasi-steady state. Make this argument systematic by introducing appropriate non-dimensional variables and then find the resulting quasi-steady state dynamics. (Segel, 1988; Frenzen and Maini, 1988; Segel and Slemrod, 1989; Sneyd and Tranchina, 1989).

12. ATP is known to inhibit its own dephosphorylation. One possible way for this to occur is if ATP binds with the enzyme, holding it in an inactive state, via

$$\mathrm{S}_1 + \mathrm{E} \underset{k_{-5}}{\overset{k_5}{\rightleftharpoons}} \mathrm{S}_1\mathrm{E}.$$

Add this reaction to the Sel'kov model for glycolysis and derive the equations governing glycolysis of the form (1.96)–(1.97). Explain from the model why this additional reaction is inhibitory.

13. In the case of noncompetitive inhibition, the inhibitor combines with the enzyme-substrate complex to give an inactive enzyme-substrate-inhibitor complex which cannot undergo further reaction, but the inhibitor does not combine directly with free enzyme or affect its reaction with substrate. Use the quasi-steady-state approximation to determine that the velocity of this reaction is

$$V = V_{\max} \frac{s}{K_m + s + \frac{i}{K_i} s}. \tag{1.145}$$

Identify all parameters. Compare this velocity with the velocity for other types of inhibition discussed in the text.

14. The following reaction scheme is a simplified version of the Goldbeter–Lefever reaction scheme:

$$\mathrm{R}_0 \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} \mathrm{T}_0,$$

$$\mathrm{S}_1 + \mathrm{R}_j \underset{k_{-2}}{\overset{k_2}{\rightleftharpoons}} \mathrm{C}_j \xrightarrow{k} \mathrm{R}_j + \mathrm{S}_2, \qquad j = 0, 1, 2,$$

$$\mathrm{S}_2 + \mathrm{R}_0 \underset{k_{-3}}{\overset{2k_3}{\rightleftharpoons}} \mathrm{R}_1, \qquad \mathrm{S}_2 + \mathrm{R}_1 \underset{2k_{-3}}{\overset{k_3}{\rightleftharpoons}} \mathrm{R}_2.$$

Show that, under appropriate assumptions about the ratios k_1/k_{-1} and $\frac{k_{-2}+k_3}{k_2}$ the equations describing this reaction are of the form (1.96)–(1.97) with $f(\sigma_1, \sigma_2)$ given by (1.113).

15. Use the law of mass action and the quasi-steady-state assumption for the enzymatic reactions to derive a system of equations of the form (1.96)–(1.97) for the Goldbeter–Lefever model. Verify (1.112).

16. When much of the ATP is depleted in a cell, a considerable amount of cAMP is formed as a product of ATP degradation. This cAMP activates an enzyme phophorylase that splits glycogen, releasing glucose that is rapidly metabolized, replenishing the ATP supply.

Devise a model for this control loop and determine conditions under which the production of ATP is oscillatory.

17. By looking for solutions to (1.37) and (1.38) of the form

$$\sigma = \sigma_0 + \epsilon\sigma_1 + \epsilon^2\sigma_2 + \cdots, \tag{1.146}$$

$$x = x_0 + \epsilon x_1 + \epsilon^2 x_2 + \cdots, \tag{1.147}$$

show that σ_0 and x_0 satisfy the quasi-steady-state approximation. Thus, the quasi-steady-state approximation is the lowest-order term in an asymptotic expansion for the solution. Typical initial conditions are $\sigma = 1$, $x = 0$. Does the lowest-order solution satisfy the initial conditions? Find σ_1 and x_1 and plot the solution to first order in ϵ. The variables σ and x are called the *outer solution*, as they are valid for times outside some boundary layer around $\tau = 0$. Now rescale time by ϵ, and use the same procedure to construct an asymptotic solution to (1.128) and (1.129), the so-called *inner solution*. Show that the inner solution satisfies the initial conditions. How can one construct a solution that satisfies the initial conditions and is valid for all times?

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 apparently 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 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. There are three passive transport mechanisms to transport molecules through the cell membrane. *Osmosis* is the process by which water is transported through the cell membrane. Simple diffusion accounts for the passage of small molecules through pores and of lipid-soluble molecules through the bilipid layer. For example, water, urea (a nitrogenous waste product of metabolism), and hydrated chloride ions diffuse through membrane pores. Oxygen and carbon dioxide diffuse through the membrane readily because they are soluble in lipids. Sodium and potassium 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

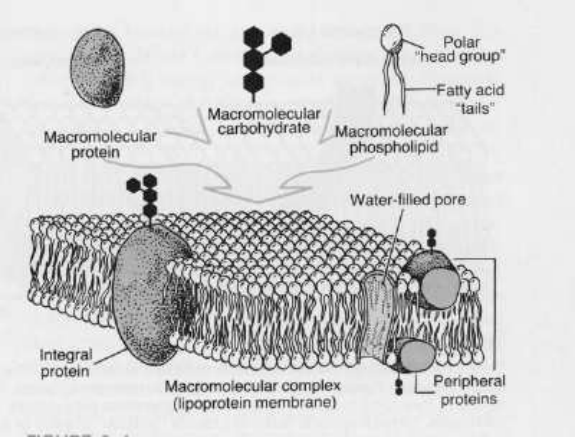

Figure 2.1: Schematic diagram of the cell membrane. (Davis et al., 1985, Fig. 3-1, p. 41.)

pores (Fig. 2.2). *Carrier-mediated diffusion* occurs when a molecule is bound to a carrier molecule that moves readily through the membrane. For example, the transport of glucose and amino acids across the cell membrane is believed to be by a carrier-mediated process.

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.

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 the amount of some chemical species, then the appropriate conservation law takes the following form (in words):

Table 2.1: Typical values for intracellular and extracellular ionic concentrations, 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. Later in this chapter we discuss Nernst potentials and resting potentials. (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

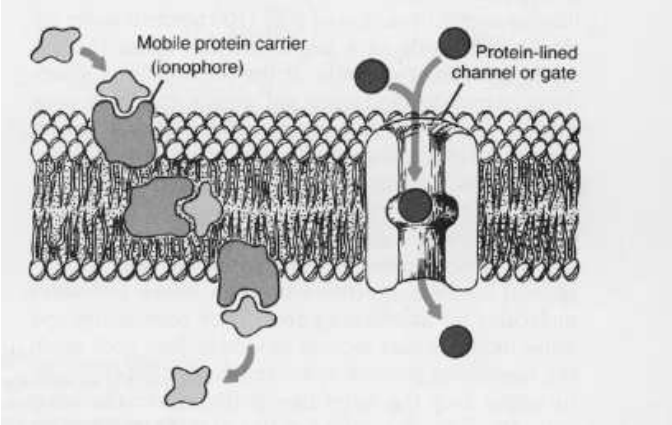

Figure 2.2: Schematic diagram of the cell membrane containing a protein carrier and a protein-lined ionic channel. (Davis et al., 1985, Fig. 3-7, p. 45.)

rate of change of u=local production of u + accumulation of u due to transport.

If Ω is a region of space, then 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 $\partial\Omega$ is the boundary of the region Ω, $\mathbf{n}$ is the outward unit normal to the boundary of Ω, f represents the local production of u per unit volume, and $\mathbf{J}$ is the flux of u. According to the divergence theorem, if $\mathbf{J}$ 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 $\mathbf{J}$ can vary. Indeed, much of our study here is involved in determining appropriate models for production and flux.

2.2.1 Fick's Law

The simplest description of the flux of a chemical species is

$$\mathbf{J} = -D\nabla u. \tag{2.4}$$

Equation (2.4) 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

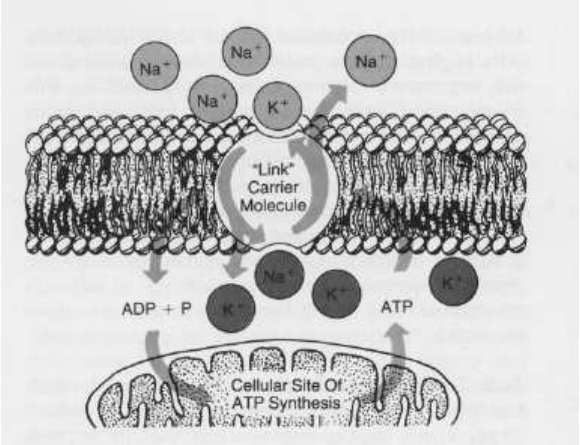

Figure 2.3: Schematic diagram of the cell membrane containing a $Na^+ - K^+$ pump. (Davis, et al., 1985, Fig. 3-11, p. 49.)

the solute and the fluid in which it is dissolved. If u represents the heat content of the volume, (2.4) is called *Newton's law of cooling.* Fick's law is not really a law, but is a reasonable 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

$$\frac{\partial u}{\partial t} = \nabla \cdot (D \nabla u) + f, \tag{2.5}$$

or, if D is a constant,

$$\frac{\partial u}{\partial t} = D \nabla^2 u + f. \tag{2.6}$$

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

$$D = \frac{kT}{6\pi\mu a}, \tag{2.7}$$

where $k = \frac{R}{N_A}$ is Boltzmann's constant, N_A is Avagadro'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

$$D = \frac{kT}{f}, \tag{2.8}$$

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}

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

$$M = \frac{4}{3}\pi a^3 \rho, \tag{2.9}$$

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

$$D = \frac{kT}{3\mu}\left(\frac{\rho}{6\pi^2 M}\right)^{1/3}. \tag{2.10}$$

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 $M^{-1/2}$.

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)

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2}, \tag{2.11}$$

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 find

$$c(x) = c_l + (c_r - c_l)\frac{x}{L}. \tag{2.12}$$

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

$$J = \frac{D}{L}(c_l - c_r). \tag{2.13}$$

Note that a flux from left to right is counted as a positive flux. The ratio L/D is the effective "resistance" (per unit area) 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-section 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

$$\frac{\partial c}{\partial t} = D\frac{\partial^2 c}{\partial x^2} \qquad 0 < x < \infty, \qquad t > 0, \tag{2.14}$$

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, $C(x,0) = 0$.

The solution of this problem is given by

$$C(x,t) = 2C_0\left(1 - \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z}\exp\left(-\frac{s^2}{2}\right)ds\right), \qquad z = \frac{x}{\sqrt{2Dt}}. \tag{2.15}$$

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

$$N = A\int_0^\infty C(x,T)dx = 2C_0A\sqrt{\frac{tD}{\pi}}. \tag{2.16}$$

From this equation it is possible to determine the diffusion coefficient by solving (2.16) for D, yielding

$$D = \frac{\pi N^2}{4C_0^2A^2T}. \tag{2.17}$$

A second useful piece of information is found from (2.15) 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

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
1μm	1 ms	size of mitochondrion
10μm	100 ms	Radius of small mammalian cell
100μm	10s	Diameter of a large muscle fiber
250μm	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

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.

An important situation in which reaction and diffusion interact to modify the behavior is when a diffusing species is buffered by a larger diffusing molecule. This occurs, for example, with oxygen in muscle (which we discuss below), or calcium, or hydrogen ions.

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

$$\text{H} + \text{B} \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \text{HB}. \tag{2.18}$$

Conservation implies

$$\frac{\partial u}{\partial t} = D_h \frac{\partial^2 u}{\partial x^2} + k_- w - k_+ uv + f(t,x,u) \tag{2.19}$$

$$\frac{\partial v}{\partial t} = D_b \frac{\partial^2 v}{\partial x^2} + k_- w - k_+ uv \tag{2.20}$$

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

where $u = [H^+]$, $v = [B]$, and $w = [HB]$. Since the buffer is a large molecule we assume that the diffusion of B and HB is the same. 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 buffer is uniformly distributed). The reaction term $f(t, x, u)$ denotes all the other reactions of u apart from the buffering.

Adding the equations (2.20) and (2.21) we find

$$\frac{\partial(v+w)}{\partial t} = D_b \frac{\partial^2(v+w)}{\partial x^2}. \tag{2.22}$$

Since $v + w$ is initially uniform, it remains uniform for all time, so that $v + w = w_0$.

If the buffering reaction is fast compared to the other reactions (i.e., those described by $f(t, x, u)$ that we are not specifiying explicitly) then we can assume u and v to be in quasi-equilibrium, so that

$$k_-(w_0 - v) - k_+ uv = 0, \tag{2.23}$$

implying that

$$v = \frac{K_{\text{eq}} w_0}{K_{\text{eq}} + u}, \qquad K_{\text{eq}} = \frac{k_-}{k_+}. \tag{2.24}$$

Subtracting (2.20) from (2.19) yields

$$\frac{\partial(u-v)}{\partial t} = D_h \frac{\partial^2 u}{\partial x^2} - D_b \frac{\partial^2 v}{\partial x^2} + f(t, x, u). \tag{2.25}$$

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

$$\frac{\partial}{\partial t}\left(u - \frac{K_{\text{eq}} w_0}{K_{\text{eq}} + u}\right) = D_h \frac{\partial^2 u}{\partial x^2} - D_b \frac{\partial^2}{\partial x^2}\left(\frac{K_{\text{eq}} w_0}{K_{\text{eq}} + u}\right) + f(t, x, u). \tag{2.26}$$

We expand some of the derivatives and find

$$\left(1 + \frac{K_{\text{eq}} w_0}{(K_{\text{eq}} + u)^2}\right) u_t = D_h \frac{\partial^2 u}{\partial x^2} + D_b \frac{\partial}{\partial x}\left(\frac{K_{\text{eq}} w_0}{K_{\text{eq}} + u} u_x\right) + f(t, x, u). \tag{2.27}$$

Now suppose that $u \ll K_{\text{eq}}$. In this limit we find that

$$\left(1 + \frac{w_0}{K_{\text{eq}}}\right) u_t = D_h u_{xx} + \frac{w_0}{K_{\text{eq}}} D_b u_{xx} + f(t, x). \tag{2.28}$$

Of course, this implies that u has an effective diffusion coefficient

$$D_{\text{eff}} = \frac{D_h + D_b \frac{w_0}{K_{\text{eq}}}}{1 + \frac{w_0}{K_{\text{eq}}}}, \tag{2.29}$$

a convex linear combination of the two diffusion coefficients, D_h and D_b.

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.

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 oxygen is much slower when it is buffered by myoglobin since the effective diffusion coefficient of oxygen is lowered substantially by diffusion.

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 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

$$\frac{\partial s}{\partial t} = D_s \frac{\partial^2 s}{\partial x^2} - f, \tag{2.30}$$

$$\frac{\partial e}{\partial t} = D_e \frac{\partial^2 e}{\partial x^2} - f, \tag{2.31}$$

$$\frac{\partial c}{\partial t} = D_c \frac{\partial^2 c}{\partial x^2} + f. \tag{2.32}$$

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 **??**), we assume that the reaction of oxygen with myoglobin is governed by the elementary reaction

$$\text{O}_2 + \text{Mb} \underset{k_-}{\overset{k_+}{\rightleftarrows}} \text{MbO}_2,$$

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 (2.31) is superfluous.

At steady state,

$$0 = s_t + c_t = D_s s_{xx} + D_c c_{xx}, \tag{2.33}$$

and thus there is a second conserved quantity, namely

$$D_s \frac{ds}{dx} + D_c \frac{dc}{dx} = -J, \tag{2.34}$$

which follows by integrating (2.33) 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 (2.34) 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

$$J = \frac{D_s}{L}(s_0 - s_L) + \frac{D_c}{L}(c_0 - c_L), \tag{2.35}$$

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 (2.30) and (2.32) become

$$\epsilon_1 \sigma_{yy} = \sigma(1-u) - u = -\epsilon_2 u_{yy}, \tag{2.36}$$

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

$$c = e_0 \frac{s}{K + s}, \tag{2.37}$$

where $K = k_-/k_+$. Now we substitute (2.37) into (2.35) to find the flux

$$\begin{aligned} J &= \frac{D_s}{L}(s_0 - s_L) + \frac{D_c}{L} e_0 \left(\frac{s_0}{K + s_0} - \frac{s_L}{K + s_L} \right) && (2.38) \\ &= \frac{D_s}{L}(s_0 - s_L)\left(1 + \frac{D_c}{D_s} \frac{e_0 K}{(s_0 + K)(s_L + K)} \right) && (2.39) \\ &= \frac{D_s}{L}(1 + \mu\rho)(s_0 - s_L), && (2.40) \end{aligned}$$

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

$$\begin{aligned} \sigma(y) + \rho u(y) &= y[\sigma(1) + \rho u(1)] + (1-y)[\sigma(0) + \rho u(0)], && (2.41) \\ u(y) &= \frac{\sigma(y)}{1 + \sigma(y)}. && (2.42) \end{aligned}$$

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 (2.13). 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

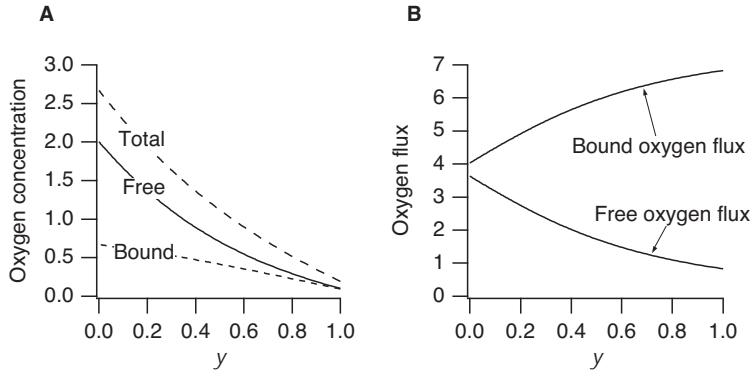

Figure 2.4: A: Free oxygen content $\sigma(y)$ and bound oxygen content $u(y)$ as a function of y. B: Free oxygen flux $-\sigma'(y)$ and bound oxygen flux $-\rho u'(y)$ plotted as a function of y.

$\rho = 560$. 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 (2.38) 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 myoglobin is free to diffuse, the total flux is thereby increased.

In Fig. 2.4A 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. 2.4B, 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 (2.37) as an approximation of (2.36). Usually, when one 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 the 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 still small enough to warrant the approximation that the quasi-steady state (2.37) 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.

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. 2.5. 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. 2.6, 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, $\rho = 5$, $\gamma/\epsilon_1 = 14$.

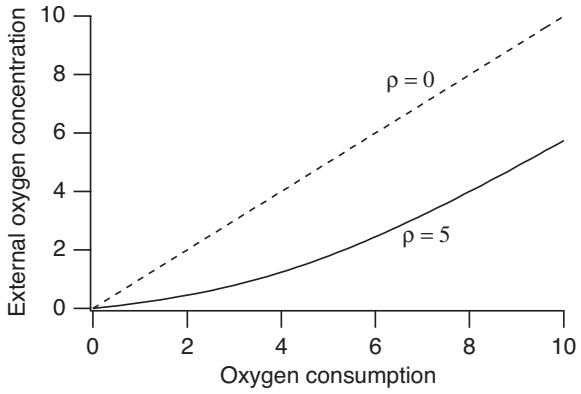

Figure 2.5: Critical concentration σ_0 plotted as a function of oxygen consumption $\frac{\gamma}{4\epsilon_1}$. The dashed curve is the critical concentration with no facilitated diffusion.

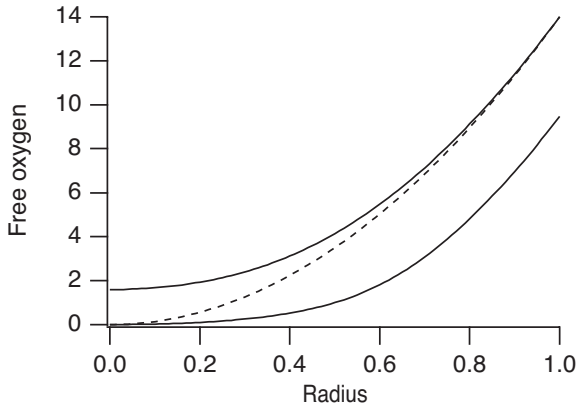

Figure 2.6: Free oxygen σ as a function of radius y. Solid curves show oxygen concentration in the presence of myoglobin ($\rho = 5$), the lower of the two having the critical external oxygen concentration. The dashed curve shows the oxygen concentration without facilitation at the critical external concentration level.

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 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. 2.7.

We now suppose that glucose (i.e., S) is supplied at the constant rate J on the exterior and taken away at the same rate from the interior. Following mass action kinetics, the differential equations describing these kinetics are

$$\frac{ds_i}{dt} = k_- p_i - k_+ s_i c_i - J, \tag{2.49}$$

$$\frac{ds_e}{dt} = k_- p_e - k_+ s_e c_e + J, \tag{2.50}$$

$$\frac{dp_i}{dt} = k p_e - k p_i + k_+ s_i c_i - k_- p_i, \tag{2.51}$$

$$\frac{dp_e}{dt} = k p_i - k p_e + k_+ s_e c_e - k_- p_e, \tag{2.52}$$

$$\frac{dc_i}{dt} = k c_e - k c_i + k_- p_i - k_+ s_i c_i, \tag{2.53}$$

$$\frac{dc_e}{dt} = k c_i - k c_e + k_- p_e - k_+ s_e c_e. \tag{2.54}$$

where $s_i = [S_i]$, $p_i = [P_i]$, etc. There are two degeneracies in these equations; first,

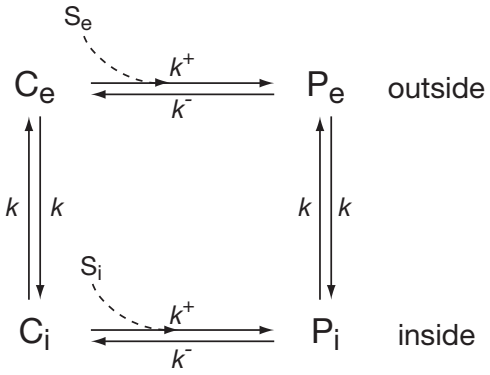

Figure 2.7: Schematic diagram of the glucose transporter described by (2.49-2.54). Note that, since S_e is required for the transition $C_e \rightarrow P_e$, the reverse transition must release S_e, although this is not indicated explicitly in the diagram.

the total amount of receptor is conserved, and thus $p_i + p_e + c_i + c_e = C_0$, where C_0 is a constant (the total transporter concentration), and second, the total amount of glucose is conserved, and thus $s_i + s_e + p_i + p_e =$ constant.

We find the steady state flux by setting all derivatives to zero and solving the resulting algebraic system. However, it is useful to break this solution process into two steps. First, find the steady solutions for p_i, p_e, c_i, c_e using any three of the last four equations along with the constraint $p_i + p_e + c_i + c_e = C_0$. Then, the flux is

$$J = k_- p_i - k_+ s_i c_i = k_+ s_e c_e - k_- p_e. \tag{2.55}$$

Following this procedure, we find J as a function of s_i and s_e to be

$$J = \frac{1}{2} K_d K k_+ C_0 \frac{s_e - s_i}{(s_i + K + K_d)(s_e + K + K_d) - K_d^2}, \tag{2.56}$$

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

$$j = \frac{\sigma_e - \sigma_i}{(\sigma_i + 1 + \kappa)(\sigma_e + 1 + \kappa) - \kappa^2}, \tag{2.57}$$

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. 2.8, 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 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.

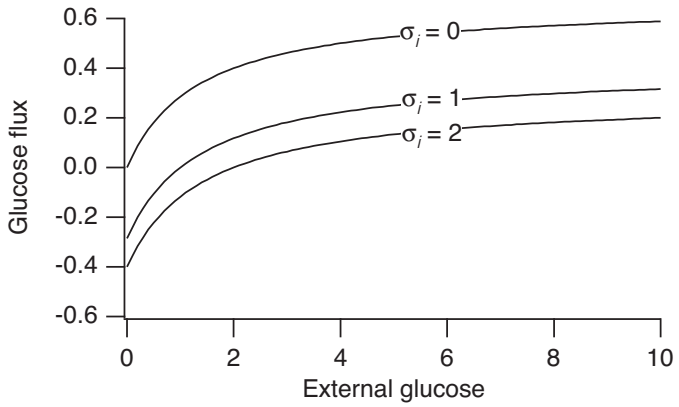

Figure 2.8: Plot of the (nondimensional) flux of glucose as a function of extracellular glucose, for fixed intracellular glucose, with $\kappa = K_d/K = 0.5$.

2.4.2 Symports and Antiports

Models for 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 sodium-driven glucose symport that transports glucose and sodium 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 **??**). 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,

$$\mathrm{C}_i \underset{k_{-c}}{\overset{k_c}{\rightleftharpoons}} \mathrm{C}_e, \qquad \mathrm{P}_i \underset{k_{-p}}{\overset{k_p}{\rightleftharpoons}} \mathrm{P}_e. \tag{2.58}$$

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. 2.9.

Unfortunately, the analysis of this fully general reaction scheme is quite complicated. However, it simplifies significantly if we assume that the intermediates can

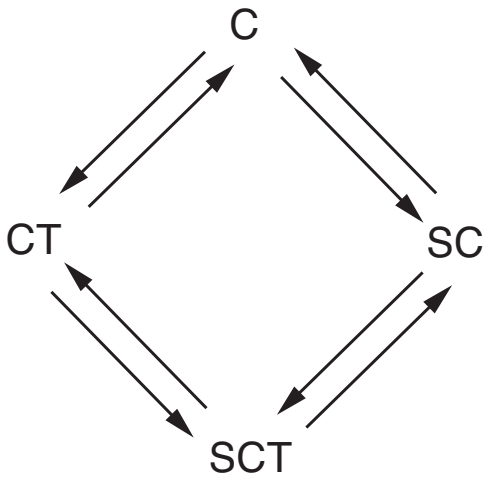

Figure 2.9: States and possible transitions of a transporter with two substrates, S and T, and one binding site for each.

be safely ignored and postulate the reaction scheme

$$m\mathrm{S} + n\mathrm{T} + \mathrm{C} \underset{k_-}{\overset{k_+}{\rightleftharpoons}} \mathrm{P}. \tag{2.59}$$

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

$$J = \frac{1}{2} K_d K k_+ C_0 \frac{s_e^m t_e^n - s_i^m t_i^n}{(s_i^m t_i^n + K + K_d)(s_e^m t_e^n + K + K_d) - K_d^2}, \tag{2.60}$$

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 $K_d = k/k_+$.

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

$$J = \frac{1}{2} K_d K k_+ C_0 \frac{s_e^m t_i^n - s_i^m t_e^n}{(s_i^m t_e^n + K + K_d)(s_e^m t_i^n + K + K_d) - K_d^2}. \tag{2.61}$$

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

$$\left(\frac{s_e}{s_i}\right)^m > \left(\frac{t_e}{t_i}\right)^n. \tag{2.62}$$

For example, for the sodium-calcium exchanger (which we discuss in more detail in the next section) for which $m = 1$ and $n = 3$, a ratio of extracellular to intracellular sodium of about 8 can be used to effectively pump calcium out of a cell even when the ratio of extracellular to intracellular calcium is 500.

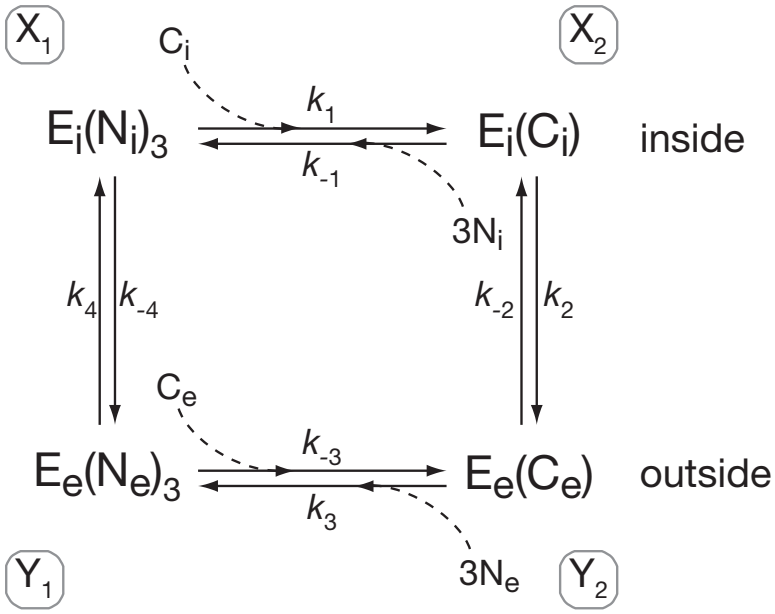

Figure 2.10: Schematic diagram of a simple model of the Na^+-Ca^{2+} exchanger.

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 (cf. Chapter **??**). It has thus been studied extensively, and a number of highly detailed models have been constructed (refs). Here we describe a simple model of this important transporter.

In our model (see Fig. 2.10), E_i is the exchanger protein in the conformation where the binding sites are exposed to the interior of the cell, and E_e is the conformation where 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 3 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 3 Na^+ ions bind or unbind the exchanger simulta-

neously. Notice that the difference between this model and the previous model of a symport/antiport is that here the exchanger protein is prohibited from changing its conformation if its binding sites are empty.

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 differential equations; three equations for exchanger states and one conservation equation. These are

$$\begin{aligned}
\frac{dx_1}{dt} &= k_{-1}n_i^3 x_2 + k_4 y_1 - (k_1 c_i + k_{-4})x_1, && (2.63)\\
\frac{dx_2}{dt} &= k_{-2}y_2 + k_1 c_i x_1 - (k_2 + k_{-1}n_i^3)x_2, && (2.64)\\
\frac{dy_1}{dt} &= k_{-4}x_1 + k_3 n_e^3 y_2 - (k_4 + k_{-3}c_e)y_1, && (2.65)\\
1 &= x_1 + x_2 + y_1 + y_2. && (2.66)
\end{aligned}$$

Here, c and n denote, respectively, Ca^{2+} and Na^+ concentration, and the subscripts e or i represent external or internal concentrations. Using a symbolic package such as Maple, the steady-state solution of these is easily calculated. The flux, J, is found to be

$$\begin{aligned}
J &= k_4 y_1 - k_{-4}x_1\\
&= \frac{k_1 k_2 k_3 k_4 (c_i n_e^3 - K_1 K_2 K_3 K_4 c_e n_i^3)}{16 \text{ positive terms}}. && (2.67)
\end{aligned}$$

An electrogenic exchanger

An important difference between the sodium-calcium exchange process and the transport processes discussed previously is that sodium and calcium 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 called *electrogenic*.

As is discussed in Section 2.6, most 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.,

$$L_i \rightarrow L_e. \tag{2.68}$$

The change in chemical potential for this reaction is

$$\begin{aligned}
\Delta G &= G^0_{L_e} + RT\ln([L_e]) + zFV_e - G^0_{L_i} - RT\ln([L_i]) - zFV_i, && (2.69)\\
&= RT\ln\left(\frac{[L_e]}{[L_i]}\right) - zFV, && (2.70)
\end{aligned}$$

where $V = V_i - V_e$ is the transmembrane potential. 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

$$K = \frac{[\mathrm{L}_i]_{\mathrm{eq}}}{[\mathrm{L}_e]_{\mathrm{eq}}} = \exp\left(\frac{-zFV}{RT}\right), \tag{2.71}$$

where K is the equilibrium constant for the reaction.

The overall transporter reaction begins with 3 Na^+ outside the cell and 1 Ca^{2+} inside the cell, and ends with 3 Na^+ inside the cell and 1 Ca^{2+} outside. We can write this as

$$3\mathrm{Na}_e^+ + \mathrm{Ca}_i^{2+} \longrightarrow 3\mathrm{Na}_i^+ + \mathrm{Ca}_e^{2+}. \tag{2.72}$$

The change in chemical potential for this reaction is

$$\Delta G = RT \ln\left(\frac{n_i^3 c_e}{n_e^3 c_i}\right) + FV. \tag{2.73}$$

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

$$\frac{n_{i,\mathrm{eq}}^3 c_{e,\mathrm{eq}}}{n_{e,\mathrm{eq}}^3 c_{i,\mathrm{eq}}} = \exp\left(-\frac{FV}{RT}\right). \tag{2.74}$$

Recall that detailed balance requires that at equilibrium, 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

$$k_1 c_{i,\mathrm{eq}} k_2 k_3 n_{e,\mathrm{eq}}^3 k_4 = n_{i,\mathrm{eq}}^3 k_{-1} k_{-4} c_{e,\mathrm{eq}} k_{-3} k_{-2}, \tag{2.75}$$

and thus

$$K_1 K_2 K_3 K_4 = \frac{c_{i,\mathrm{eq}}}{c_{e,\mathrm{eq}}} \frac{n_{e,\mathrm{eq}}^3}{n_{i,\mathrm{eq}}^3}. \tag{2.76}$$

Combining (2.75) and (2.76), we get

$$K_1 K_2 K_3 K_4 = \exp\left(\frac{FV}{RT}\right), \tag{2.77}$$

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

It follows from (2.67) that the flux is given by

$$J = \frac{k_1 k_2 k_3 k_4 (c_i n_e^3 - e^{\frac{FV}{RT}} c_e n_i^3)}{16 \text{ positive terms}}. \tag{2.78}$$

Note that because of (2.79), the denominator of this expression contains terms that depend on the membrane potential difference. Note also that all the terms in the denominator are cubic products of rate constants, so that the flux J has units of a rate constant.

In writing down (2.77) we have made no assumption about where the charge transfer takes place. From Fig. 2.10 it might look like the charge transfer takes place during the transitions $\mathrm{Y}_1 \rightarrow \mathrm{X}_1$ and $\mathrm{X}_2 \rightarrow \mathrm{Y}_2$. However, this is not necessarily true. Those conformational changes could be accompanied by no actual charge transfer, which could all occur during the other transitions. However, commonly

one assumes that one Ca^{2+} ion is transferred from inside to outside during the $X_2 \to Y_2$ transition, with three Na^+ ions being transferred during the $Y_1 \to X_1$ transition. With these additional modeling assumptions we would have the further constraints

$$\frac{k_{-2}}{k_2} = \exp\left(\frac{-2FV}{RT}\right), \qquad \frac{k_4}{k_{-4}} = \exp\left(\frac{-3FV}{RT}\right), \tag{2.79}$$

and thus $K_1K_3 = 1$.

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 sodium is 8, and the potential difference $V = -85$ mV (which is typical), calcium is removed provided

$$\frac{c_i}{c_e} > \frac{n_i^3}{n_e^3} e^{\frac{FV}{RT}} = 7 \times 10^{-5}. \tag{2.80}$$

Notice that the difference in potential gives an improvement in the capability of the pump by a factor of 27 over a pump 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 and not susceptible to export. 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. As we saw, this is the mechanism for the sodium-calcium exchanger as well as numerous other exchangers that use to advantage the energy stored in the sodium 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 sodium–potassium ATPase pump (Na,K-ATPase). This pump acts

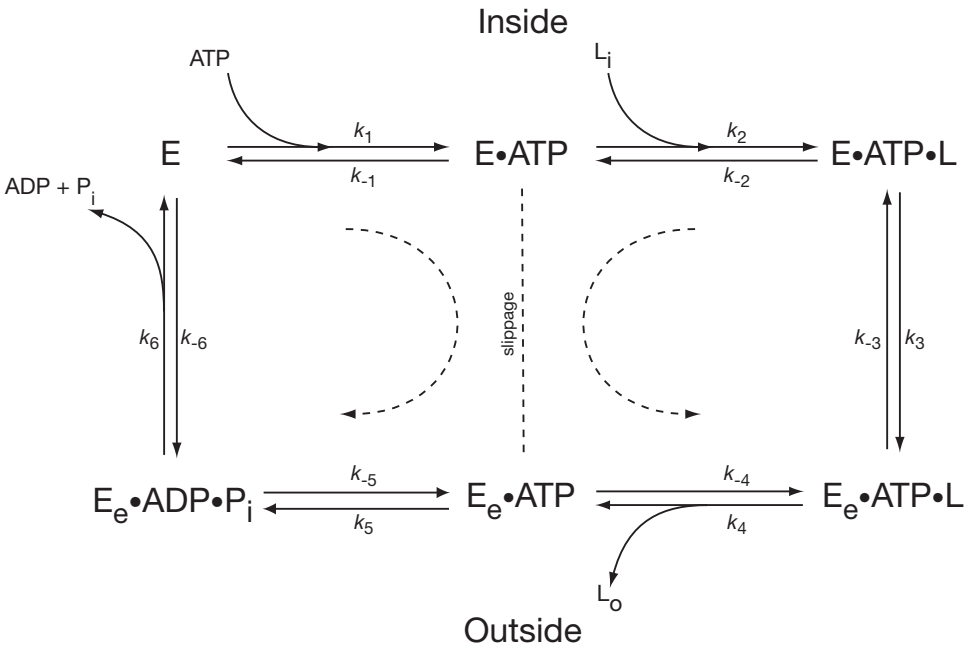

Figure 2.11: Schematic diagram of an ATPase pump that transports one ligand, L, from the inside to the outside against its concentration gradient. For each L transported, one molecule of ATP is hydrolysed. A subscript e denotes the ATPase conformation in which the L binding sites are exposed to the exterior of the cell.

as an antiport, actively pumping sodium ions out of the cell and pumping potassium 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 we see later in this chapter, the Na,K-ATPase is used to regulate the cell volume and to maintain 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 a simple ATPase that pumps a ligand, L, up its concentration gradient (Fig. 2.11). 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. 2.11. 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, exposes the L binding site to the outside, and at the same time the affinity of L for its binding site is decreased. 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. 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 sub-cycles as indicated. This is called *slippage*, as each of the sub-cycles accomplishes nothing toward the goal of pumping L. The sub-cycle on the left goes naturally in a clockwise direction and hydrolyses ATP to ADP and inorganic phosphate, P_i, without using this energy to pump L. Similarly, the sub-cycle 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 whole cycle.

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

$$\frac{d[\mathrm{E}]}{dt} = k_{-1}[\mathrm{E}\cdot\mathrm{ATP}] + k_6[\mathrm{E}_e\cdot\mathrm{ADP}\cdot\mathrm{P}_i] - (k_1[\mathrm{ATP}] + k_{-6}[\mathrm{P}_i][\mathrm{ADP}])[\mathrm{E}], \quad (2.81)$$

with similar equations for each of the other ATPase states. Using the steady state solution, the steady-state flux, J, is calculated from

$$J = k_1[\mathrm{ATP}][\mathrm{E}] - k_{-1}[\mathrm{E}\cdot\mathrm{ATP}]. \quad (2.82)$$

(Ways to calculate steady states and fluxes in a model of this type, for example, the method of King and Altman (1956), are discussed in the book by Hill (1989). However, these methods were developed before programs like Maple which are now the preferred method for these kinds of calculations.)

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

$$J = \frac{\frac{[\mathrm{ATP}][\mathrm{L}_i]}{[\mathrm{ADP}][\mathrm{P_i}][\mathrm{L}_e]} - K_1K_2K_3K_4K_5K_6}{\phi}, \quad (2.83)$$

where $\phi > 0$ is a complicated function of the rate constants and concentrations, and where, as usual, $K_i = k_{-i}/k_i$.

Detailed balance requires

$$k_1[\mathrm{ATP}]_{\mathrm{eq}}k_2[\mathrm{L}_i]_{\mathrm{eq}}k_3k_4k_5k_6 = k_{-1}k_{-2}k_{-3}k_{-4}[\mathrm{L}_e]_{\mathrm{eq}}k_{-5}k_{-6}[\mathrm{ADP}]_{\mathrm{eq}}[\mathrm{P}_i]_{\mathrm{eq}}, \quad (2.84)$$

so that

$$\prod_{i=1}^{6} K_i = \frac{[\mathrm{L}_i]_{\mathrm{eq}}}{[\mathrm{L}_e]_{\mathrm{eq}}}\frac{[\mathrm{ATP}]_{\mathrm{eq}}}{[\mathrm{ADP}]_{\mathrm{eq}}[\mathrm{P}_i]_{\mathrm{eq}}}. \quad (2.85)$$

It follows immediately that

$$J = \frac{\frac{[\mathrm{L}_i]}{[\mathrm{L}_e]}\frac{[\mathrm{ATP}]}{[\mathrm{ADP}][\mathrm{P}_i]} - \frac{[\mathrm{L}_i]_{\mathrm{eq}}}{[\mathrm{L}_e]_{\mathrm{eq}}}\frac{[\mathrm{ATP}]_{\mathrm{eq}}}{[\mathrm{ADP}]_{\mathrm{eq}}[\mathrm{P}_i]_{\mathrm{eq}}}}{\phi}. \quad (2.86)$$

We see from the numerator that the flux is either positive or negative depending on how far the concentrations of L, ATP, ADP and P_i are from their equilibrium concentrations. In general, [ATP] is much higher than its equilibrium concentration (due to the other processes in the cell that are continually generating ATP) and it is this that causes a positive pump flux, pumping L against its gradient. However, if $[L_e]$ 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 observe that the overall reaction is

$$L_i + \text{ATP} \longrightarrow L_e + \text{ADP} + P_i. \tag{2.87}$$

The change in free energy is given by

$$\begin{aligned} \Delta G &= G_{\text{ADP}} + G_{\text{P}} + G_{L_e} - G_{\text{ATP}} - G_{L_i} && (2.88)\\ &= G^0_{\text{ADP}} + G^0_{\text{P}} - G^0_{\text{ATP}} - RT \ln\left(\frac{[\text{ATP}]L_i]}{[\text{ADP}][P_i]L_e]}\right) && (2.89)\\ &= \Delta G^0_{\text{ATP}} - RT \ln\left(\frac{[\text{ATP}]L_i]}{[\text{ADP}][P_i]L_e]}\right). && (2.90) \end{aligned}$$

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

$$\Delta G^0_{\text{ATP}} = RT \ln\left(\frac{[L_i]_{\text{eq}}}{[L_e]_{\text{eq}}} \frac{[\text{ATP}]_{\text{eq}}}{[\text{ADP}]_{\text{eq}}[P_i]_{\text{eq}}}\right). \tag{2.91}$$

Combining this with (2.85) gives

$$\prod_{i=1}^{6} K_i = e^{\frac{\Delta G^0_{\text{ATP}}}{RT}}. \tag{2.92}$$

Finally, from (2.83) it follows that

$$J = \frac{\frac{[L_i]}{[L_e]} \frac{[\text{ATP}]}{[\text{ADP}][P_i]} - e^{\frac{\Delta G^0_{\text{ATP}}}{RT}}}{\phi}. \tag{2.93}$$

Since the free energy released by the hydrolysis of ATP is well known to be -31 kJ mole^{-1}, (2.92) provides a constraint on the rate constants, analogous to the constraint provided by detailed balance (see Section 1.3) but now involving the change in free energy. Of course, if there is no change in free energy, $\Delta G = 0$, in which case we recover $\prod_{i=1}^{6} K_i = 1$, which is just the usual constraint due to detailed balance.

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

$$\Delta G = \Delta G^0_{\text{ATP}} - RT \ln\left(\frac{[\text{ATP}]L_i]}{[\text{ADP}][P_i]L_e]}\right) - zFV, \tag{2.94}$$

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

An identical argument to before gives that

$$\prod_{i=1}^{6} K_i = e^{\frac{\Delta^0 G_{\mathrm{ATP}}}{RT}} e^{\frac{-zFV}{RT}}, \tag{2.95}$$

and thus

$$J = \frac{\frac{[\mathrm{L}_i]}{[\mathrm{L}_e]} \frac{[\mathrm{ATP}]}{[\mathrm{ADP}][\mathrm{P}_i]} - e^{\frac{\Delta^0 G_{\mathrm{ATP}}}{RT}} e^{\frac{-zFV}{RT}}}{\phi(V)}. \tag{2.96}$$

From this we see that if $z > 0$, a positive V enhances the flow (i.e. increases J), and similarly $V < 0$ decreases J by making it more difficult to move positive ions from inside to outside.

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 $\mathrm{E} \cdot \mathrm{ATP} \cdot \mathrm{L}$ to $\mathrm{E}_e \cdot \mathrm{ATP} \cdot \mathrm{L}$ 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. With different models there can be no change in the sign of the flux, although the voltage dependence of the function $\phi(V)$ is strongly model dependent.

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

$$\mathrm{ATP} + 3\mathrm{Na}_i^+ + 2\mathrm{K}_e^+ \longrightarrow \mathrm{ADP} + \mathrm{P} + 3\mathrm{Na}_e^+ + 2\mathrm{K}_i^+. \tag{2.97}$$

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 we discuss at length in Chapter **??**. A great deal of work has been done to discover the mechanisms that underly Na^+ and K^+ transport by this ATPase; the most widely accepted model is the Post-Albers model that was developed by two independent groups in the 1960s (Albers *et al.*, 1963; Charnock and Post, 1963). More recent reviews can be found in Apell (2004) or Taniguchi and Kaya (2000), while a history of investigations into the Na,K-ATPase is given by Glynn (2002). 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 it is accessible from neither side of the membrane. Occlusion prevents slippage, thus increasing the efficiency of the pump.

This is illustrated in Fig. 2.12. 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

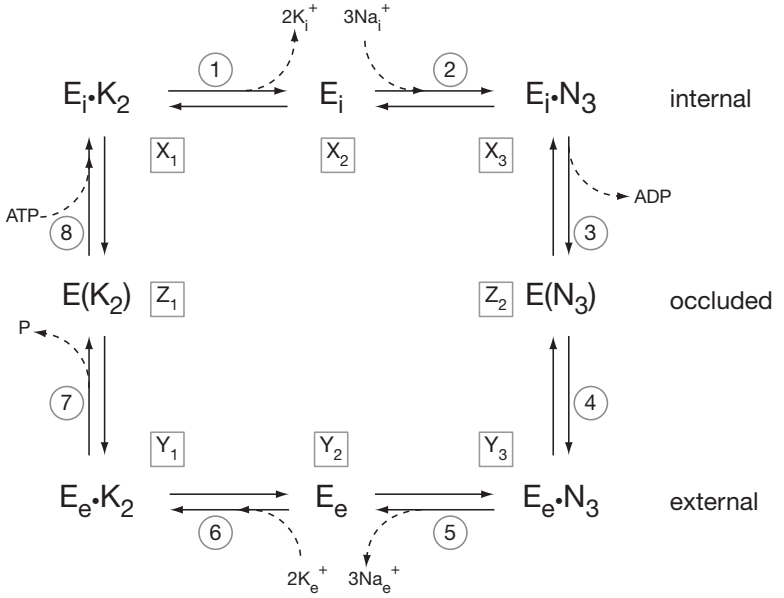

Figure 2.12: Model of the Na,K-ATPase based on the Post-Albers scheme. The model is simplified by assuming fast equilibrium between the boxed states.

we assume to happen in a single step), and gains three Na^+ ions, again in a single step, to move through states X_2 and X_3. ATP is bound to each 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 conformation 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 number inside a circle. For each $i = 1 \ldots 8$, transition i has two rate constants, k_i in the clockwise direction and k_{-i} in the anti-clockwise direction.

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

$$\frac{dx_1}{dt} = k_{-1}[K_i^+]^2 x_2 + k_8[\mathrm{ATP}]z_1 - (k_{-8} + k_1)x_1, \tag{2.98}$$

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. It is of the form

$$J = \frac{[\mathrm{ATP}]n_i^3\kappa_e^2 - [\mathrm{ADP}][\mathrm{P}_i]\kappa_i^2 n_e^3(\prod_{i=1}^{8} K_i)}{\phi}, \tag{2.99}$$

where ϕ is the sum of a large number of terms involving products of the rate constants and concentrations. Here, n denotes a Na^+ concentration, and κ denotes a 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 must be 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.

2.5.4 Nuclear Transport

The transport of proteins from the cytoplasm to the nucleus (or the reverse) is accomplished by means that combines features of the each of the above transport mechanisms. The nuclear membrane contains protein structures called nuclear pore complexes (NPCs) which 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 if its much lower binding affiinity, unbinds from the carrier molecule, completing the cycle. Of course, all of the reactions are reversible, however, the directionality is maintained by the free energy of GTP hydrolysis and the high concentration of GTP in the nucleus.

We defer 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 we assume that hydrolysis of RanGTP to RanGDP is 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.

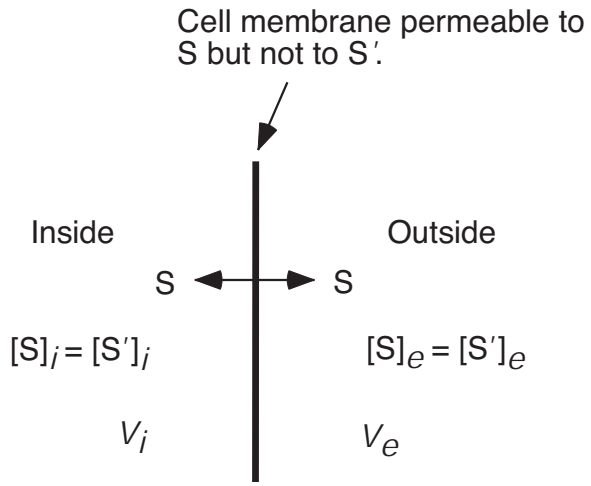

Figure 2.13: Schematic diagram of a membrane separating two solutions with different ionic concentrations.

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 for 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 we have two reservoirs containing the same ion S, but at different concentrations, as shown schematically in Fig. 2.13. The reservoirs are separated by a 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 will be 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 (**??**)

$$G_{\mathrm{S},i} = G^0_{\mathrm{S}} + RT\ln([\mathrm{S}]_i) + zFV_i, \tag{2.100}$$

while on the outside it is

$$G_{\mathrm{S},e} = G^0_{\mathrm{S}} + RT\ln([\mathrm{S}]_e) + zFV_e. \tag{2.101}$$

The chemical potential difference is

$$\Delta G_{\mathrm{S}} = G_{\mathrm{S},i} - G_{\mathrm{S},e} = RT\ln\left(\frac{[\mathrm{S}]_i}{[\mathrm{S}]_e}\right) + zFV. \tag{2.102}$$

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

$$V_{\mathrm{S}} = \frac{RT}{zF}\ln\left(\frac{[\mathrm{S}]_e}{[\mathrm{S}]_i}\right) = \frac{kT}{zq}\ln\left(\frac{[\mathrm{S}]_e}{[\mathrm{S}]_i}\right), \tag{2.103}$$

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.

Typical concentrations (in this case, for squid axon) are 397, 50, and 40 mM for potassium, sodium, and chloride, 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 $V_{\mathrm{Na}} = 56$ mV, $V_{\mathrm{K}} = -77$ mV, $V_{\mathrm{Cl}} = -68$ 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 $V = V_{\mathrm{S}}$. 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 no net 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, we suppose that the macromolecules are negatively charged, S is positively charged, and S′ is negatively charged. Outside the cell, S and S′ must have the same concentration (assuming the charge amplitude on S and S′ is the same), to maintain charge neutrality. Inside, charge neutrality requires more S than S′, in order to balance the negative charge on the macromolecules. At equilibrium, there must be a potential difference across the membrane which opposes the movement of S from inside to outside and also opposes the movement of S′ from outside to inside. Thus, the membrane potential is the Nernst potential for both S and S′, namely

$$V_{\rm S} = \frac{RT}{zF} \ln\left(\frac{[{\rm S}]_e}{[{\rm S}]_i}\right) = -\frac{RT}{zF} \ln\left(\frac{[{\rm S}']_e}{[{\rm S}']_i}\right) \tag{2.104}$$

where

$$z_x[{\rm X}] + z[{\rm S}']_i = z[{\rm S}]_i \quad \text{and} \quad [{\rm S}']_e = [{\rm S}]_e. \tag{2.105}$$

It follows that

$$[{\rm S}']_e[{\rm S}]_e = [{\rm S}']_i[{\rm S}]_i, \tag{2.106}$$

and thus

$$[{\rm S}]_i([{\rm S}]_i - \frac{z_x}{z}[{\rm X}]) - ([{\rm S}]_e)^2 = 0. \tag{2.107}$$

This quadratic equation can be solved to find a unique positive value for $[{\rm S}]_i$

$$[{\rm S}]_i = \sigma[{\rm S}]_e, \qquad \sigma = \frac{1}{2}(Z + \sqrt{Z^2+4}), \tag{2.108}$$

where $Z = \frac{z_x[{\rm X}]}{z[{\rm S}]_e}$.and from this the transmembrane potential can be determined using (2.104).

This equilibrium is called the *Gibbs-Donnan equilibrium*. The potential difference generated this way is known to occur across cell membranes and also across the edge of a gel in aqueous solution. This potential drop occurs in 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*

$$\mathbf{J} = -u\frac{z}{|z|}c\nabla\phi, \tag{2.109}$$

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:

$$D = \frac{uRT}{|z|F}. \tag{2.110}$$

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

$$\mathbf{J} = -D\left(\nabla c + \frac{zF}{RT}c\nabla\phi\right). \tag{2.111}$$

If the flow of ions and the electric field are transverse to the membrane, we can view (2.111) as the one-dimensional relation

$$J = -D\left(\frac{dc}{dx} + \frac{zF}{RT}c\frac{d\phi}{dx}\right). \tag{2.112}$$

The Nernst equation

The Nernst equation can also be derived from the Nernst–Planck electrodiffusion equation (2.112). When the flux J is zero, we find

$$-D\left(\frac{dc}{dx} + \frac{zF}{RT}c\frac{d\phi}{dx}\right) = 0, \tag{2.113}$$

so that

$$\frac{1}{c}\frac{dc}{dx} + \frac{zF}{RT}\frac{d\phi}{dx} = 0. \tag{2.114}$$

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 from $x = 0$ to $x = L$ we get

$$\ln(c)\big|_{c_i}^{c_e} = \frac{zF}{RT}(\phi_i - \phi_e), \tag{2.115}$$

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

$$V = \frac{RT}{zF}\ln\left(\frac{c_e}{c_i}\right), \tag{2.116}$$

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 (this is 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 we have two reservoirs 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, we have $\partial\phi/\partial x = -V/L$, 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 (2.111) is an ordinary differential equation for the concentration c,

$$\frac{dc}{dx} - \frac{zFV}{RTL}c + \frac{J}{D} = 0, \tag{2.117}$$

whose solution is

$$\exp\left(\frac{-zVFx}{RTL}\right)c(x) = \frac{JRTL}{DzVF}\left[\exp\left(\frac{-zVFx}{RTL}\right) - 1\right] + c_i, \tag{2.118}$$

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

$$J = \frac{D}{L}\frac{zFV}{RT}\frac{c_i - c_e\exp\left(\frac{-zVF}{RT}\right)}{1 - \exp\left(\frac{-zVF}{RT}\right)}, \tag{2.119}$$

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

$$I_{\mathrm{S}} = P_{\mathrm{S}}\frac{z^2F^2}{RT}V\frac{c_i - c_e\exp\left(\frac{-zFV}{RT}\right)}{1 - \exp\left(\frac{-zFV}{RT}\right)}, \tag{2.120}$$

where $P_{\mathrm{S}} = D/L$ is the permeability of the membrane to S. This is the famous Goldman–Hodgkin–Katz (GHK) current equation. It plays an important role in models of cellular electrical activity.

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

$$V = V_{\mathrm{S}} = \frac{RT}{zF}\ln\left(\frac{c_e}{c_i}\right), \tag{2.121}$$

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

$$0 = \sum_{z=1} P_j \frac{c_i^j - c_e^j \exp\left(\frac{-VF}{RT}\right)}{1 - \exp\left(\frac{-VF}{RT}\right)} + \sum_{z=-1} P_j \frac{c_i^j - c_e^j \exp\left(\frac{VF}{RT}\right)}{1 - \exp\left(\frac{VF}{RT}\right)}, \tag{2.122}$$

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

$$V = \frac{RT}{F} \ln - \left(\frac{\sum_{z=-1} P_j c_e^j + \sum_{z=1} P_j c_i^j}{\sum_{z=-1} P_j c_i^j + \sum_{z=1} P_j c_e^j} \right). \tag{2.123}$$

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

$$V_r = \frac{RT}{F} \ln - \left(\frac{P_{Na}[Na^+]_i + P_K[K^+]_i + P_{Cl}[Cl^-]_e}{P_{Na}[Na^+]_e + P_K[K^+]_e + P_{Cl}[Cl^-]_i} \right). \tag{2.124}$$

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 present a detailed discussion of 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

$$C_m = \frac{Q}{V}. \tag{2.125}$$

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

$$C_m = \frac{k\epsilon_0}{d}, \tag{2.126}$$

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 found to be 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. 2.14. It is assumed that the membrane acts like a capacitor in parallel with a resistor (although not necessarily ohmic). Since the current is defined by dQ/dt, it follows

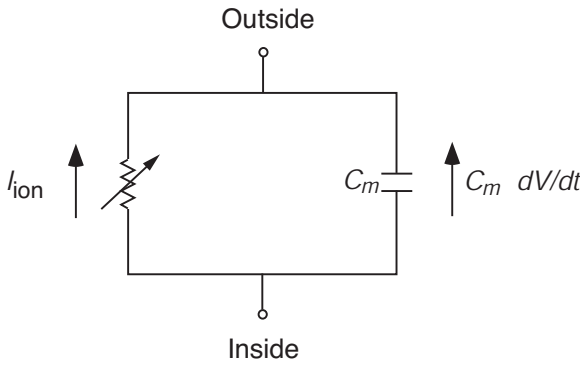

Figure 2.14: Electrical circuit model of the cell membrane.

from (2.125) 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

$$C_m \frac{dV}{dt} + I_{\text{ion}} = 0, \tag{2.127}$$

where $V = V_i - V_e$.

We will meet this equation 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 (2.120), and others will be discussed in Chapter 3.

Another common model describes I_{ion} as a linear function of the membrane potential. In Chapter 3 we will see 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

$$V_{\text{S}} = \frac{RT}{zF} \ln\left(\frac{[\text{S}]_e}{[\text{S}]_i}\right), \tag{2.128}$$

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 find

$$V = rI_{\text{S}} + V_{\text{S}}, \tag{2.129}$$

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

$$I_{\text{S}} = g(V - V_{\text{S}}), \tag{2.130}$$

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 $V = V_S$, 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

$$rQ = P_1 - P_2, \tag{2.131}$$

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 (2.131) 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, and therefore it is useful but not universally true.

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

$$\Delta G = RT \ln \frac{S_1}{S_2}, \tag{2.132}$$

where S_i is the mole fraction of solvent in the i^{th} 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 will tend to diffuse 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

$$RT \ln \frac{S_1}{S_2} + \pi_s v_s = 0, \tag{2.133}$$

where v_s is the molar volume (L/mole) of the solvent. Thus, since $S_2 = 1$,

$$\pi_s = -\frac{RT}{v_s} \ln(S_1) = -\frac{RT}{v_s} \ln(1 - N) \approx \frac{RT}{v_s} N, \tag{2.134}$$

where N is the mole fraction of solvent. 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

$$\pi_s = \frac{RT}{v_s} \frac{n}{n_s} \approx RcT, \tag{2.135}$$

since $n_s v_s$ is quite close to the volume of solution. Here c is the concentration of solvent in units of moles per liter. Using that $c = n/v$, where v is the volume, (2.135) becomes

$$\pi_s v = nkT, \tag{2.136}$$

which is the same as the ideal gas law. Equation (2.135) was first found empirically by van't Hoff.

As with all things derived from ideal properties, the expression (2.135) is an approximation, which unfortunately, is worst for some of the most common small ions, like sodium and potassium. (It similarly turns out that the expression for free energy (1.12) is least valid for small ions such as sodium and potassium.) The formula

$$\pi_s v = \gamma nkT, \tag{2.137}$$

works much better, where γ is a correction factor found experimentally. In Table **??** the correction factor γ is listed for several of the most common chemical species of physiological importance. **Use info from Berne and Levy, pg. 9, to construct this table.**

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 is modified by osmotic pressure to be

$$rQ = P_1 - \pi_s - P_2, \tag{2.138}$$

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 (a 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

$$\rho g h_2 = \rho g h_1 - \frac{nkT}{h_1 A}, \tag{2.139}$$

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:

$$h_1^2 - h_0 h_1 - \frac{nkT}{2\rho g A} = 0. \tag{2.140}$$

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

$$h_1 - h_2 = \sqrt{h_0^2 + \frac{2nkT}{\rho g A}} - h_0. \tag{2.141}$$

When the solute is at a high enough concentration, physical solutions of (2.141) 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 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, 1992).

Because the cell membrane is a thin lipid bilayer, it is incapable of withstanding any hydrostatic pressure differences. This is a potentially fatal weakness. If the intracellular concentrations of various ions and larger molecules become too large, osmotic forces cause 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; Reuss, 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 urinary 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

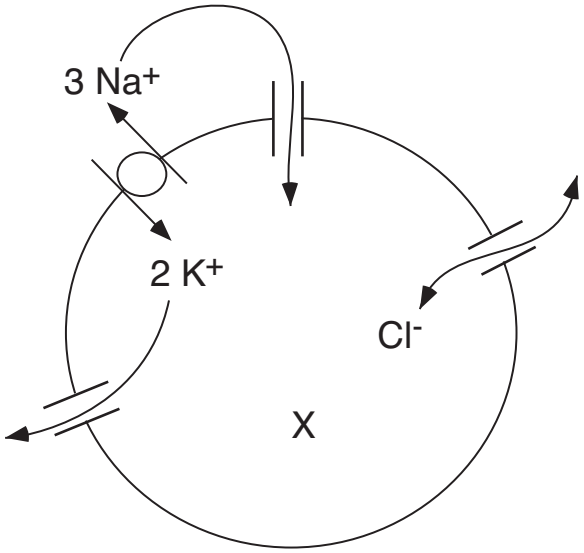

Figure 2.15: Schematic diagram of the pump–leak model.

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 (2.120) and the linear relationship (2.130). 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).

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

$$I_{\mathrm{Na}} = g_{\mathrm{Na}} \left[V - \frac{RT}{F} \ln \left(\frac{[\mathrm{Na}^+]_e}{[\mathrm{Na}^+]_i} \right) \right] + 3pq, \tag{2.142}$$

$$I_{\mathrm{K}} = g_{\mathrm{K}} \left[V - \frac{RT}{F} \ln \left(\frac{[\mathrm{K}^+]_e}{[\mathrm{K}^+]_i} \right) \right] - 2pq, \tag{2.143}$$

$$I_{\mathrm{Cl}} = g_{\mathrm{Cl}} \left[V + \frac{RT}{F} \ln \left(\frac{[\mathrm{Cl}^-]_e}{[\mathrm{Cl}^-]_i} \right) \right], \tag{2.144}$$

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

$$I_A = -\frac{d}{dt}(zFw[A^{z+}]), \tag{2.145}$$

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

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

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

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

The flow of water across the membrane is driven by osmotic pressure, so that the change of cell volume is given by

$$r\frac{dw}{dt} = RT\left([Na^+]_i - [Na^+]_e + [K^+]_i - [K^+]_e + [Cl^-]_i - [Cl^-]_e + \frac{X}{w}\right). \tag{2.149}$$

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.

Finally, we assume that both the extracellular and intracellular media are each in electroneutrality. The only place where electroneutrality is violated is in a thin region (the Debye layer) near the membrane, and the amount of charge here is so small, relatively speaking, that it does not affect the overall assumption of electroneutrality. To see that this stored charge is quite small, 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 potassium ions at 400 mM is about 0.1 π C/cm, showing a relative charge deflection of about 10^{-7}.

For this model, we assume that sodium, potassium, and chloride are in electroneutrality in the extracellular region. 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, sodium, potassium, and chloride 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.

The assumption of electroneutrality gives the two equations

$$[Na^+]_e + [K^+]_e - [Cl^-]_e = 0, \qquad [Na^+]_i + [K^+]_i - [Cl^-]_i + z_x\frac{X}{w} = 0. \tag{2.150}$$

It is also convenient to assume that the cell is in an infinite bath, so that ionic currents do not change the external concentrations, and therefore the external concentrations are assumed to be fixed and known.

The differential equations (2.146), (2.147), (2.148), and (2.149) together with the electrostatic balance laws (2.150) describe the changes of cell volume and membrane potential as functions of time. 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 4 for the discussion of excitability.

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

$$\alpha y - \frac{1}{y} + \frac{z_x}{\mu} = 0, \tag{2.151}$$

and the equation of osmotic pressure balance becomes

$$\alpha y + \frac{1}{y} + \frac{1}{\mu} - 2 = 0, \tag{2.152}$$

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

$$\frac{[\mathrm{Na}^+]_i}{[\mathrm{Na}^+]_e} = e^{-3P} y, \tag{2.153}$$

$$\frac{[\mathrm{K}^+]_i}{[\mathrm{K}^+]_e} = e^{2P\gamma} y, \tag{2.154}$$

$$\frac{[\mathrm{Cl}^-]_i}{[\mathrm{Cl}^-]_e} = \frac{1}{y}. \tag{2.155}$$

Solving (2.151) for its unique positive root, we obtain

$$y = \frac{-z_x + \sqrt{z^2 + 4\alpha\mu^2}}{2\alpha\mu}, \tag{2.156}$$

and when we substitute for y back into (2.152), we find the quadratic equation for μ:

$$4(1-\alpha)\mu^2 - 4\mu + 1 - z_x^2 = 0. \tag{2.157}$$

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

$$\rho(P) = \frac{[\mathrm{Na}^+]_e e^{-3P} + [\mathrm{K}^+]_e e^{2P\gamma}}{[\mathrm{Na}^+]_e + [\mathrm{K}^+]_e} < 1. \tag{2.158}$$

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

$$\frac{3[\mathrm{Na}^+]_e}{g_{\mathrm{Na}}} > \frac{2[\mathrm{K}^+]_e}{g_{\mathrm{K}}}, \tag{2.159}$$

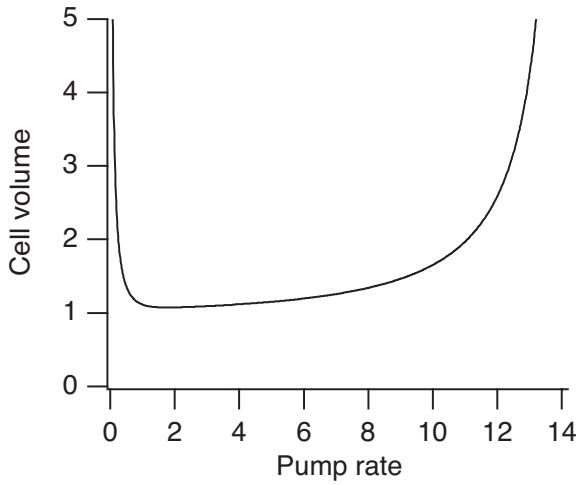

Figure 2.16: Cell volume as a function of the pump rate.

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 we will see, 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 $g_{\mathrm{K}} = 0.367$ mS/cm^2 and $g_{\mathrm{Na}} = 0.01$ mS/cm^2. For these values, and at the extracellular concentrations of 437 and 20 mM for sodium and potassium, respectively, the condition (2.159) 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 (2.148) or (2.155). 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. 2.16 and 2.17 the 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. 2.17 are shown the sodium and potassium equilibrium potentials. For these plots, γ was chosen to be 0.11, and $z_x = -1$. Then, at $P = 1.6$, the sodium and potassium 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 potassium equilibrium potential is seen to decrease rapidly as a function of pump rate until it reaches a plateau at a minimum value. The sodium equilibrium potential increases monotonically.

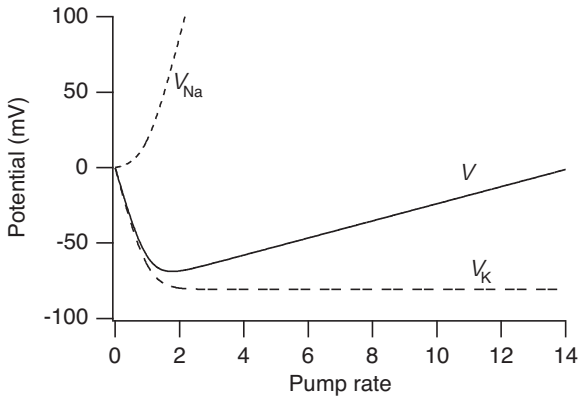

Figure 2.17: Membrane potential, sodium equilibrium potential, and potassium equilibrium potential as functions of the pump rate.

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

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 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 for control of volume 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} were treated as constants. In Chapter 4 we will see 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

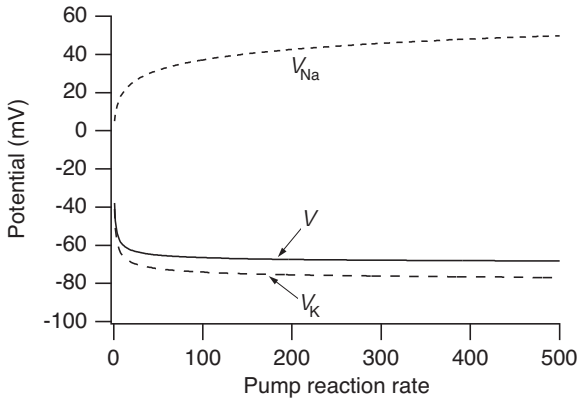

Figure 2.18: Membrane potential, sodium equilibrium potential, and potassium equilibrium potential as functions of the pump rate.

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 2.17 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.

A different model of the pump activity might be beneficial. Recall from (2.97) that with each "stroke" of the ion exchange pump, three intracellular sodium ions are exchanged for two extracellular potassium ions. The analysis of the Na,K-ATPase suggests that at low internal sodium concentrations, the pump rate can be represented in nondimensional variables as

$$P = \rho u^3, \tag{2.160}$$

where $u = [\mathrm{Na}^+]_i/[\mathrm{Na}^+]_e$. 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 sodium concentration becomes

$$u \exp(3\rho u^3) = y, \tag{2.161}$$

and this must be solved together with the quadratic polynomials (2.151) and (2.152) (replacing (2.142) for y and μ).

In Fig. 2.18 are shown the membrane potential, and the sodium and potassium equilibrium potentials, plotted as functions of the nondimensional reaction rate ρ. Here we see something qualitatively different from what is depicted in Fig. 2.17. 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 sodium concentration, increasing the speed of the pumping rate has little effect when the internal sodium is depleted, because of the diminished number of sodium 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 sodium binding site on the extracellular side. This causes an increase in internal sodium, which has an inhibitory effect on the sodium–calcium antiport exchanger, slowing down the rate by which calcium exits the cells. Increased levels of calcium 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 for 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. 2.19). (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.

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 down 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

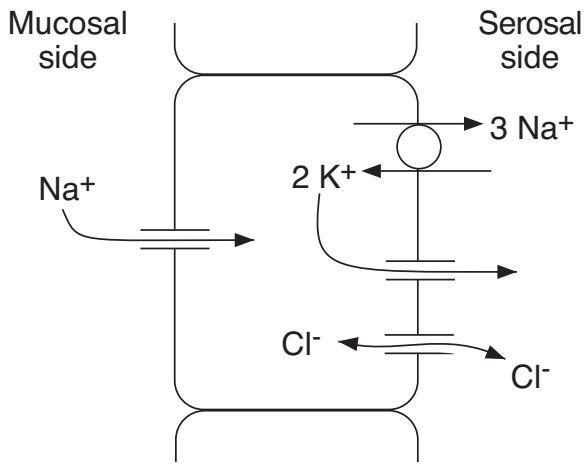

Figure 2.19: Schematic diagram of the model of a Na^+-transporting epithelial cell, based on the model of Koefoed-Johnsen and Ussing (1958).

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,

$$P_{\mathrm{Na}} v \frac{N_i - N_m e^{-v}}{1 - e^{-v}} + 3qpN_i = 0, \tag{2.162}$$

$$P_{\mathrm{K}} v \frac{K_i - K_s e^{-v}}{1 - e^{-v}} - 2qpN_i = 0, \tag{2.163}$$

$$P_{\mathrm{Cl}} v \frac{C_i - C_s e^{v}}{1 - e^{v}} = 0. \tag{2.164}$$

Note that the voltage, v, has been scaled by $F/(RT)$ 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,

$$w(N_i + K_i - C_i) + z_x X = 0, \tag{2.165}$$

$$N_i + K_i + C_i + \frac{X}{w} = N_s + K_s + C_s, \tag{2.166}$$

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

$$N_s + K_s = C_s. \tag{2.167}$$

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 (2.162), (2.163), and (2.164) can be solved for N_i, K_i, and C_i, respectively, to get

$$N_i(v) = \frac{vN_m e^{-v}}{v + 3\rho_n(1 - e^{-v})}, \tag{2.168}$$

$$K_i(v) = 2\rho_k N_i(v)\frac{1 - e^{-v}}{v} + K_s e^{-v}, \tag{2.169}$$

$$C_i(v) = C_s e^{v}, \tag{2.170}$$

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

Next, eliminating $N_i + K_i$ between (2.165) and (2.166), we find that

$$2\mu(C_i - C_s) = z_x - 1. \tag{2.171}$$

We now use (2.170) to find that

$$z_x - 1 = 2\mu C_s(e^v - 1), \tag{2.172}$$

and thus, using (2.172) to eliminate μ from (2.165), we get

$$N_i(v) + K_i(v) = \frac{C_s}{1 - z_x}[-2z_x + e^v(1 + z_x)] \equiv \phi(v). \tag{2.173}$$

Since $z_x - 1 < 0$, it must be (from (2.172)) 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 (2.173), with $N_i(v)$ and $K_i(v)$ specified by (2.168) and (2.169).

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. 2.20. Next, from (2.168) and (2.169) 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

$$\frac{N_m}{1 + 3\rho_n} + \frac{2\rho_k N_m}{1 + 3\rho_n} + K_s < C_s. \tag{2.174}$$

Since $K_s + N_s = C_s$, this becomes

$$\frac{N_m}{N_s} < \frac{1 + 3\rho_n}{1 + 2\rho_k}. \tag{2.175}$$

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

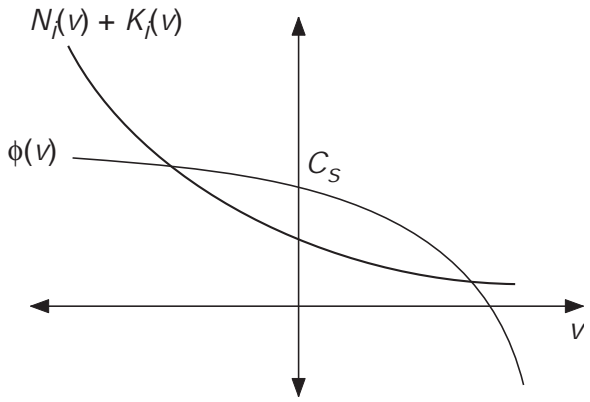

Figure 2.20: Sketch (not to scale) of the function $\phi(v)$, defined as the right-hand side of (2.173), and of $N_i(v) + K_i(v)$, where N_i and K_i are defined by (2.168) and (2.169). $\phi(v)$ is sketched for $z_x < -1$.

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 (2.168) and (2.169) 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

$$N_i'(0) + K_i'(0) < \phi'(0), \tag{2.176}$$

or that

$$4(1+3\rho_n)C_s + N_s(1-z_x)\frac{3\rho_n - 2\rho_k}{1+2\rho_k} > 0. \tag{2.177}$$

For the remainder of this discussion we assume that this condition holds, so that the failure of (2.175) also implies that the cell bursts.

According to (2.175), a transporting epithelial cell can maintain its cell volume, provided that 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. 2.21.

Obviously, this state of affairs is unsatisfactory. In fact, some epithelial cells, such as those in the loop of Henle in the nephron (Chapter **??**), must work in environments with extremely high mucosal sodium 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.

From (2.175) 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

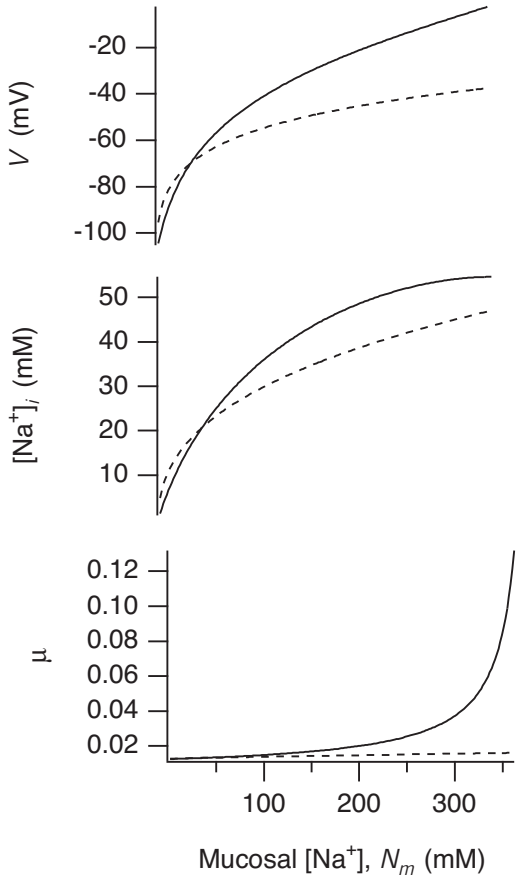

Figure 2.21: Numerical solutions of the model for epithelial cell volume regulation and Na^+ transport. The membrane potential V, the scaled cell volume μ, and the intracellular Na^+ concentration $[Na^+]_i$ are plotted as functions of the mucosal Na^+ concentration. The solid lines are the solutions 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 (2.168) and (2.169), 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 sodium ions from the mucosal side helps to control this swelling. Similarly, increasing the conductance of potassium ions allows more potassium ions to flow out of the cell, thereby decreasing the osmotic pressure from potassium 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. 2.21. 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 for 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).

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 of $D = 10^{-7}$ cm^2/s and binding equilibrium of $K_{\mathrm{eq}} = 30$ 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 the dye is a fast buffer of hydrogen.)

3. Segel, Chet and Henis (1977) used (2.17) to estimate the diffusion coefficient for bacteria. With the external concentration C_0 at 7×10^7ml^{-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 $k_+ = 1.4 \times 10^{10}\text{cm}^3\,\text{M}^{-1}\text{s}^{-1}$, and $k_- = 11\ \text{s}^{-1}$.

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

$$\frac{1}{y^2}(y^2\sigma')' - \alpha^2 \frac{\sigma}{\sigma+1} = 0, \tag{2.178}$$

$$\sigma'(0) = 0, \tag{2.179}$$

$$\sigma(1) = \sigma_0, \tag{2.180}$$

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

8. Suppose a membrane that contains water-filled pores separates two solutions, as sketched in Fig. 2.22.

   (a) Suppose that the solution on either side of the membrane contains an impermeant solute. Show that at steady-state the hydrostatic pressure of the water within the pores must be less than the hydrostatic pressure of the solutions on either side of the membrane. Hint: note that the solute cannot get inside the pore, but water can. What would happen if the hydrostatic pressure of water was the same inside the pore as out?

   (b) If the impermeant solute occurs at the same concentration, c, on both sides of the membrane, what is the steady-state hydrostatic pressure drop between the external solution and the water inside the pore?

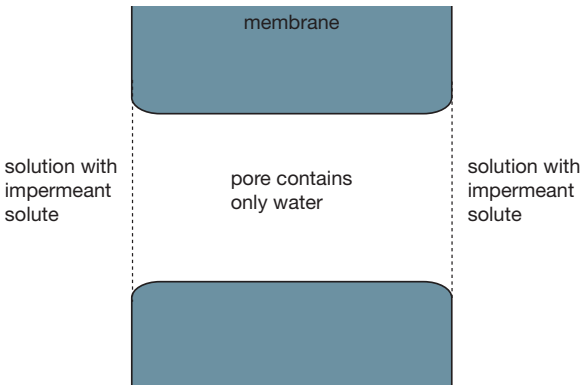

Figure 2.22: Sketch of a pore in a membrane, with an impermeant solute on both sides of the membrane.

(c) Show that if the solute can permeate the pore freely, osmosis can drive a flow of water in the absence of any hydrostatic pressure gradient.

(d) Now suppose that the impermeant solute occurs at different concentrations on either side of the membrane. We expect there to be flow of water across the membrane, from the region of low solute concentration to the region of high solute concentration. How can this happen, seeing as how there cannot be any gradient of solute concentration inside the pore?

(e) If the impermeant solute occurs at different concentrations on either side of the membrane what is the steady-state pressure gradient within the pore? What are the steady-state pressure drops at the ends of the pore?

(f) Show that if the solution on one side of the membrane contains both permeant and impermeant solutes, while the solution on the other side contains only the permeant solute, it is possible for water to flow against its chemical potential gradient (at least temporarily). This problem of wrong-way water flow has been observed experimentally, and is discussed in detail by Dawson (1992). Hint: consider the case where the left side of the membrane has a much higher concentration of the permeable solute, while the right side contains a smaller concentration of the impermeant solute. Keep the pressure the same on both sides of the membrane. What hydrostatic pressure gradient is established inside the pore? What effect does this have on the bulk flow of water? Can this overwhelm the diffusion of water from right to left?

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

10. 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 (2.56)? How is this reaction affected by the concentration of ATP?

11. In the model for the glucose transporter (Fig. 2.7) 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.

12. Consider the model of a non-electrogenic Na^+-Ca^{2+} exchanger. At equilibrium the concentrations on either side of the membrane are related by the equation

$$\frac{n_e^3 c_i}{n_i^3 c_e} = 1. \tag{2.181}$$

    Clearly, this equation alone does not determine the equilibrium concentrations uniquely. For a given set of initial conditions 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.

13. Simplify the model of the Na^+-Ca^{2+} exchanger 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

$$k_1 c_i x_1 = k_{-1} n_i^3 x_2, \tag{2.182}$$
$$k_3 n_e^3 y_2 = k_{-3} c_e y_1. \tag{2.183}$$

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

14. Simplify the model of Fig. 2.12 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

$$x_1 = K_1 x_2 \kappa_i^2, \tag{2.184}$$
$$n_i^3 x_2 = K_2 x_3, \tag{2.185}$$

$$y_3 = K_5 n_e^3 y_2, \tag{2.186}$$
$$\kappa_e^2 y_2 = K_6 y_1, \tag{2.187}$$

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

15. The process by which calcium is taken up into the sarcoplasmic reticulum (SR) in muscle and cardiac cells is similar to the sodium–potassium ATPase, but simpler. Two intracellular calcium ions bind with a carrier protein with high affinity for calcium. ATP is dephosphorylated, with the phosphate bound to the carrier. There is a conformational change of the carrier protein that exposes the calcium to the interior of the SR and reduces the affinity of the binding sites, thereby releasing the two ions of calcium. The phosphate is released and the conformation changed so that the calcium binding sites are once again exposed to the intracellular space.

    Formalize this reaction and find the rate of calcium uptake by this pump.

16. 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.

17. Suppose the intracellular macromolecule X can bind b molecules of the ion S via $X + bS \rightleftarrows XS$. 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 sodium. When eaten and absorbed into the body, how many osmoles does this bag of potato chips represent?

19. (a) Confirm that π_s in (2.135) 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 semi-permeable 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 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 potassium. 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 sodium 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. Write and analyze the balance equations for a cell in a finite bath. Hint: In a finite bath the total volume is conserved as are the total number of sodium, potassium, and chloride ions.

25. Simulate the time-dependent differential equations governing cell volume and ionic concentrations.

26. Ouabain is known to compete with K^+ for external potassium 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?

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 we see 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 (1992), to which the reader is referred for a more detailed presentation than that given here. The bibliography given there will also serve as a starting point for more detailed studies.

3.1 Current–Voltage Relations

Before we discuss specific models for ion channels, we emphasize an important fact that can be a source of confusion to the novice. Although the Nernst equation (2.103) 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 membrane Na^+ channels. Already we have seen two different models for ionic currents. In the previous chapter we discussed two common models for Na^+ current as a function of the membrane potential and the

internal and external Na^+ concentrations. In the simpler model, we assumed that the Na^+ current across the cell membrane was 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^+. (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 must be the case. However, we also discussed an alternative model, where integration of the Nernst–Planck equation (2.111), 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. 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 a current–voltage relation different from the rest. 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 we discuss 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 4). 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 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 (2.124) that the reversal potential V_r at which there is no net current flow is given by

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

However, if we assume instead that the I–V curves for Na^+ and K^+ are linear, then

the reversal potential is given by

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

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 an already open channel, 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 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

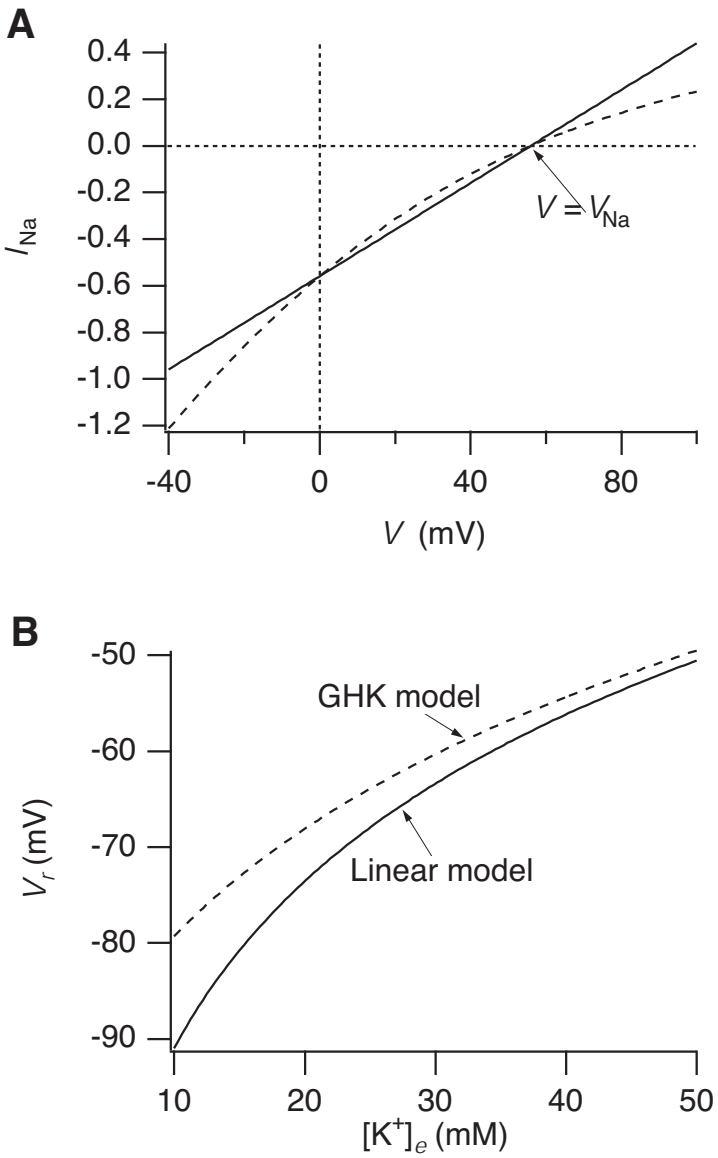

Figure 3.1: A: I–V curves of the linear and GHK models for Na^+ flux through a membrane. Both curves have the same reversal potential as expected, but the GHK model (dashed curve) gives a nonlinear I–V curve. Typical concentrations and conductances of the squid axon were used: $[Na^+]_i = 50\,\text{mM}$,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

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).

Finally, we discuss simple models for the kinetics of channel gating. These models will be of fundamental importance in Chapter 4, where we use an early model for the voltage-dependent gating of ion channels proposed by Hodgkin and Huxley as part of their model for the action potential in the squid giant axon. More detailed recent models for channel gating are not discussed at any length. The interested reader is referred to Hille (1992), 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 (1992) 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 introduce a hypothetical concentration c_e^* defined as that external concentration necessary to maintain a Nernst potential V. Thus

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

and

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

When the external concentration is c_e^* and the internal concentration is c_i, then the voltage is V, and there is 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

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

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

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

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

As an illustration of the application of the Ussing flux ratio, suppose the Na^+ current is measured when the cell is immersed in a high Na^+ solution and then compared to the Na^+ current measured in a low Na^+ solution. The membrane potential and the internal Na^+ concentration are assumed to be the same in both cases. We let a prime denote quantities measured in the high Na^+ solution, and then

$$\frac{I'_{\text{Na}}}{I_{\text{Na}}} = \frac{J'_{\text{out}} - J'_{\text{in}}}{J_{\text{out}} - J_{\text{in}}}. \tag{3.12}$$

Since the internal concentrations are the same, it follows from (3.6) that $J_{\text{out}} = J'_{\text{out}}$, and from (3.5) we find $J'_{\text{in}}/J_{\text{in}} = [\text{Na}^+]'_e/[\text{Na}^+]_e$. Substituting these into (3.12) and using the Ussing flux ratio, we find

$$\frac{I'_{\text{Na}}}{I_{\text{Na}}} = \frac{([\text{Na}^+]'_e/[\text{Na}^+]_e)\exp[\frac{(V_{\text{Na}}-V)F}{RT}] - 1}{\exp[\frac{(V_{\text{Na}}-V)F}{RT}] - 1}. \tag{3.13}$$

Alternatively, this can be written as

$$\frac{I'_{\text{Na}}}{I_{\text{Na}}} = \frac{[\text{Na}^+]_i - [\text{Na}^+]'_e \exp(\frac{-VF}{RT})}{[\text{Na}^+]_i - [\text{Na}^+]_e \exp(\frac{-VF}{RT})}. \tag{3.14}$$

By measuring the current ratio as a function of membrane potential, the Na^+ channel can thus conveniently be tested for independence.

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 still 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 $[A_rB_{n-r}]$ is converted to $[A_{r+1}B_{n-r-1}]$ is $\alpha[A_rB_{n-r}]$, 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,

$$\frac{\alpha}{\beta} = \frac{c_e}{c_i} \exp\left(\frac{-VF}{RT}\right). \tag{3.15}$$

This assumption is justified by the fact that a flux of one ion involves the movement of a single charge 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 $\alpha[A_nB_0]$, and the rate at which B ions are transferred to the left hand side is $\beta[A_0B_n]$. Thus, the flux ratio of labeled ions is

$$\frac{J_{\text{in}}}{J_{\text{out}}} = \frac{\alpha}{\beta}\frac{[A_nB_0]}{[A_0B_n]}. \tag{3.16}$$

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

$$\frac{[A_{r+1}B_{n-r-1}]}{[A_rB_{n-r}]} = \frac{\alpha}{\beta}. \tag{3.17}$$

Thus,

$$\frac{J_{\text{in}}}{J_{\text{out}}} = \frac{\alpha}{\beta}\frac{[A_nB_0]}{[A_0B_n]} = \left(\frac{\alpha}{\beta}\right)^2 \frac{[A_{n-1}B_1]}{[A_0B_n]} = \cdots = \left(\frac{\alpha}{\beta}\right)^{n+1}, \tag{3.18}$$

so that

$$\frac{J_{\text{in}}}{J_{\text{out}}} = \left[\frac{c_e}{c_i} \exp\left(\frac{-VF}{RT}\right)\right]^{n+1}. \tag{3.19}$$

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

$$\frac{J_{\text{in}}}{J_{\text{out}}} = \left[\frac{c_e}{c_i} \exp\left(\frac{-VF}{RT}\right)\right]^{n}. \tag{3.20}$$

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 (1992), 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.

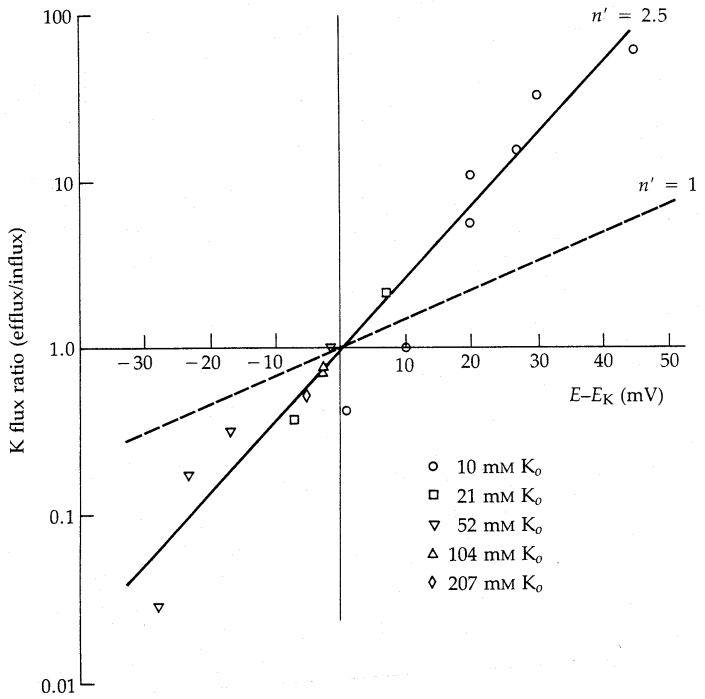

Figure 3.2: K^+ flux ratios as measured by Hodgkin and Keynes (1955), Fig. 7. Slightly modified into modern conventions by Hille (1992), page 375. K_o is the external K^+ concentration, and n' is the flux-ratio exponent, denoted by n in (3.20). (Hille, 1992, Fig. 7, p. 375.)

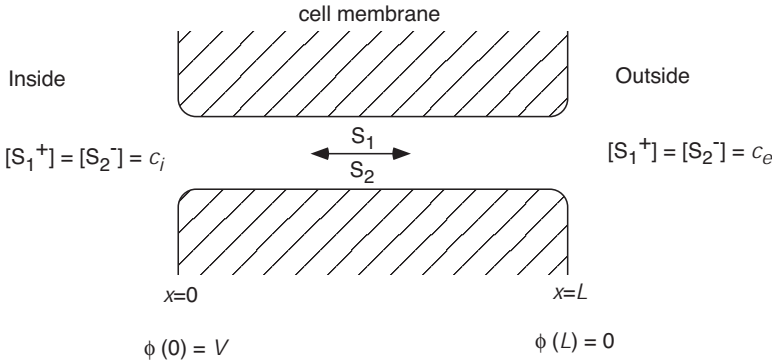

Figure 3.3: Schematic diagram of the electrodiffusion model for current through an ionic channel. Each side of the channel is electrically neutral, and both ion types can diffuse through the channel.

3.3 Electrodiffusion Models

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

$$J = -D\left(\frac{dc}{dx} + \frac{zF}{RT}c\frac{d\phi}{dx}\right), \tag{3.21}$$

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 (3.21) can be solved to give the Goldman–Hodgkin–Katz current and voltage equations (2.120) and (2.123). 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 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 we have 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 1 and that of the second is -1. Then the potential in the channel $\phi(x)$ must satisfy Poisson's equation,

$$\frac{d^2\phi}{dx^2} = -\frac{q}{\epsilon}(c_1 - c_2), \tag{3.22}$$

where q is the unit electric charge and ϵ is the dielectric constant of the channel medium (usually assumed to be an aqueous solution). 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 any charge buildup within the channel. Hence, the steady-state flux through the channel is described by (3.22) coupled with

$$J_1 = -D_1\left(\frac{dc_1}{dx} + \frac{F}{RT}c_1\frac{d\phi}{dx}\right), \tag{3.23}$$

$$J_2 = -D_2\left(\frac{dc_2}{dx} - \frac{F}{RT}c_2\frac{d\phi}{dx}\right), \tag{3.24}$$

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,

$$\begin{aligned} c_1(0) &= c_i, & c_1(L) &= c_e, \\ c_2(0) &= c_i, & c_2(L) &= c_e, \\ \phi(0) &= V, & \phi(L) &= 0. \end{aligned} \tag{3.25}$$

Note that we have specified 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 (3.22)–(3.25). 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, Barcilon, and Eisenberg, 1992; Barcilon, 1992; Barcilon, Chen, and Eisenberg, 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, rather than three, through the channel. Similar models have also been discussed by Peskin (1991).

It is convenient first to nondimensionalize the PNP equations. We let $x^* = x/L$, $\phi^* = \phi F/RT$, $v = VF/RT$, $c_1^* = c_1/\tilde{c}$, and similarly for c_2, c_i, and c_e, where $\tilde{c} = c_e + c_i$. Substituting into (3.22)–(3.24) and dropping the stars, we find

$$-\bar{J}_1 = \frac{dc_1}{dx} + c_1\frac{d\phi}{dx}, \tag{3.26}$$

$$-\bar{J}_2 = \frac{dc_2}{dx} - c_2\frac{d\phi}{dx}, \tag{3.27}$$

$$\frac{d^2\phi}{dx^2} = -\lambda^2(c_1 - c_2), \tag{3.28}$$

where $\lambda^2 = L^2qF\tilde{c}/(\epsilon RT)$, $\bar{J}_1 = J_1L/(\tilde{c}D_1)$, and similarly for $\bar{J}_2$. The boundary conditions are

$$\begin{aligned} c_1(0) &= c_i, & c_1(1) &= c_e, \\ c_2(0) &= c_i, & c_2(1) &= c_e, \\ \phi(0) &= v, & \phi(1) &= 0. \end{aligned}$$

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

$$\frac{d^2\phi}{dx^2} = 0, \tag{3.29}$$

and thus

$$\frac{d\phi}{dx} = -v. \tag{3.30}$$

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 c_1 is then

$$\frac{dc_1}{dx} - vc_1 = -\bar{J}_1, \tag{3.31}$$

and thus

$$c_1 = \frac{\bar{J}_1}{v} + K_1 e^{vx}. \tag{3.32}$$

From the boundary conditions $c_1(0) = c_i, c_1(1) = c_e$ it follows that

$$\bar{J}_1 = v \cdot \frac{c_i - c_e e^{-v}}{1 - e^{-v}}. \tag{3.33}$$

In dimensional form, this is

$$I_1 = FJ_1 = \frac{D_1}{L}\frac{F^2}{RT} \cdot V \cdot \left(\frac{c_i - c_e \exp(\frac{-VF}{RT})}{1 - \exp(\frac{-VF}{RT})} \right), \tag{3.34}$$

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

$$-\bar{J}_1 = \frac{dc_1}{dx} + c_1\frac{d\phi}{dx}, \tag{3.35}$$

$$-\bar{J}_2 = \frac{dc_2}{dx} - c_2\frac{d\phi}{dx}, \tag{3.36}$$

$$-\eta^2\frac{d^2\phi}{dx^2} = (c_1 - c_2). \tag{3.37}$$

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,

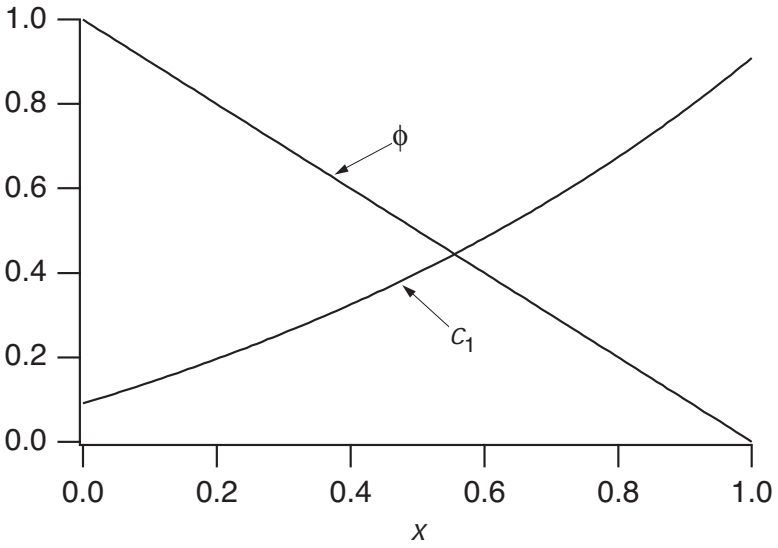

Figure 3.4: Graphs of the concentration and potential profiles for the short-channel limit of the Poisson–Nernst–Planck equations. Dimensionless parameters were set arbitrarily at $c_i = 50/550 = 0.091$, $c_e = 500/550 = 0.909$, $v = 1$. In this limit the electric field is constant through the channel (the potential has a constant slope), the concentration profile is nonlinear, and the GHK I–V curve is obtained.

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 (3.37) gives $c_1 = c_2$, which happens to satisfy both the left and right boundary conditions. Thus, c_1 and c_2 are identically equal throughout the channel. From (3.35) and (3.36) it follows that

$$\frac{d}{dx}(c_1 + c_2) = -\bar{J}_1 - \bar{J}_2. \tag{3.38}$$

Since both $\bar{J}_1$ and $\bar{J}_2$ are constants, it follows that dc_1/dx is a constant, and hence, from the boundary conditions,

$$c_1 = c_2 = c_i + (c_e - c_i)x. \tag{3.39}$$

We are now able to solve for ϕ. Subtracting (3.37) from (3.36) gives

$$2c_1 \frac{d\phi}{dx} = 2\tilde{J}, \tag{3.40}$$

where $2\tilde{J} = \bar{J}_2 - \bar{J}_1$, and hence

$$\phi = \frac{\tilde{J}}{c_e - c_i} \ln[c_i + (c_e - c_i)x] + K, \tag{3.41}$$

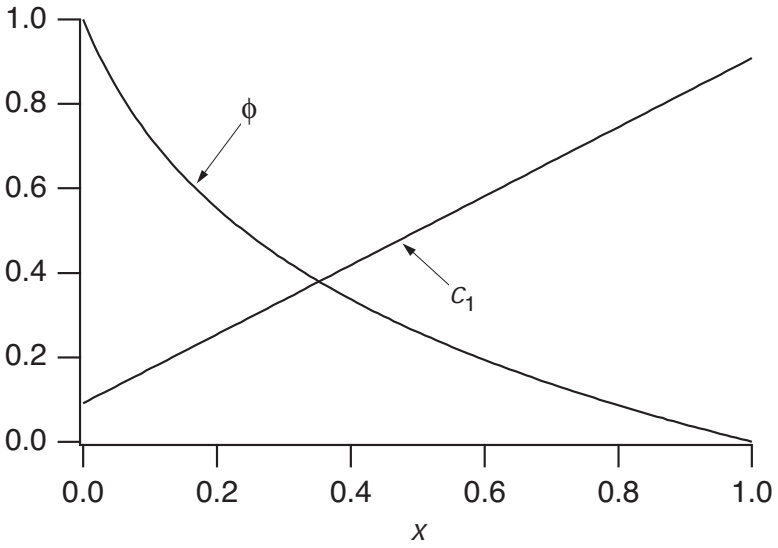

Figure 3.5: Graphs of the concentration and potential profiles for the long-channel limit of the Poisson–Nernst–Planck equations. Dimensionless parameters were set arbitrarily at $c_i = 50/550 = 0.091$, $c_e = 500/550 = 0.909$, $v = 1$. In this limit the concentration profile has a constant slope, the potential profile is nonlinear, and the linear I–V curve is obtained.

for some other constant K. Applying the boundary conditions $\phi(0) = v, \phi(1) = 0$ we determine $\tilde{J}$ and K, with the result that

$$\phi = -\frac{v}{v_1} \ln \left[\frac{c_i}{c_e} + \left(1 - \frac{c_i}{c_e} \right) x \right], \tag{3.42}$$

where $v_1 = \ln(c_e/c_i)$ 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 c_1 and ϕ into (3.35) to get

$$\bar{J}_1 = \frac{c_e - c_i}{v_1}(v - v_1), \tag{3.43}$$

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, ϕ 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 I–V curve obeys neither. Given the above derivation of the linear 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.

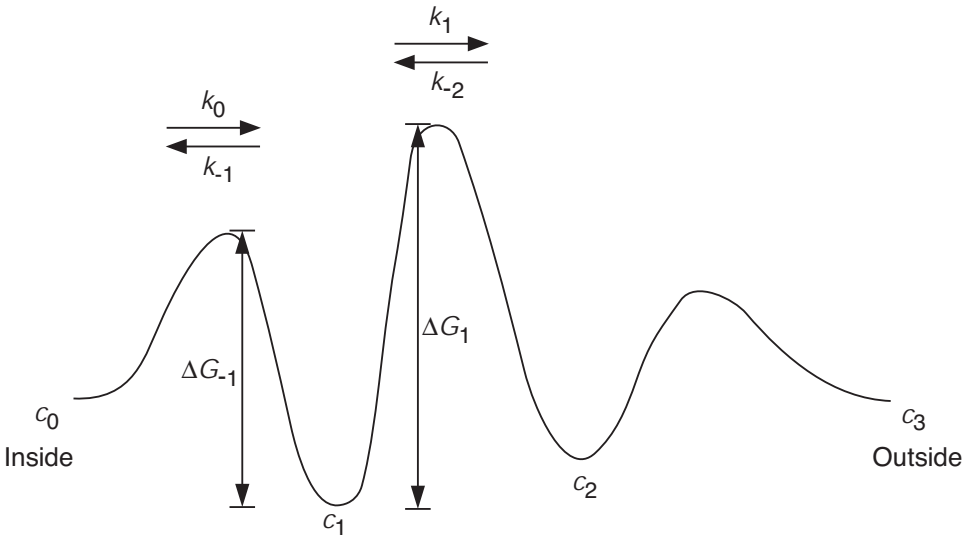

Figure 3.6: General potential energy profile for barrier models. The local minima correspond to binding sites within the channel, and the local maxima are barriers that impede the ion flow. An ion progresses through the channel by hopping over the barriers from one binding site to another.

In either case, one does not expect the movement of each ion through the channel to be independent of other ions, and so one expects the independence principle 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 unexpected.

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 barrier that it must cross. Thus, in the notation of the diagram,

$$k_j = \kappa \exp\left(\frac{-\Delta G_j}{RT}\right), \tag{3.44}$$

for some factor κ with units of 1/time. One of the most difficult questions in the use of this expression is deciding on the precise form of the factor. According to Eyring rate theory (as used in this context by Hille (1992), 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 whether at room temperature 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). We do not enter this debate here, as it is unnecessary for our purposes. All we require is some factor, of plausible value, that can be used to fit the rate expressions to experimental data.

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

$$k_j = \kappa \exp\left[\frac{1}{RT}(-\Delta G_j + zF\Delta V_{j+1}/2)\right], \tag{3.45}$$

$$k_{-j} = \kappa \exp\left[\frac{1}{RT}(-\Delta G_{-j} - zF\Delta V_j/2)\right]. \tag{3.46}$$

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

$$\sum_{j=0}^{n-1} \Delta G_j - \sum_{j=1}^{n} \Delta G_{-j} = 0. \tag{3.47}$$

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

$$\Delta G_j = \Delta G_0 - j\delta G, \tag{3.48}$$

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

$$J = \lambda(k_{j-1}c_{j-1} - k_{-j}c_j). \tag{3.49}$$

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 get a system of linear equations,

$$k_0c_0 - k_{-1}c_1 = k_1c_1 - k_{-2}c_2 = \cdots = k_{n-1}c_{n-1} - k_{-n}c_n = M, \tag{3.50}$$

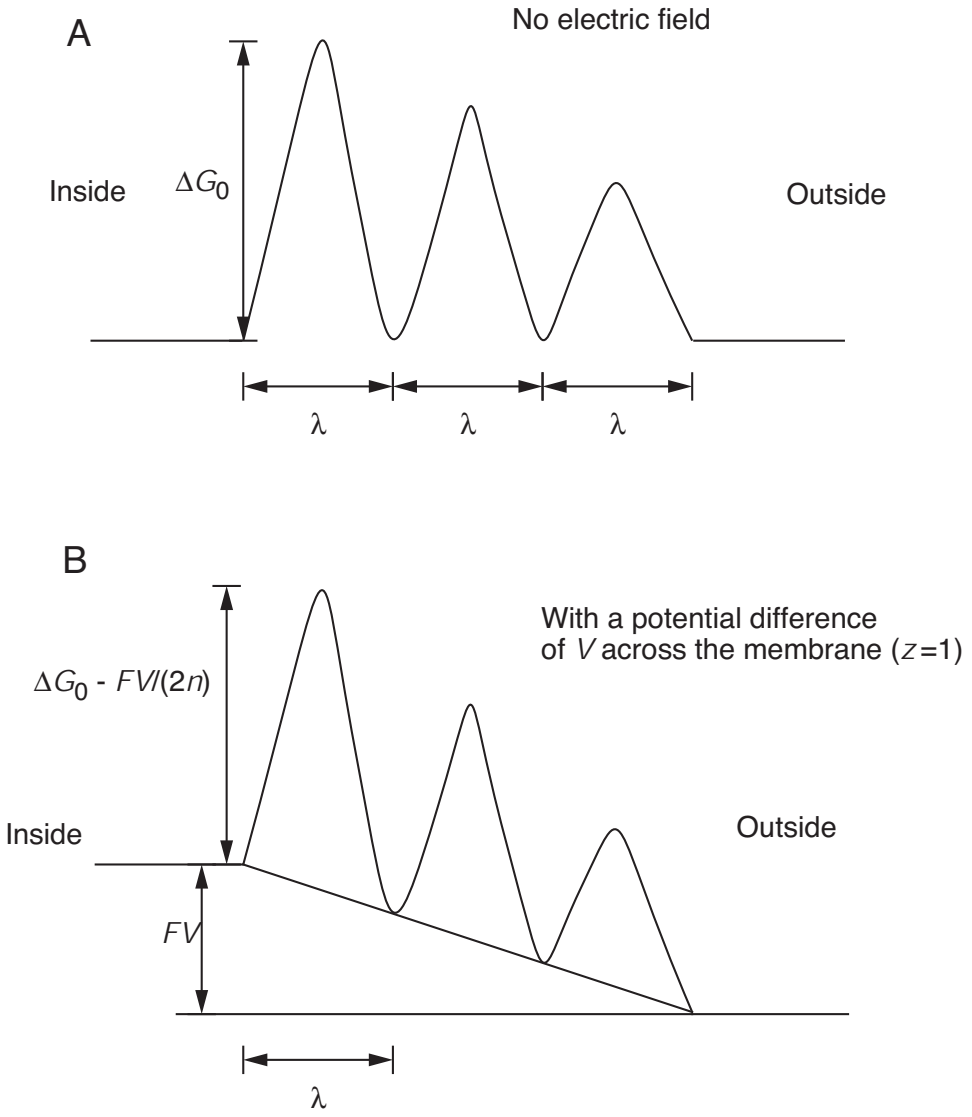

Figure 3.7: The potential energy diagram used in the nonsaturating model of Woodbury (1971). There is an equal distance between the binding sites, and the barriers are symmetrical. A. In the absence of an electric field the barrier height decreases linearly through the membrane. B. The presence of a constant electric field skews the energy profile, bringing the outside end 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.

where $M = J/\lambda$ is a constant. Hence

$$\begin{aligned} k_0 c_0 &= (k_1 + k_{-1})c_1 - k_{-2}c_2, && (3.51)\\ k_1 c_1 &= (k_2 + k_{-2})c_2 - k_{-3}c_3, && (3.52)\\ k_2 c_2 &= (k_3 + k_{-3})c_3 - k_{-4}c_4, && (3.53)\\ &\vdots \end{aligned}$$

Eventually, we need to determine J in terms of the concentrations on either side of the membrane, c_0 and c_n. Solving (3.52) for c_1 and substituting into (3.51) gives

$$k_0 c_0 = c_2 k_2 \left(1 + \frac{k_{-1}}{k_1} + \frac{k_{-1}k_{-2}}{k_1 k_2}\right) - c_3 k_{-3}\left(1 + \frac{k_{-1}}{k_1}\right), \tag{3.54}$$

and then solving (3.53) for c_2 and substituting into (3.54) gives

$$k_0 c_0 = c_3 k_3 \left(1 + \frac{k_{-1}}{k_1} + \frac{k_{-1}k_{-2}}{k_1 k_2} + \frac{k_{-1}k_{-2}k_{-3}}{k_1 k_2 k_3}\right) - c_4 k_{-4}\left(1 + \frac{k_{-1}}{k_1} + \frac{k_{-1}k_{-2}}{k_1 k_2}\right). \tag{3.55}$$

Repeating this process of sequential substitutions, and letting

$$\phi_n = 1 + \frac{k_{-1}}{k_1} + \frac{k_{-1}k_{-2}}{k_1 k_2} + \cdots + \frac{k_{-1}\cdots k_{-n}}{k_1 \cdots k_n}, \tag{3.56}$$

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

$$J = \lambda M = \frac{\lambda k_0 \left(c_0 - c_n \frac{k_{-1}\cdots k_{-n}}{k_0 \cdots k_{n-1}}\right)}{1 + \frac{k_{-1}}{k_1} + \frac{k_{-1}k_{-2}}{k_1 k_2} + \cdots + \frac{k_{-1}\cdots k_{-(n-1)}}{k_1 \cdots k_{n-1}}}. \tag{3.60}$$

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

$$\begin{aligned} \Delta G_j &= \Delta G_0 - j\delta G - \frac{FzV}{2n}, && (3.61)\\ \Delta G_{-j} &= \Delta G_0 - (j-1)\delta G + \frac{FzV}{2n}. && (3.62) \end{aligned}$$

Now we use (3.44) 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

$$\begin{aligned} J &= \frac{k_0\lambda(c_0 - c_n e^{-v})}{1 + e^{-(g+v/n)} + e^{-2(g+v/n)} + \cdots + e^{-(n-1)(g+v/n)}}, \\ &= k_0\lambda(c_0 - c_n e^{-v})\frac{e^{-(g+v/n)} - 1}{e^{-n(g+v/n)} - 1}. \end{aligned} \tag{3.64}$$

As expected, (3.64) 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,

$$J = \frac{k_{00}\lambda^2}{L} \cdot v \cdot \frac{c_0 - c_n e^{-v}}{1 - e^{-v}}, \tag{3.65}$$

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.

It follows that in the homogeneous membrane case,

$$\begin{aligned} J &= \frac{D_{\mathrm{S}}}{L} \cdot v \cdot \frac{c_0 - c_n e^{-v}}{1 - e^{-v}}, \\ &= P_{\mathrm{S}} \cdot v \cdot \frac{c_0 - c_n e^{-v}}{1 - e^{-v}}, \end{aligned} \tag{3.66}$$

which is exactly the GHK current equation (2.119) 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 we will see, 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, 1992). 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

$$X + S_i \underset{k_{-1}}{\overset{k_0}{\rightleftharpoons}} XS \underset{k_{-2}}{\overset{k_1}{\rightleftharpoons}} X + S_e. \tag{3.67}$$

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,

$$k_0 c_0 x - k_{-1} c_1 = k_1 c_1 - k_{-2} c_2 x, \tag{3.68}$$

where x denotes the probability that the binding site is empty. Note that (3.68) is similar to the corresponding equation for the nonsaturating pore, (3.50), with the only difference that x appears in the saturating model. In addition, we have a conservation equation for x,

$$x + c_1 = 1. \tag{3.69}$$

Solution of (3.68) and (3.69) gives the flux J as

$$J = k_0 c_0 x - k_{-1} c_1 = \frac{k_0 k_1 c_0 - k_{-1} k_{-2} c_2}{k_0 c_0 + k_{-2} c_2 + k_{-1} + k_1}. \tag{3.70}$$

It is important to note that J, as defined by (3.70), does not have the same units (concentration $\times$ distance/time) as we used previously, but instead has units of number of ions crossing the membrane per unit time. The corresponding transmembrane current, 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 (3.44) 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

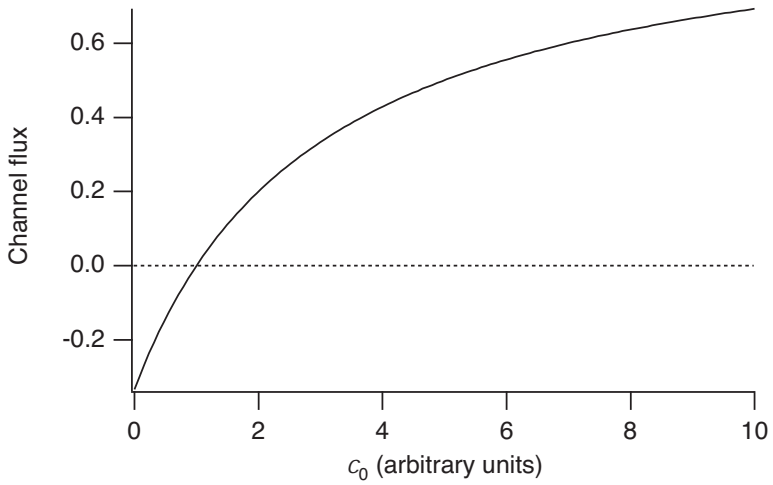

Figure 3.8: Plot of J against c_0 for the simplest saturating model with one binding site. When c_0 is small, the flux is approximately a linear function of c_0, but as c_0 increases, the flux saturates to a maximum value.

side, it follows that

$$k_0 = \kappa \exp\left[\frac{1}{RT}\left(-\Delta G_0 + \delta FV/2\right)\right], \tag{3.71}$$

$$k_1 = \kappa \exp\left[\frac{1}{RT}\left(-\Delta G_1 + (1-\delta)FV/2\right)\right], \tag{3.72}$$

$$k_{-1} = \kappa \exp\left[\frac{1}{RT}\left(-\Delta G_{-1} - \delta FV/2\right)\right], \tag{3.73}$$

$$k_{-2} = \kappa \exp\left[\frac{1}{RT}\left(-\Delta G_{-2} - (1-\delta)FV/2\right)\right]. \tag{3.74}$$

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 $\bar{\mathrm{S}}$, 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 $\bar{\mathrm{S}}$ on the right. We let c denote [S] and $\bar{c}$ denote [$\bar{\mathrm{S}}$]. Since S and $\bar{\mathrm{S}}$ have identical energy profiles in the channel, the rate constants for the passage of $\bar{\mathrm{S}}$ through the channel are the same as those for S. From the kinetic schemes for S and $\bar{\mathrm{S}}$ we get

$$k_0 c_0 x - k_{-1} c_1 = k_1 c_1 - k_{-2} c_2 x = J_{\mathrm{S}}, \tag{3.75}$$

$$k_0 \bar{c}_0 x - k_{-1} \bar{c}_1 = k_1 \bar{c}_1 - k_{-2} \bar{c}_2 x = J_{\bar{\mathrm{S}}}, \tag{3.76}$$

but here the conservation equation for x is

$$x + \bar{c}_1 + c_1 = 1. \tag{3.77}$$

To calculate the individual fluxes of S and $\bar{\mathrm{S}}$ it is necessary to eliminate x from (3.75) and (3.76) using the conservation equation (3.77). However, to calculate the flux ratio this is not necessary. Solving (3.75) for J_{S} in terms of x, c_0, and c_2, we find

$$J_{\mathrm{S}} = x \left(\frac{k_0 c_0 - \dfrac{k_{-1} k_{-2}}{k_1} c_2}{1 + k_{-1}/k_1} \right), \tag{3.78}$$

and similarly,

$$J_{\bar{\mathrm{S}}} = x \left(\frac{k_0 \bar{c}_0 - \dfrac{k_{-1} k_{-2}}{k_1} \bar{c}_2}{1 + k_{-1}/k_1} \right). \tag{3.79}$$

The variable x cancels when we calculate the flux ratio, and so we do not need to use the conservation equation. If S is present only on the left-hand side and $\bar{\mathrm{S}}$ only on the right, we then have $c_2 = 0$ and $\bar{c}_0 = 0$, in which case

$$\frac{J_{\mathrm{S}}}{J_{\bar{\mathrm{S}}}} = -\frac{k_0 k_1}{k_{-1} k_{-2}} \cdot \frac{c_0}{\bar{c}_2}. \tag{3.80}$$

The minus sign on the right-hand side appears because the fluxes are in different directions. Now we substitute for the rate constants, (3.71) to (3.74), 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

$$\left| \frac{J_{\mathrm{S}}}{J_{\bar{\mathrm{S}}}} \right| = \exp\left(\frac{VF}{RT} \right) \cdot \frac{c_0}{\bar{c}_2}, \tag{3.81}$$

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

$$k_0 c_0 x - k_{-1} c_1 = k_1 c_1 - k_{-2} c_2 = \cdots = k_{n-1} c_{n-1} - k_{-n} c_n x = J, \tag{3.82}$$

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

$$x = 1 - \sum_{i=1}^{n-1} c_i. \tag{3.83}$$

For convenience we define

$$\pi_j = \frac{k_{-1} \cdots k_{-j}}{k_1 \cdots k_j}, \qquad \pi_0 = 1, \tag{3.84}$$

$$\phi_j = \sum_{i=0}^{j} \pi_i. \tag{3.85}$$

Note that this definition for ϕ_j is the same as that given in (3.56). It is left as an exercise to show that

$$J = \frac{k_0 c_0 - k_{-n} c_n \pi_{n-1}}{\phi_{n-1} + \beta k_0 c_0 + k_{-n} c_n (\alpha \phi_{n-1} - \beta \phi_{n-2})}, \tag{3.86}$$

where

$$\alpha = \sum_{j=1}^{n-1} \frac{\phi_{n-2} - \phi_{j-1}}{k_j \pi_j}, \tag{3.87}$$

$$\beta = \sum_{j=1}^{n-1} \frac{\phi_{n-1} - \phi_{j-1}}{k_j \pi_j}. \tag{3.88}$$

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).

A model for sodium channels

Saturable one-ion barrier models have been used by a large number of authors to describe ion channels that show deviations from the independence principle (see, for example, Hille, 1975; Ciani and Ribalet, 1988; Robello et al., 1987). Here we focus our attention on the model of Hille (1975).

An example of Hille's experimental results is given in Fig. 3.9. A frog myelinated nerve was voltage clamped, held at -80 mV, and then depolarized in brief steps. The peak of the Na^+ current is plotted as a function of the voltage. If the Na^+ channels obeyed the independence principle, the resultant currents would then be

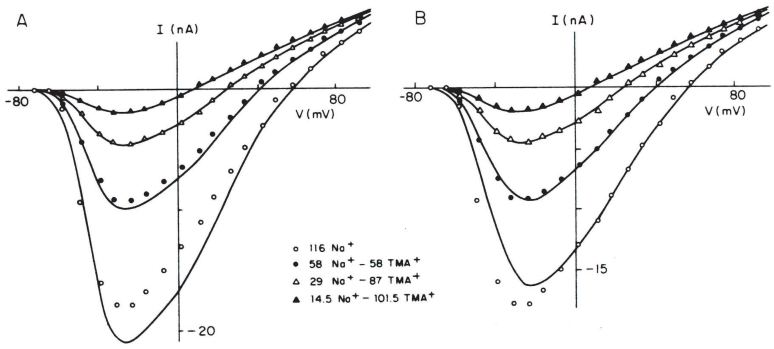

Figure 3.9: Peak current plotted against voltage, for a range of external Na^+ concentrations (given in mM in the legend). For the model calculations, the internal Na^+ concentration was assumed to be 11.2 mM. A: The smooth curves are the predictions from the independence relation. B: the smooth curves are the predictions from the Hille four-barrier model, showing saturation of the flux at high Na^+ concentrations. (Hille, 1975, Fig. 5.)

predicted by (3.14). Thus, if the external Na^+ concentration is changed from $[Na^+]_e$ to $[Na^+]'_e$ while keeping the same internal $[Na^+]_i$, we expect

$$\frac{I'_{\rm Na}}{I_{\rm Na}} = \frac{[{\rm Na}^+]_i - [{\rm Na}^+]'_e \exp(\frac{-VF}{RT})}{[{\rm Na}^+]_i - [{\rm Na}^+]_e \exp(\frac{-VF}{RT})}. \tag{3.89}$$

The smooth curves in Fig. 3.9A are drawn from this relation, using the curve for $[Na^+]_e = 14.5$ mM as a reference curve. It is necessary to use one of the curves as a reference curve, because (3.89) determines only the ratio of the currents, not their absolute values. Clearly, the curves predicted from the independence principle agree with the data at low Na^+ concentrations, but do not agree with the data at high $[Na^+]$. In the latter case the observed Na^+ currents are smaller than predicted, suggesting that the Na^+ channel is saturated.

To explain the observed deviation from independence, Hille proposed a 4-barrier, 3 binding site model, sketched schematically in Fig. 3.10. As usual, each rate constant is described in terms of the free-energy profile and the voltage drop across the barrier, as in (3.71)–(3.74). For example,

$$k_0 = \kappa \exp\left[\frac{1}{RT}(-\Delta G_0 + \delta_1 FV/2)\right], \tag{3.90}$$

$$k_1 = \kappa \exp\left[\frac{1}{RT}(-\Delta G_1 + \delta_2 FV/2)\right], \tag{3.91}$$

$$k_{-1} = \kappa \exp\left[\frac{1}{RT}(-\Delta G_{-1} - \delta_1 FV/2)\right], \tag{3.92}$$

$$k_{-2} = \kappa \exp\left[\frac{1}{RT}(-\Delta G_{-2} - \delta_2 FV/2)\right], \tag{3.93}$$

and similarly for the other rate constants. The model is completely specified by

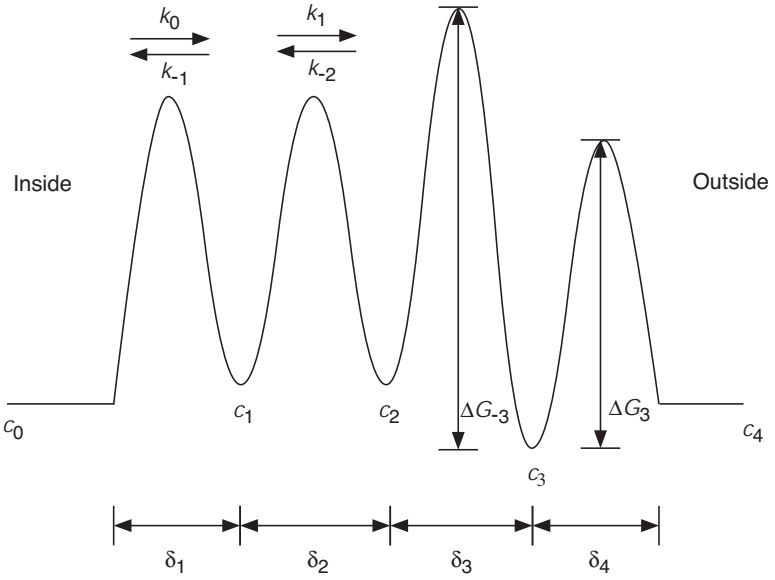

Figure 3.10: Schematic diagram of the potential energy profile of the Hille four-barrier model of the Na^+ channel.

the parameters ΔG_i, ΔG_{-i}, and δ_i. All but two of these parameters were fixed at reasonable values, and the values of ΔG_3 and ΔG_{-3} were varied to obtain agreement with experimental data. The standard parameter set used to plot the curves in Figs. 3.9B and 3.11 is given in Table 3.1.

With these parameters, the channel flux and current ratios are calculated using the methods of the previous two sections. Plotting the current ratio against the voltage gives the smooth curves shown in Fig. 3.9B. Clearly, the saturating model shows better agreement with the experimental data at high $[Na^+]$.

A typical I–V curve of the model with the standard parameter set is shown in Fig. 3.11. A linear I–V curve and the GHK current equation are also given for the sake of comparison. The I–V curve of the Hille model has characteristics reminiscent of both the linear and the GHK curves. At large negative voltages the Hille model is similar to the GHK curve, but at large positive voltages it is more similar to the linear I–V curve. All three curves have the same reversal potential, as indeed they must, since the Nernst potential is model-independent.

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

$\Delta G_0 = 7$	$\Delta G_{-1} = 6.5$	$\delta_1 = 0.25$
$\Delta G_1 = 6.5$	$\Delta G_{-2} = 6.5$	$\delta_2 = 0.25$
$\Delta G_2 = 8.5$	$\Delta G_{-3} = 10$	$\delta_3 = 0.23$
$\Delta G_3 = 7$	$\Delta G_{-4} = 6$	$\delta_4 = 0.27$

Table 3.1: Standard parameter set for the Hille model of the Na^+ channel (Hille, 1975). The ΔG_i and ΔG_{-i} are in units of RT. Note that $\sum_{i=1}^{4}(\Delta G_{i-1} - \Delta G_{-i}) = 0$, since the ends of the energy profile are at the same level, arbitrarily set to be $G = 0$. Also, $\sum_{i=1}^{4} \delta_i = 1$, since δ_i represents the fraction of the total voltage drop over the ith barrier. To determine these parameters, it was assumed that $\kappa = kT/h$, where k is Boltzmann's constant, and h is Planck's constant.

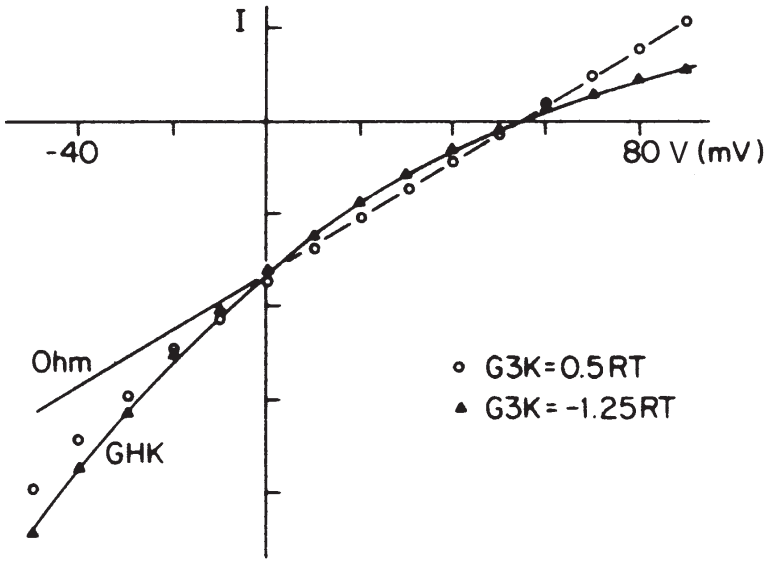

Figure 3.11: The I–V curve of the Hille model (open circles), compared to a linear I–V curve (labeled Ohm), and the GHK I–V curve. The current scale marks correspond to 4–5 pA per channel. (Hille, 1975, Fig. 4.)

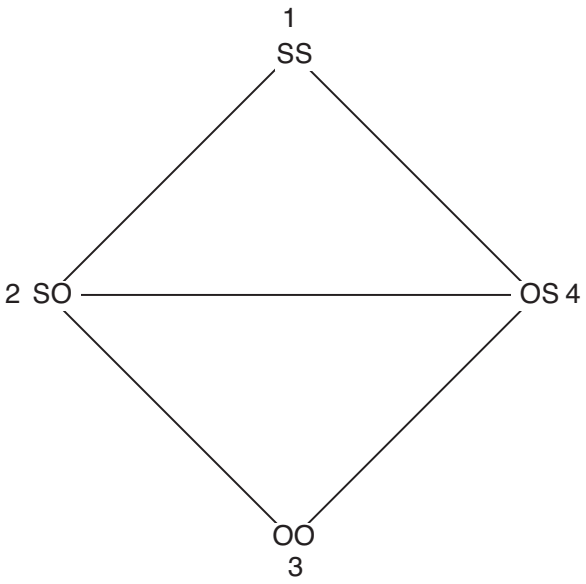

Figure 3.12: State diagram for a multi-ion barrier model with two binding sites and a single ion.

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). We do not have space for a detailed discussion of the properties of these models, but present only a brief discussion of the simplest model. Hille and Schwarz (1978) and Hille (1992) 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. 3.12). Arbitrary move-

ments from one state to another are not possible. For example, the state OO (where both binding 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. 3.12 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

$$\frac{dP_1}{dt} = -(k_{12}+k_{14})P_1 + k_{21}c_eP_2 + k_{41}c_iP_4, \tag{3.94}$$

$$\frac{dP_2}{dt} = -(k_{21}c_e+k_{23}+k_{24})P_2 + k_{12}P_1 + c_ik_{32}P_3 + k_{42}P_4, \tag{3.95}$$

$$\frac{dP_3}{dt} = -(c_ik_{32}+c_ek_{34})P_3 + k_{43}P_4 + k_{23}P_2, \tag{3.96}$$

$$\frac{dP_4}{dt} = -(k_{41}c_i+k_{42}+k_{43})P_4 + k_{14}P_1 + k_{24}P_2 + c_ek_{34}P_3. \tag{3.97}$$

The probabilities must also satisfy the conservation equation

$$\sum_{i=1}^{4} P_i = 1. \tag{3.98}$$

Using the conservation equation in place of the equation for P_4, the steady-state probability distribution is given by the linear system

$$\begin{pmatrix} -k_{12}-k_{14} & k_{21} & 0 & k_{41} \\ k_{12} & -k_{21}-k_{23}-k_{24} & c_ek_{32} & k_{42} \\ 0 & k_{23} & -c_ek_{32}-c_ik_{34} & k_{43} \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} P_1 \\ P_2 \\ P_3 \\ P_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{3.99}$$

Since each rate constant is determined as a function of the voltage in the same way as one-ion models (as in, for example, (3.90)–(3.93)), solution of (3.98) gives each P 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

$$J = P_2k_{24} - P_4k_{42}. \tag{3.100}$$

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 one-ion models.

3.4.4 Electrogenic pumps and exchangers

In Chapter 2 we saw how detailed balance and chemical potential required that rate constants in models for electrogenic exchanger and pumps were dependent on the

membrane potential. See, for instance, (2.77) or (2.95). However, although the arguments from chemical equilibrium shown that such voltage-dependency must exist, they don't tell us exactly which rate constants depend on the voltage, or what the functional dependency is. As we have come to expect by now, it is much more difficult to answer these questions, as they rely on how one models non-equilibrium rates rather than merely equilibrium situations. 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 exchanger or pump model could conceivably depend on membrane potential. In other words, not only do we have a number of ways to model the voltage-dependence when it occurs, we also have 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. 2.10. 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 of V, then we can model the rate constants as

$$\begin{aligned} k_2 &= \kappa \exp\left[\frac{1}{RT}(-\Delta G_+ + 2FV/2)\right] && (3.101)\\ &= \bar{k}_2 \exp\left(\frac{FV}{RT}\right), && (3.102)\\ k_{-2} &= \kappa \exp\left[\frac{1}{RT}(-\Delta G_- - 2FV/2)\right] && (3.103)\\ &= \bar{k}_{-2} \exp\left(\frac{-FV}{RT}\right), && (3.104) \end{aligned}$$

where $\bar{k}_2 = \kappa \exp[-\Delta G_+/(RT)]$ and similarly for $\bar{k}_{-2}$. In (3.101) and (3.103) $2FV$ 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 get similar equations for those rate constants. Then

$$\begin{aligned} \frac{k_{-2}}{k_2}\frac{k_{-4}}{k_4} &= \frac{\bar{k}_{-2}}{\bar{k}_2}\frac{\bar{k}_{-4}}{\bar{k}_4} \exp\left(\frac{-2FV}{RT}\right) \exp\left(\frac{3FV}{RT}\right) && (3.105)\\ &= \frac{\bar{k}_{-2}}{\bar{k}_2}\frac{\bar{k}_{-4}}{\bar{k}_4} \exp\left(\frac{FV}{RT}\right), && (3.106) \end{aligned}$$

and thus the necessary equilibrium condition, $K_1K_2K_3K_4 = \exp\left(\frac{FV}{RT}\right)$ holds (cf. (2.77)).

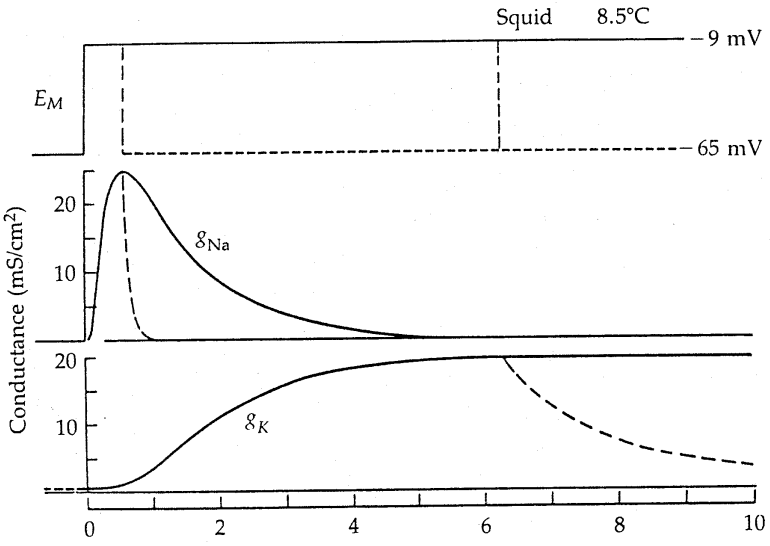

Figure 3.13: Na^+ and K^+ conductances as a function of time after a step change in voltage from −65 mV to −9 mV. The dashed line shows how after repolarization g_{Na} recovers quickly, and g_K recovers more slowly. (Hille, 1992, Fig. 11, p. 40.)

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 probably of equally great interest to determine how ionic channels open and close in response to voltage. As we will see in Chapter 4, 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 two terms,

$$I = g(V,t)\phi(V), \tag{3.107}$$

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. In the previous sections we discussed electrodiffusion and barrier models for $\phi(V)$; in this section we discuss models for the dependence of g on voltage and time.

Consider, for example, the curves in Fig. 3.13, 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. From this data we can draw the following 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 moves to an inactivated state. Thus, Na^+ channels first activate and then inactivate.

3.5.1 A Two-State K^+ Channel

The simplest model for 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,

$$\mathrm{C} \underset{\beta(V)}{\overset{\alpha(V)}{\rightleftharpoons}} \mathrm{O}. \tag{3.108}$$

Letting g denote the proportion of channels in the open state, we can write the differential equation for the rate of change of g as

$$\frac{dg}{dt} = \alpha(V)(1-g) - \beta(V)g, \tag{3.109}$$

where we have used the fact that because channels are conserved, the proportion of closed channels is $1-g$. Under voltage-clamp conditions (i.e., where the voltage is held fixed, as in Fig. 3.13), α and β are constants, and thus we can readily solve for g as a function of time. It is often convenient to write (3.109) as

$$\tau_g(V)\frac{dg}{dt} = g_\infty(V) - g, \tag{3.110}$$

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 that shown in Fig. 3.13, one can obtain values for g_∞ and τ_g, and thus α and β can be unambiguously determined.

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, $\mathrm{S}_{00}, \mathrm{S}_{10}, \mathrm{S}_{01}$, or S_{11}, where the subscripts denote the different subunits, with 1 and 0 denoting open and closed subunits, respectively. A general model for 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 $\mathrm{S}_{00} + \mathrm{S}_{10} + \mathrm{S}_{01} + \mathrm{S}_{11} = 1$), but we can simplify the model by grouping the channel states with the same number of closed and open subunits. For example, 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

$$S_0 \underset{\beta}{\overset{2\alpha}{\rightleftarrows}} S_1 \underset{2\beta}{\overset{\alpha}{\rightleftarrows}} S_2. \tag{3.111}$$

The corresponding differential equations are

$$\frac{dx_0}{dt} = \beta x_1 - 2\alpha x_0, \tag{3.112}$$

$$\frac{dx_2}{dt} = \alpha x_1 - 2\beta x_2, \tag{3.113}$$

where x_i denotes the proportion of channels in state S_i, and $x_0 + x_1 + x_2 = 1$. We make the change of variables $x_2 = n^2$, where n satisfies the differential equation

$$\frac{dn}{dt} = \alpha(1-n) - \beta n. \tag{3.114}$$

A simple substitution then shows that (3.112) and (3.113) are satisfied by $x_0 = (1-n)^2$ and $x_1 = 2n(1-n)$. Thus, (3.112) and (3.113) are equivalent to $x_0 = (1-n)^2, x_1 = 2n(1-n), x_2 = n^2$, where n satisfies (3.114).

In fact, we can derive a stronger result. We let

$$x_0 = (1-n)^2 + y_0, \tag{3.115}$$

$$x_2 = n^2 + y_2, \tag{3.116}$$

so that of necessity, $x_1 = 2n(1-n) - y_0 - y_2$. It follows that

$$\frac{dy_0}{dt} = -2\alpha y_0 - \beta(y_0 + y_2), \tag{3.117}$$

$$\frac{dy_2}{dt} = -\alpha(y_0 + y_2) - 2\beta y_2. \tag{3.118}$$

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 (3.114) (see Exercise 16). Thus, the channel conductance is proportional to n^k, where n satisfies the simple equation (3.114). This multiple subunit model for channel gating provides the basis for the model of excitability that we examine in the next chapter.

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

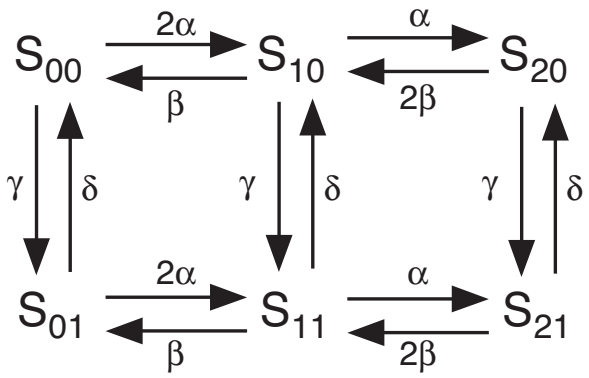

Figure 3.14: Diagram of the possible states in a model of the Na^+ channel.

analysis to the case of multiple subunits of two different types, m and h, say, where each subunit can be either closed or 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. 3.14. 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}. As above, a simple substitution shows that the reaction scheme is equivalent to

$$x_{21} = m^2 h, \tag{3.119}$$

$$\frac{dm}{dt} = \alpha(1-m) - \beta m, \tag{3.120}$$

$$\frac{dh}{dt} = \gamma(1-h) - \delta h, \tag{3.121}$$

where the other variables are given by $x_{00} = (1-m)^2(1-h)$, $x_{10} = 2m(1-m)(1-h)$, $x_{20} = m^2(1-h)$, $x_{01} = (1-m)^2 h$, and $x_{11} = 2m(1-m)h$. Furthermore, the invariant manifold is again 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 4.

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. 3.15A). Transitions between states are described by

$$\mathrm{C} \underset{\beta}{\overset{\alpha}{\rightleftharpoons}} \mathrm{O}, \qquad \mathrm{O} \xrightarrow{\gamma} \mathrm{I}, \qquad \mathrm{C} \xrightarrow{\delta} \mathrm{I}. \tag{3.122}$$

Thus, the state I is absorbing. While this is clearly not true in general, it is a reasonable approximation at high depolarizations. As before, we let g denote the proportion of open channels and let c denote the proportion of closed channels. Then,

$$\frac{dc}{dt} = -(\alpha+\delta)c + \beta g, \tag{3.123}$$

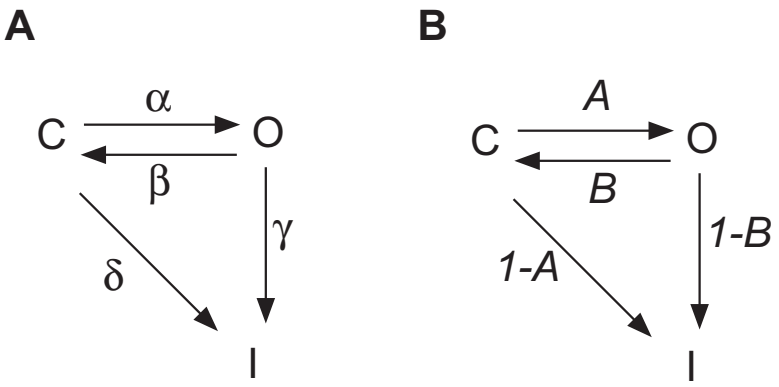

Figure 3.15: A: Schematic diagram of the states of the Na^+ channel. C, O, and I denote the closed, open, and inactivated states, respectively. B: Time-independent transition probability diagram.

$$\frac{dg}{dt} = \alpha c - (\beta + \gamma)g, \tag{3.124}$$

where as before, we have used the conservation of channels to eliminate the proportion of channels in the inactivated state. Initial conditions are $c(0) = 1$, $g(0) = 0$, i.e., all the channels are initially closed. By differentiating the equation for s we eliminate c to get

$$\frac{d^2g}{dt^2} + (\alpha + \beta + \gamma + \delta)\frac{dg}{dt} + [(\alpha + \delta)(\beta + \gamma) - \alpha\beta]g = 0, \tag{3.125}$$

which now has the initial conditions $g(0) = 0$, $g'(0) = \alpha$. This can be solved directly to give

$$g = a(e^{\lambda_1 t} - e^{\lambda_2 t}), \tag{3.126}$$

where $\lambda_2 < \lambda_1 < 0$ are the roots of

$$\lambda^2 + (\alpha + \beta + \gamma + \delta)\lambda + (\alpha + \delta)(\beta + \gamma) - \alpha\beta = 0, \tag{3.127}$$

and where

$$\alpha = a(\lambda_1 - \lambda_2) > 0. \tag{3.128}$$

As in the simple two-state model, a, λ_1, and λ_2 can be determined from experimental data. However, the rate constants cannot be determined uniquely. For since λ_1 and λ_2 are the roots of (3.127), it follows that

$$\alpha + \beta + \gamma + \delta = -\lambda_1 - \lambda_2, \tag{3.129}$$

$$(\alpha + \delta)(\beta + \gamma) - \alpha\beta = \lambda_1\lambda_2. \tag{3.130}$$

Here we have only three equations for the four unknowns, α, β, γ, and δ, so the system is underdetermined (see Exercise 18). This problem cannot be resolved using the macroscopic data that has been discussed so far, but requires data collected from a single channel, as described in the next section.

Single-channel recordings

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, 1983; Neher and Sakmann received the 1991 Nobel Prize in physiology for their development of the patch-clamp technique). An example of an experimental record is given in Fig. 3.16. The current through an individual channel is stochastic (panel A) and cannot be predicted as 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. 3.13. However, the single-channel recordings contain more information than does the ensemble average.

To describe a channel with n possible states (open, closed, inactivated, etc.), we introduce a stochastic variable $S(t) \in 1, 2, \ldots, n$ such that $S(t) = i$ if the channel is in state i at time t. Further, if k_{ij} (independent of time) is the rate constant for transitions from state i to state j, then the probability that the channel changes from state i to state j in the time interval $(t, t + dt)$ is $k_{ij}dt$. In more condensed notation we write

$$P[S(t + dt) = j|S(t) = i] = k_{ij}dt. \tag{3.131}$$

Note that this is valid only approximately and for small dt, since for large dt and k_{ij} fixed, this probability will exceed 1. Here, $P[x|y]$ is a conditional probability, meaning the probability of x given y. Also, the probability that the channel does not change state in the time interval $(t, t + dt)$ is given by

$$P[S(t + dt) = i|S(t) = i] = 1 - K_i dt, \tag{3.132}$$

where $K_i = \sum_{j=1, j\neq i}^{n} k_{ij}$.

We now calculate the probability that the channel stays in state i for time t. Let $M_i(t)$ be the logical random variable $\{S(\tau) = i, 0 < \tau < t\}$. If the interval $(0, t)$ is divided into m subintervals, each of length t/m, then the channel stays in state i only if it does not change state during any of the subintervals. Therefore,

$$P[M_i(t)|S(0) = i] = \left(1 - \frac{K_i t}{m}\right)^m. \tag{3.133}$$

Taking the limit $m \to \infty$, we find

$$P[M_i(t)|S(0) = i] = e^{-K_i t}. \tag{3.134}$$

To apply this theory to the determination of channel kinetics, we consider the probability diagram for the channel states (Fig. 3.15B). Here A denotes the time-independent probability that a channel in the closed state moves to the open state (rather than to the inactivated state), and B denotes the probability that a channel in the open state moves to the closed state. Comparing Fig. 3.15A with Fig. 3.15B, we see that

$$A = \frac{\alpha}{\alpha + \delta}, \qquad B = \frac{\beta}{\beta + \gamma}. \tag{3.135}$$

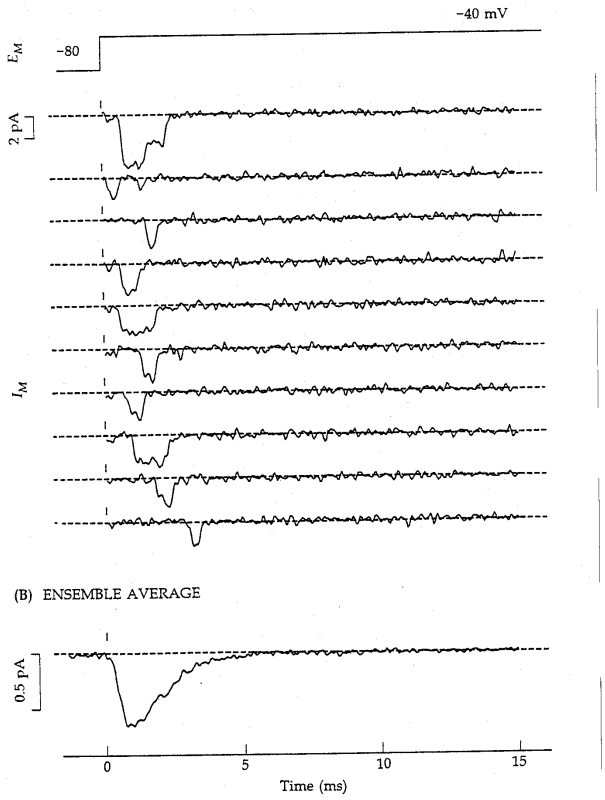

Figure 3.16: A: Na^+ currents from a single channel (or possibly two in the first trace) following a voltage step from -80 mV to -40 mV. B: Average open probability of the Na^+ channel, obtained by averaging over many traces of the type shown in A. (Hille, 1992, Fig. 6, p. 68.)

The probability $1 - A$ is easily determined experimentally, as it is the probability that a channel in the closed state inactivates without ever opening. 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.

Now let T denote the time to first opening of the channel, often called the *latency* of the channel. T is easily measured experimentally. Then,

$$\begin{aligned} P[T > t] &= P[\text{First transition is to state I}] \\ &\quad + P[\text{First transition is to state O and } T > t] \\ &= 1 - A + P[\text{the transition is to state O}] \cdot P[\text{no transitions for time } t] \\ &= 1 - A + A \cdot e^{-(\alpha+\delta)t}. \end{aligned} \tag{3.136}$$

Thus, $P[T > t]$ is a decreasing exponential, and so $1-A$ and $\alpha+\delta$ can be determined by fitting an exponential to the experimental measurements of $P[T > t]$. Hence, α and δ are unambiguously determined from the latency of the channel.

To determine the remaining two rate constants, let N be the number of times the channel opens before it finally inactivates and determine the probability distribution for N. Clearly, $P[N = 0] = 1 - A$. Furthermore,

$$\begin{aligned} P[N = k] &= P[N = k \text{ and channel enters I from O}] \\ &\quad + P[N = k \text{ and channel enters I from C}] \\ &= A^k B^{k-1}(1 - B) + A^k B^k (1 - A) \\ &= (AB)^k \left(\frac{1 - AB}{B} \right). \end{aligned} \tag{3.137}$$

Since A can be determined from the latency, B can be determined from an experimental plot of $P[N = k]$ vs. k. Finally, the distribution of open times is given by $\exp[-(\beta + \gamma)t]$, and so $\beta + \gamma$ can be determined from the open time distribution of the channel. This completes the characterization of the channel rate constants.

Since the work of Hodgkin and Huxley (described in Chapter 4), 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. 3.13 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 $\alpha = 1/\text{ms}$, $\beta = 0.4/\text{ms}$, $\gamma = 1.6/\text{ms}$ and $\delta = 1/\text{ms}$ 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.

3.5.4 Drugs and Toxins

Many drugs act by blocking a specific ion channel. There are numerous specific channel blockers, such as sodium channel blockers, potassium channel blockers, calcium 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 (calcium-channel blocker), quinidine, sotolol, nicotine, DDT, various barbiturates (potassium-channel blockers), tetrodotoxin (TTX, the primary ingredient of puffer fish toxin), and scorpion toxins (sodium-channel blockers).

To include the effects of a drug or toxin like TTX in a model of a sodium channel is a relatively simple matter. We assume that a population P of sodium channels is available for ionic conduction and that a population B is blocked because they are bound by the toxin. Thus,

$$\mathrm{P} + \mathrm{D} \underset{k_-}{\overset{k_+}{\rightleftarrows}} \mathrm{B}, \tag{3.138}$$

where D represents the concentration of the drug. Clearly, $P + B = P_0$, so that

$$\frac{dP}{dt} = k_-(P_0 - P) - k_+ DP, \tag{3.139}$$

and the original channel conductance must be modified by multiplying by the percentage of unbound channels, P/P_0.

In steady state, we have

$$\frac{P}{P_0} = \frac{K_d}{K_d + D}. \tag{3.140}$$

The remarkable potency of TTX is reflected by its small equilibrium constant K_d, as $K_d \approx$ 1–5 nM for sodium channels in nerve cells, and $K_d \approx$ 1–10 μM for sodium 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* sodium-channel blockers, in that they interfere with the sodium channel only when it is open. Thus, the more the channel is used, the more likely it will be 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

$$\mathrm{C} \underset{\beta}{\overset{\alpha}{\rightleftarrows}} \mathrm{O}, \qquad \mathrm{CB} \underset{\beta}{\overset{\alpha}{\rightleftarrows}} \mathrm{OB} \tag{3.141}$$

$$\mathrm{CB} \xrightarrow{k_+} \mathrm{C} + \mathrm{D}, \qquad \mathrm{O} + \mathrm{D} \underset{k_-}{\overset{k_+}{\rightleftarrows}} \mathrm{OB}. \tag{3.142}$$

Notice that we have assumed that the drug does not interfere with the process of opening and closing, only with the actual 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.

Exercises

1. Derive the extended independence principle. Assume that there are more than one species of ion present, all with the same valence, and assume that the reversal potential is given by the GHK potential. Show that

$$\frac{I'}{I} = \frac{\sum_j P_j [S_j]'_i - \sum_j P_j [S_j]'_e \exp(\frac{-VF}{RT})}{\sum_j P_j [S_j]_i - \sum_j P_j [S_j]_e \exp(\frac{-VF}{RT})}, \tag{3.143}$$

   where the sum over j is over all the ionic species. Subscripts i and e denote internal and external concentrations, respectively.

2. 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.

3. In Section 3.3.1 we used the PNP equations 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 still 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 (3.64) satisfies the independence principle and the Ussing flux ratio.

   (b) Show that (3.64) can be made approximately linear by choosing g such that

$$ng = \ln\left(\frac{c_n}{c_0}\right). \tag{3.144}$$

   Although we know that 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 (3.82) and (3.83). First show that

$$J = x \cdot \frac{k_0 c_0 - k_{-n} c_n \pi_{n-1}}{\phi_{n-1}}. \tag{3.145}$$

Then show that

$$k_0 c_0 x = k_{n-1} c_{n-1} \phi_{n-1} - x k_{-n} c_n \phi_{n-2}, \tag{3.146}$$

$$k_j c_j = \frac{k_{n-1} c_{n-1}}{\pi_j} \cdot (\phi_{n-1} - \phi_{j-1}) - \frac{k_{-n} c_n x}{\pi_j} \cdot (\phi_{n-2} - \phi_{j-1}), \tag{3.147}$$

for $j = 1, \ldots, n-1$. Substitute these expressions into the conservation equation and solve for x.

7. Numerically plot some I–V curves of the Hille Na^+ channel model for a selection of values for $[Na^+]_e$ and $[Na^+]_i$ and for a range of parameter values, not only those in Table 3.1. Compare to the linear and GHK I–V curves.

8. By making a guess at the shape of the curve for $\left[Na^+\right] = 14.5$ mM in Fig. 3.9A, repeat the calculations to obtain the smooth curves in A and B of that figure. In other words, take an arbitrary curve of approximately the same shape as the $\left[Na^+\right] = 14.5$ mM curve and calculate the other smooth curves, first by using the independence principle, and second, by using the Hille model. (This is best done numerically.)

9. Write down 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 down the corresponding system of linear equations that determine the steady-state ionic concentrations at the channel binding sites.

10. By 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 8-state channel diagram must be used.) Hodgkin and Keynes predicted that the actual flux ratio is the Ussing ratio raised to the $(n+1)$st power (cf. (3.19)). How does n depend on the ionic concentrations on either side of the membrane, and on the energy profile?

11. 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 4-state multi-ion model is needed.

12. Suppose the sodium Nernst potential of a cell is 56 mV, its resting potential is -70 mV, and the extracellular calcium concentration is 1 mM. At what intracellular calcium concentration is the flux of a three-for-one sodium–calcium exchanger zero? (Use that $RT/F = 25.8$ mV at $27°C$.)

13. Modify the pump–leak model of Chapter 2 to include a calcium current and the 3-for-1 sodium–calcium exchanger. What effect does this modification have on the relationship between pump rate and membrane potential?

14. Because there is a net current, the sodium–potassium pump current must be voltage dependent. Determine this dependence by including voltage dependence in the rates of conformational change in expression (**??**). How does voltage dependence affect the pump–leak model of Chapter 2?

15. Intestinal epithelial cells have a glucose–sodium symport that transports one sodium 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?

16. 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

    $$S_0 \underset{\beta}{\overset{k\alpha}{\rightleftarrows}} S_1, \ldots, S_{k-1} \underset{k\beta}{\overset{\alpha}{\rightleftarrows}} S_k. \tag{3.148}$$

    (b) Derive the differential equation for x_j, the proportion of channels in state j.

    (c) 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 a stable invariant manifold for the system of differential equations, provided that

    $$\frac{dn}{dt} = \alpha(1-n) - \beta n. \tag{3.149}$$

17. Consider the model of the Na^+ channel shown in Fig. 3.14. Show that if α and β are large compared to γ and δ, then x_{21} is given (approximately) by

    $$x_{21} = \left(\frac{\alpha}{\alpha+\beta}\right)^2 h, \tag{3.150}$$

    $$\frac{dh}{dt} = \gamma(1-h) - \delta h, \tag{3.151}$$

    while conversely, if γ and δ are large compared to α and β, then (approximately)

    $$x_{21} = m^2 \left(\frac{\gamma}{\gamma+\delta}\right), \tag{3.152}$$

    $$\frac{dm}{dt} = \alpha(1-m) - \beta m. \tag{3.153}$$

18. Show that (3.127) has two negative real roots. Show that when $\beta = 0$ and $a \leq \frac{-\lambda_1}{\lambda_1 - \lambda_2}$, then (3.128)–(3.130) 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.

19. Write a computer program to simulate the response of a stochastic three-state Na^+ channel (Fig. 3.15A) to a voltage step. Take the ensemble average of many runs to reproduce the macroscopic behavior of Fig. 3.13. 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 calculate the rate constants of the simulation.

20. Find the differential equations describing the interaction of a two-state channel with a use-dependent blocker.

Chapter 4

Excitability

We have seen in previous chapters how the control of cell volume results in a potential difference across the cell membrane, and how this potential difference causes ionic currents to flow through channels in the cell membrane. Regulation of this membrane potential by control of the ionic channels is one of the most important cellular functions. Many cells, such as neurons and muscle cells, use the membrane potential as a signal, and thus the operation of the nervous system and muscle contraction (to name but two examples) are both dependent on the generation and propagation of electrical signals.

To understand electrical signaling in cells, it is helpful (and not too inaccurate) to divide all cell types into two groups: excitable cells and nonexcitable cells. Many cells maintain a stable equilibrium potential. For some, if currents are applied to the cell for a short period of time, the potential returns directly to its equilibrium value after the applied current is removed. Such cells are called nonexcitable, typical examples of which are the epithelial cells that line the walls of the gut. Photoreceptors are also nonexcitable, although in their case, membrane potential plays an extremely important signaling role nonetheless.

However, there are cells for which, if the applied current is sufficiently strong, the membrane potential goes through a large excursion, called an *action potential*, before eventually returning to rest. Such cells are called *excitable*. Excitable cells include cardiac cells, smooth and skeletal muscle cells, secretory cells, and most neurons. The most obvious advantage of excitability is that an excitable cell either responds in full to a stimulus or not at all, and thus a stimulus of sufficient amplitude may be reliably distinguished from background noise. In this way, noise is filtered out, and a signal is reliably transmitted.

There are many examples of excitability that occur in nature. A simple example of an excitable system is a household match. The chemical components of the match head are stable to small fluctuations in temperature, but a sufficiently large temperature fluctuation, caused, for example, by friction between the head and a rough surface, triggers the abrupt oxidation of these chemicals with a dramatic release of heat and light. The fuse of a stick of dynamite is a one-dimensional continuous version of an excitable medium, and a field of dry grass is its two-dimensional version. Both of these spatially extended systems admit the possibility

of wave propagation. The field of grass has one additional feature that the match and dynamite fuse fail to have, and that is recovery. While it is not very rapid by physiological standards, given a few months of growth, a burned-over field of grass will regrow enough fuel so that another fire may spread across it.

Although the generation and propagation of signals have been extensively studied by physiologists for at least the past 100 years, the most important landmark in these studies is the work of Alan Hodgkin and Andrew Huxley, who developed the first quantitative model of the propagation of an electrical signal along a squid giant axon (deemed "giant" because of the size of the axon, *not* the size of the squid). Their model was originally used to explain the action potential in the long giant axon of a squid nerve cell, but the ideas have since been extended and applied to a wide variety of excitable cells. Hodgkin–Huxley theory is remarkable, not only for its influence on electrophysiology, but also for its influence, after some filtering, on applied mathematics. FitzHugh (in particular) showed how the essentials of the excitable process could be distilled into a simpler model upon which mathematical analysis could make some progress. Because this simplified model turned out to be of such great theoretical interest, it contributed enormously to the formation of a new field of applied mathematics, the study of excitable systems, a field that continues to stimulate a vast amount of research.

Because of the central importance of cellular electrical activity in physiology, because of the importance of the Hodgkin–Huxley model in the study of electrical activity, and because it forms the basis for the study of excitability, it is no exaggeration to say that the Hodgkin–Huxley model is the most important model in all of the physiological literature.

4.1 The Hodgkin–Huxley Model

In Chapter 2 we described how the cell membrane can be modeled as a capacitor in parallel with an ionic current, resulting in the equation

$$C_m \frac{dV}{dt} + I_{\text{ion}}(V,t) = 0, \tag{4.1}$$

where V, as usual, denotes the internal minus the external potential ($V = V_i - V_e$). In the squid giant axon, as in many neural cells, the principal ionic currents are the sodium current and the potassium current. Although there are other ionic currents, primarily the chloride current, in the Hodgkin–Huxley theory they are small and lumped together into one current called the *leakage current*. Since the instantaneous I–V curves of open Na^+ and K^+ channels in the squid giant axon are approximately linear, (4.1) becomes

$$C_m \frac{dV}{dt} = -g_{\text{Na}}(V - V_{\text{Na}}) - g_{\text{K}}(V - V_{\text{K}}) - g_{\text{L}}(V - V_{\text{L}}) + I_{\text{app}}, \tag{4.2}$$

where I_{app} is the applied current. During an action potential there is a measured influx of 3.7 pmoles/cm^2 of sodium and a subsequent efflux of 4.3 pmoles/cm^2 of potassium. These amounts are so small that it is realistic to assume that the ionic concentrations, and hence the equilibrium potentials, are constant and unaffected

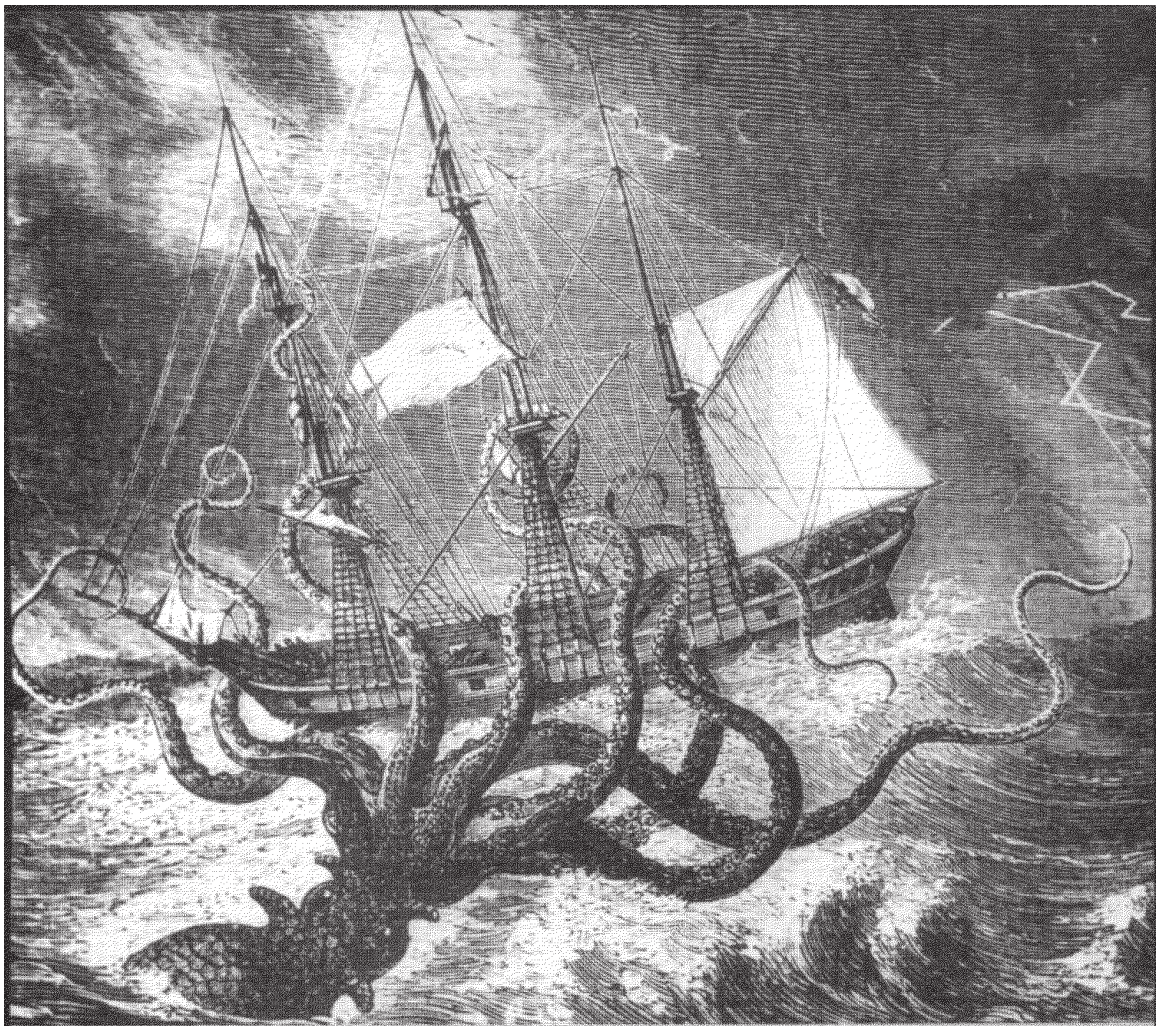

Figure 4.1: The infamous giant squid, having nothing to do with the work of Hodgkin and Huxley on squid giant axon. From *Dangerous Sea Creatures*, © 1976, 1977 Time-Life Films, Inc.

by an action potential. It is important to emphasize that our choice of linear I–V curves for the three different channel types is dictated largely by experimental data. Axons in other species (such as vertebrates) have ionic channels that are better described by other I–V curves, such as the GHK current equation (2.120). However, the qualitative nature of the results remains largely unaffected, and so the discussion in this chapter, which is mostly of a qualitative nature, remains correct for models that use more complex I–V curves to describe the ionic currents.

Equation (4.2) is a first-order ordinary differential equation and can be rewritten in the form

$$C_m \frac{dV}{dt} = -g_{\text{eff}}(V - V_{\text{eq}}) + I_{\text{app}}, \tag{4.3}$$

where $g_{\text{eff}} = g_{\text{Na}} + g_{\text{K}} + g_{\text{L}}$ and $V_{\text{eq}} = (g_{\text{Na}}V_{\text{Na}} + g_{\text{K}}V_{\text{K}} + g_{\text{L}}V_{\text{L}})/g_{\text{eff}}$. V_{eq} is the membrane resting potential and is a balance between the reversal potentials for the three ionic currents. In fact, at rest, the sodium and leakage conductances are small compared to the potassium conductance, so that the resting potential is close to the potassium equilibrium potential.

The quantity $R_m = 1/g_{\text{eff}}$, the passive membrane resistance, is on the order of 1000 Ωcm^2. The time constant for this equation is

$$\tau_m = C_m R_m, \tag{4.4}$$

on the order of 1 msec. It follows that with a steady applied current, the membrane potential should equilibrate quickly to

$$V = V_{\text{eq}} + R_m I_{\text{app}}. \tag{4.5}$$

For sufficiently small applied currents, this is indeed what happens. However, for larger applied currents, the response is quite different. Assuming that the model (4.2) is correct, the only possible explanation for these differences is that the conductances are not constant but depend in some way on the voltage. Historically, the key step to determining the conductances was being able to measure the individual ionic currents and from this to deduce the changes in conductances. This was brilliantly accomplished by Hodgkin and Huxley in 1952.

4.1.1 History of the Hodgkin–Huxley Equations

(This section is adapted from Rinzel, 1990.) In a series of five articles that appeared in the *Journal of Physiology* in 1952, Alan Lloyd Hodgkin and Andrew Fielding Huxley, along with Bernard Katz, who was a coauthor of the lead paper and a collaborator in several related studies, unraveled the dynamic ionic conductances that generate the nerve action potential (Hodgkin et al., 1952; Hodgkin and Huxley, 1952a,b,c,d). They were awarded the 1963 Nobel Prize in physiology and medicine (shared with John C. Eccles, for his work on potentials and conductances at motorneuron synapses).

Before about 1939, the membrane potential was believed to play an important role in the membrane's state, but there was no way to measure it. It was known that a cell's membrane separated different ionic concentrations inside and outside the cell. Applying the Nernst equation, Bernstein (1902) was led to suggest that the

resting membrane was semipermeable to potassium, implying that at rest, V should be around -70 mV. He believed that during activity there was a breakdown in the membrane's resistance to all ionic fluxes, and potential differences would disappear, i.e., V would approach zero.

In 1940, Cole and Curtis, using careful electrode placement coupled with biophysical and mathematical analysis, obtained the first convincing evidence for a substantial transient increase in membrane conductivity during passage of the action potential. While they estimated a large conductance increase, it was not infinite, so without a direct measurement of membrane potential it was not possible to confirm or nullify Bernstein's hypothesis. During a postdoctoral year in the U.S. in 1937–1938, Hodgkin established connections with Cole's group at Columbia and worked with them at Woods Hole in the summer. He and Curtis nearly succeeded in measuring V directly by tunneling along the giant axon with a glass micropipette. When each succeeded later (separately, with other collaborators), they found, surprisingly, that V rose transiently toward zero, but with a substantial overshoot. This finding brought into serious question the hypothesis of Bernstein and provided much food for thought during World War II, when Hodgkin, Huxley, and many other scientists were involved in the war effort.

By the time postwar experimental work was resuming in England, Cole and Marmont had developed the *space clamp technique*. This method allowed one to measure directly the total transmembrane current, uniform through a known area, rather than spatially nonuniform as generated by a capillary electrode. To achieve current control with space clamping, the axon was threaded with a metallic conductor (like a thin silver wire) to provide low axial resistance and thereby eliminate voltage gradients along the length of the axon. Under these conditions the membrane potential is no longer a function of distance along the axon, only of time. In addition, during the 1947 squid season, Cole and company made substantial progress toward controlling the membrane potential as well.

In 1948, Hodgkin went to visit Cole (then at Chicago) to learn directly of their methods. With some further developments of their own, Hodgkin, Huxley, and Katz applied the techniques with great success to record transient ionic fluxes over the physiological ranges of voltages. Working diligently, they collected most of the data for their papers in the summer of 1949. Next came the step of identifying the individual contributions of the different ion species. Explicit evidence that both sodium and potassium were important came from the work of Hodgkin and Katz (1949). This also explained the earlier puzzling observations that V overshoots zero during an action potential, opposing the suggestion of Bernstein. Instead of supposing that there was a transient increase in permeability identical for all ions, Hodgkin and Katz realized that different changes in permeabilities for different ions could account for the V time course, as V would approach the Nernst potential for the ion to which the membrane was predominantly permeable, and this dominance could change with time. For example, at rest the membrane is most permeable to K^+, so that V is close to V_K. However, if g_K were to decrease and g_{Na} were to increase, then V would be pushed toward V_{Na}, which is positive, thus depolarizing the cell.

The question of how the changes in permeability were dynamically linked to V

was not completely stated until the papers of 1952. In fact, the substantial delay from data collection in 1949 until final publication in 1952 can be attributed to the considerable time devoted to data analysis, model formulation, and testing. Computer downtime was also a factor, as some of the solutions of the Hodgkin–Huxley equations were computed on a desktop, hand-cranked calculator. As Hodgkin notes, "The propagated action potential took about three weeks to complete and must have been an enormous labour for Andrew [Huxley]" (Hodgkin, 1976, p. 19).

The final paper of the 1952 series is a masterpiece of the scientific art. Therein they present their elegant experimental data, a comprehensive theoretical hypothesis, a fit of the model to the experimental data (obtained for fixed values of the membrane potential), and then, presto, a prediction (from their numerical computations) of the time course of the propagated action potential. In biology, where quantitatively predictive theories are rare, this work stands out as one of the most successful combinations of experiment and theory.

4.1.2 Voltage and Time Dependence of Conductances

The key step to sorting out the dynamics of the conductances came from the development of the *voltage clamp*. A voltage clamp fixes the membrane potential, usually by a rapid step from one voltage to another, and then measures the current that must be supplied in order to hold the voltage constant. Since the supplied current must equal the transmembrane current, the voltage clamp provides a way to measure the transient transmembrane current that results. The crucial point is that the voltage can be stepped from one constant level to another, and so the ionic currents can be measured at a constant, known, voltage. Thus, even when the conductances are functions of the voltage (as is actually the case), a voltage clamp eliminates any voltage changes and permits measurement of the conductances as functions of time only.

Hodgkin and Huxley found that when the voltage was stepped up and held fixed at a higher level, the total ionic current was initially inward, but at later times an outward current developed (Fig. 4.2). For a number of reasons, not discussed here, they argued that the initial inward current is carried almost entirely by Na^+ ions, while the outward current that develops later is carried largely by K^+ ions. With these assumptions, Hodgkin and Huxley were able to use a clever trick to separate the total ionic current into its constituent ionic parts. They replaced 90% of the extracellular sodium in the normal seawater bath with choline (a viscous liquid vitamin B complex found in many animal and vegetable tissues), which rendered the axon nonexcitable but changed the resting potential only slightly. Since it is assumed that immediately after the voltage has been stepped up, the ionic current is all carried by Na^+, it is possible to measure the initial Na^+ currents in response to a voltage step. Note that although the Na^+ currents can be measured directly immediately after the voltage step, they cannot be measured directly over a longer time period, as the total ionic current begins to include a contribution from the K^+ current. If we denote the Na^+ currents for the two cases of normal extracellular Na^+ and zero extracellular Na^+ by I^1_{Na} and I^2_{Na} respectively, then the ratio of the two currents,

$$I^1_{Na}/I^2_{Na} = K, \tag{4.6}$$

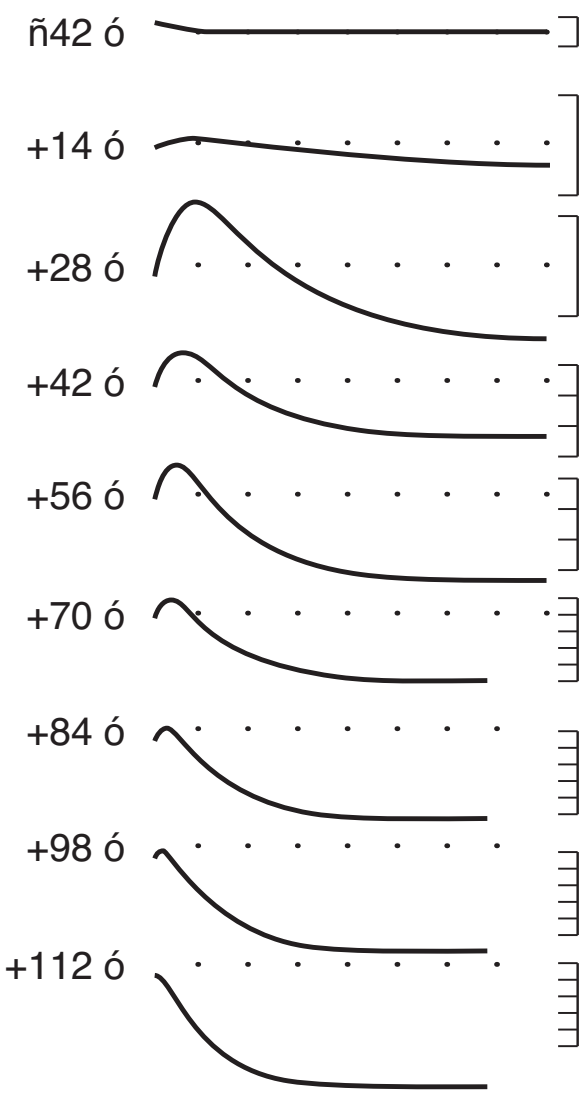

Figure 4.2: Experimental results describing the total membrane current in response to a step depolarization. The numbers on the left give the final value of the membrane potential, in mV. The interval between dots on the horizontal scale is 1 ms, while one division on the vertical scale represents 0.5 mA/cm^2. (Hodgkin and Huxley, 1952a, Fig. 2a.)

say, can be measured directly from the experimental data.

Next, Hodgkin and Huxley made two further assumptions. First, they assumed that the sodium current ratio K is independent of time and is thus constant over the course of each voltage clamp experiment. In other words, the amplitude and direction of the Na^+ current may be affected by the low extracellular Na^+ solution, but its time course is not. Second, they assumed that the potassium channels are unaffected by the change in extracellular sodium concentration. There is considerable evidence that the sodium and potassium channels are independent. Tetrodotoxin (TTX) is known to block sodium currents while leaving the potassium currents almost unaffected, while tetraethylammonium (TEA) has the opposite effect of blocking the potassium current but not the sodium current. To complete the argument, since $I_{\rm ion} = I_{\rm Na} + I_{\rm K}$, and $I^1_{\rm K} = I^2_{\rm K}$, it follows that $I^1_{\rm ion} - I^1_{\rm Na} = I^2_{\rm ion} - I^2_{\rm Na}$, and thus

$$I^1_{\rm Na} = \frac{K}{K-1}(I^1_{\rm ion} - I^2_{\rm ion}), \tag{4.7}$$

$$I_{\rm K} = \frac{I^1_{\rm ion} - KI^2_{\rm ion}}{1-K}. \tag{4.8}$$

Hence, given measurements of the total ionic currents in the two cases, and given the ratio K of the Na^+ currents, it is possible to determine the complete time courses of both the Na^+ and K^+ currents.

Finally, from knowledge of the individual currents, one obtains the conductances as

$$g_{\rm Na} = \frac{I_{\rm Na}}{V - V_{\rm Na}}, \qquad g_{\rm K} = \frac{I_{\rm K}}{V - V_{\rm K}}. \tag{4.9}$$

Note that this result relies on the specific (linear) model used to describe the I–V curve of the Na^+ and K^+ channels, but, as we discussed above, we assume throughout that the instantaneous I–V curves of the Na^+ and K^+ channels are linear.

Samples of Hodgkin and Huxley's data are shown in Fig. 4.3. The plots show ionic conductances as functions of time following a step increase or decrease in the membrane potential. The important observation is that with voltages fixed, the conductances are time dependent. For example, when V is stepped up and held fixed at a higher level, $g_{\rm K}$ does not increase instantaneously, but instead increases over time to a final steady level. Both the time constant of the increase and the final value of $g_{\rm K}$ are dependent on the value to which the voltage is stepped. Further, $g_{\rm K}$ increases in a sigmoidal fashion, with a slope that first increases and then decreases (Fig. 4.3A and B). Following a step decrease in the voltage, $g_{\rm K}$ falls in a simple exponential fashion (Fig. 4.3A). This particular feature of $g_{\rm K}$ — a sigmoidal increase coupled with an exponential decrease — will be important when we model $g_{\rm K}$. The behavior of $g_{\rm Na}$ is more complex. Following a step increase in voltage, $g_{\rm Na}$ first increases, but then decreases again, *all at the same fixed voltage* (Fig. 4.3C). Hence, the time dependence of $g_{\rm Na}$ requires a more complex model than for that of $g_{\rm K}$.

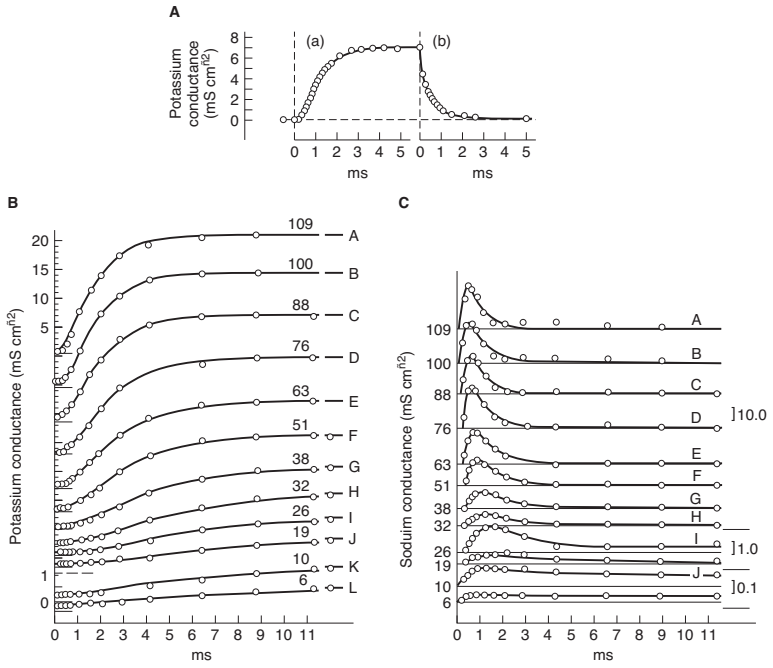

Figure 4.3: Conductance changes as a function of time at different voltage clamps. A: The response of g_{K} to a step increase in V and then a step decrease. B: Responses of g_{K} to step increases in V of varying magnitudes. The number on each curve gives the depolarization in mV, and the smooth curves are calculated from solution of (4.11) and (4.12), with the initial condition $g_{\mathrm{K}}(t = 0) = 0.24$ mS/cm^2. The vertical scale is the same in curves A–J, but is increased by a factor of four in the lower two curves. For clarity, the baseline of each curve has been shifted up. C: Responses of g_{Na} to step increases in V of magnitudes given by the numbers on the left, in mV. The smooth curves are the model solutions. The vertical scales on the right are in units of mS/cm^2. (Hodgkin and Huxley, 1952d, Figs. 2, 3, and 6.)

The potassium conductance

From the experimental data shown in Fig. 4.3A and B, it is reasonable to expect that $g_{\rm K}$ obeys some differential equation,

$$\frac{dg_{\rm K}}{dt} = f(v,t), \tag{4.10}$$

say, where $v = V - V_{\rm eq}$; i.e., v is the difference between the membrane potential and the resting potential. (Of course, since $V_{\rm eq}$ is a constant, $dv/dt = dV/dt$.) However, for $g_{\rm K}$ to have the required sigmoidal increase and exponential decrease, Hodgkin and Huxley realized that it would be easier to write $g_{\rm K}$ as some power of a different variable, n say, where n satisfies a first-order differential equation. Thus, they wrote

$$g_{\rm K} = \bar{g}_{\rm K} n^4, \tag{4.11}$$

for some constant $\bar{g}_{\rm K}$. The fourth power was chosen not for physiological reasons, but because it was the smallest exponent that gave acceptable agreement with the experimental data. The secondary variable n obeys the differential equation

$$\tau_n(v)\frac{dn}{dt} = n_\infty(v) - n, \tag{4.12}$$

for some functions $\tau_n(v)$ and $n_\infty(v)$ that must be determined from the experimental data in a manner that we describe soon. Equation (4.12) is often written in the form

$$\frac{dn}{dt} = \alpha_n(v)(1-n) - \beta_n(v)n, \tag{4.13}$$

where

$$n_\infty(v) = \frac{\alpha_n(v)}{\alpha_n(v)+\beta_n(v)}, \tag{4.14}$$

$$\tau_n(v) = \frac{1}{\alpha_n(v)+\beta_n(v)}. \tag{4.15}$$

At elevated potentials $n(t)$ increases monotonically and exponentially toward its resting value, thereby turning on, or *activating*, the potassium current. Since the Nernst potential is below the resting potential, the potassium current is an outward current at potentials greater than rest. The function $n(t)$ is called the *potassium activation*.

It is instructive to consider in detail how such a formulation for $g_{\rm K}$ results in the required sigmoidal increase and exponential decrease. Suppose that at time $t = 0$, v is increased from 0 to v_0 and then held constant, and suppose further that $n(0) = 0$. Solving (4.12) then gives

$$n(t) = n_\infty(v_0)\left[1 - \exp\left(\frac{-t}{\tau_n(v_0)}\right)\right], \tag{4.16}$$

which is an increasing curve (with monotonically decreasing slope) that approaches its maximum at $n_\infty(v_0)$. Raising n to the fourth power gives a sigmoidally increasing

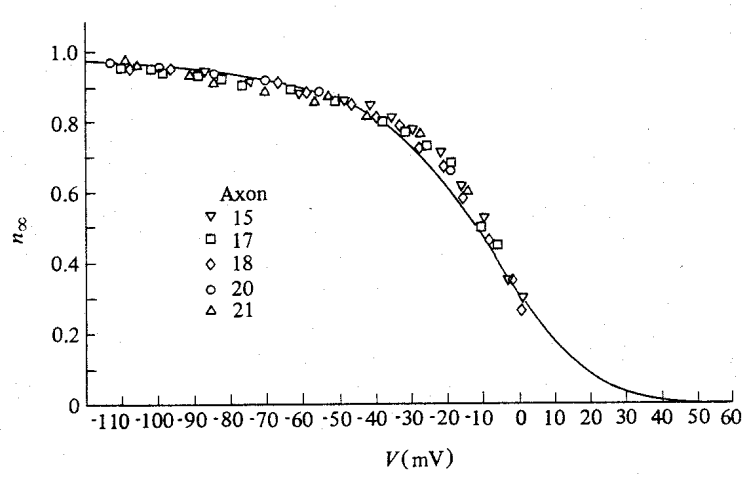

Figure 4.4: Data points (symbols) of n_∞, determined by fitting (4.16) to the experimental time courses. The smooth curve through the symbols provides a continuous description of n_∞, and its functional form has no physiological significance. (Hodgkin and Huxley, 1952d, Fig. 5.)

curve as required. Higher powers of n result in curves with a greater maximum slope at the point of inflection. In response to a step decrease in v, from v_0 to 0 say, the solution for n is

$$n(t) = n_\infty(v_0) \exp\left(\frac{-t}{\tau_n(v_0)}\right), \tag{4.17}$$

in which case n^4 is exponentially decreasing, with no inflection point.

It remains to describe how the functions n_∞ and τ_n are determined from the experimental data. For any given voltage step, the time constant τ_n, and the final value of n, namely n_∞, can be determined by fitting (4.16) to the experimental data. By this procedure one can determine τ_n and n_∞ at a discrete set of values for v, i.e., those values used experimentally. Typical data points for n_∞ are shown in Fig. 4.4 as symbols. To obtain a complete description of $g_{\rm K}$, valid for all voltages and not only those used in the experiments, Hodgkin and Huxley fitted a smooth curve through the data points. The functional form of the smooth curve has no physiological significance, but is a convenient way of providing a continuous description of n_∞. A similar procedure is followed for τ_n. The continuous descriptions of n_∞ and τ_n (expressed in terms of α_n and β_n) are given in (4.28) and (4.29) below.

The sodium conductance

The time dependence for the sodium conductance is more difficult to unravel. From the experimental data it is suggested that there are two processes at work, one that turns on the sodium current and one that turns it off. Hodgkin and Huxley

proposed that the sodium conductance is of the form

$$g_{\mathrm{Na}}(v) = \bar{g}_{\mathrm{Na}} m^3 h, \tag{4.18}$$

and they fit the time-dependent behavior of m and h to exponentials with dynamics

$$\frac{dw}{dt} = \alpha_w(1-w) - \beta_w w, \tag{4.19}$$

where $w = m$ or h. Because m is small at rest and first increases, it is called the *sodium activation*, and because h shuts down, or inactivates, the sodium current, it is called the *sodium inactivation*. When $h = 0$, the sodium current is completely inactivated. The overall procedure is similar to that used in the specification of g_{K}. For any fixed voltage step, the unknown functions α_w and β_w are determined by fitting to the experimental curves (Fig. 4.3C), and then smooth curves, with arbitrary functional forms, are fitted through the data points for α_w and β_w.

Summary of the equations

In summary, the Hodgkin–Huxley equations for the space clamped axon are

$$\begin{aligned}
C_m \frac{dv}{dt} &= -\bar{g}_{\mathrm{K}} n^4 (v - v_{\mathrm{K}}) - \bar{g}_{\mathrm{Na}} m^3 h (v - v_{\mathrm{Na}}) - \bar{g}_{\mathrm{L}} (v - v_{\mathrm{L}}) + I_{\mathrm{app}}, && (4.20)\\
\frac{dm}{dt} &= \alpha_m(1-m) - \beta_m m, && (4.21)\\
\frac{dn}{dt} &= \alpha_n(1-n) - \beta_n n, && (4.22)\\
\frac{dh}{dt} &= \alpha_h(1-h) - \beta_h h. && (4.23)
\end{aligned}$$

The specific functions α and β proposed by Hodgkin and Huxley were, in units of $(\mathrm{ms})^{-1}$,

$$\begin{aligned}
\alpha_m &= 0.1 \frac{25 - v}{\exp\left(\frac{25-v}{10}\right) - 1}, && (4.24)\\
\beta_m &= 4 \exp\left(\frac{-v}{18}\right), && (4.25)\\
\alpha_h &= 0.07 \exp\left(\frac{-v}{20}\right), && (4.26)\\
\beta_h &= \frac{1}{\exp(\frac{30-v}{10}) + 1}, && (4.27)\\
\alpha_n &= 0.01 \frac{10 - v}{\exp(\frac{10-v}{10}) - 1}, && (4.28)\\
\beta_n &= 0.125 \exp\left(\frac{-v}{80}\right). && (4.29)
\end{aligned}$$

For these expressions, the potential v is the deviation from rest ($V = V_{\mathrm{eq}} + v$), measured in units of mV, current density is in units of $\mu\mathrm{A/cm}^2$, conductances are

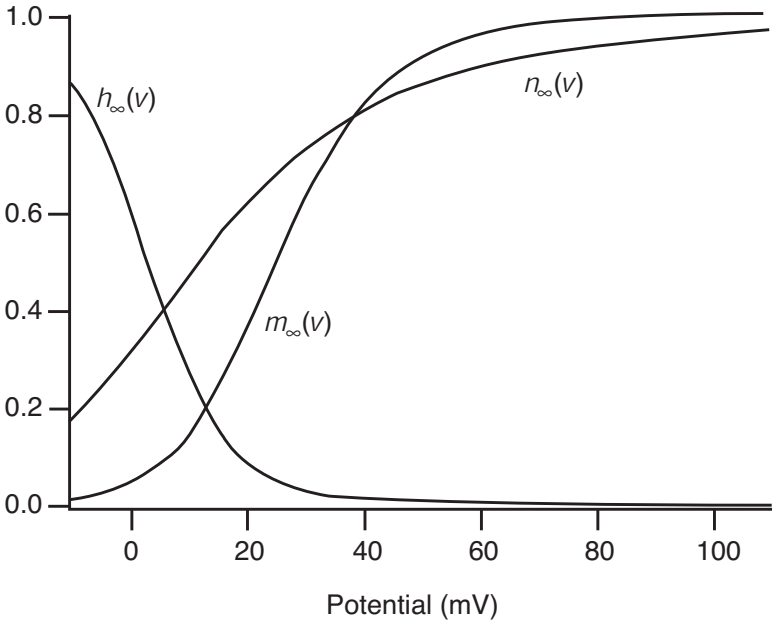

Figure 4.5: Steady-state functions $m_\infty(v)$, $n_\infty(v)$ and $h_\infty(v)$.

in units of mS/cm^2, and capacitance is in units of μF/cm^2. The remaining constants are

$$\bar{g}_{\rm Na} = 120, \qquad \bar{g}_{\rm K} = 36, \qquad \bar{g}_{\rm L} = 0.3, \tag{4.30}$$

with (adjusted) equilibrium potentials $v_{\rm Na} = 115, v_{\rm K} = -12$, and $v_{\rm L} = 10.6$. In Fig. 4.5 are shown the steady-state functions, and the time constants are shown in Fig. 4.6.

In Chapter 3 we discussed simple models for the gating of Na^+ and K^+ channels and showed how the rate constants in simple kinetic schemes could be determined from whole-cell or single-channel data. We also showed how models of the form (4.20)–(4.23) can be derived by modeling the ionic channels as consisting of multiple subunits, each of which obeys a simple two-state model. In the Hodgkin–Huxley equations, it is assumed that the Na^+ channel consists of three "m" gates and one "h" gate, each of which can be either closed or open. If the gates operate independently, then the fraction of open Na^+ channels is m^3h, where m and h obey the equation of the two-state channel model. Similarly, if there are four "n" gates per potassium channel, all of which must be open for potassium to flow, then the fraction of open K^+ channels is n^4.

Now comes the most interesting challenge facing these equations. Having incorporated the measurements of conductance found from voltage-clamp experiments, one wonders whether these equations reproduce a realistic action potential, and if so, by what mechanism is the action potential produced? We can describe in qualitative terms how the Hodgkin–Huxley equations should work. If small currents are applied to a cell for a short period of time, the potential returns rapidly to its

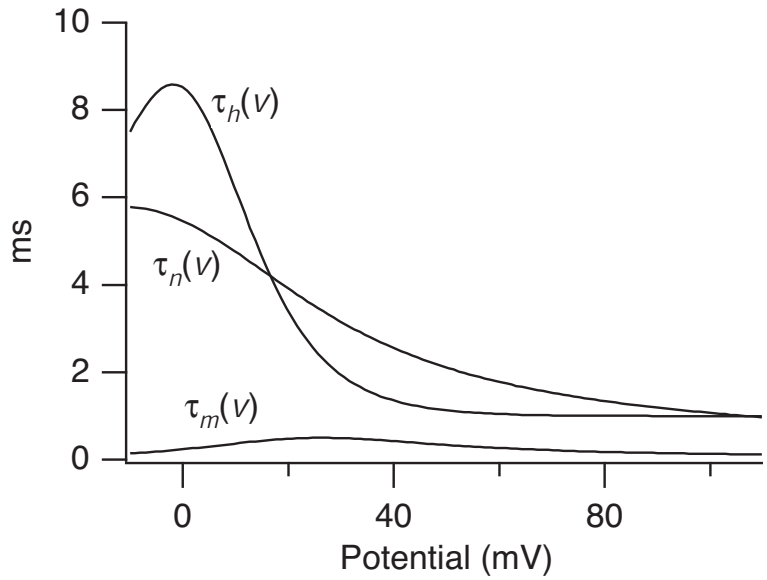

Figure 4.6: Time constants $\tau_m(v)$, $\tau_n(v)$, and $\tau_h(v)$.

equilibrium $v = 0$ after the applied current is removed. The equilibrium potential is close to the potassium Nernst potential $v_{\rm K} = -12$, because at rest, the sodium and leakage conductances are small. There is always competition among the three ionic currents to drive the potential to the corresponding resting potential. For example, if the potassium and leakage currents could be blocked or the sodium conductance dramatically increased, then the term $g_{\rm Na}(V - V_{\rm Na})$ should dominate (4.2), and as long as v is below $v_{\rm Na}$, an inward sodium current will drive the potential toward $v_{\rm Na}$. Similarly, while v is above $v_{\rm K}$, the potassium current is outward in an attempt to drive v toward $v_{\rm K}$. Notice that since $v_{\rm K} < v_{\rm L} < v_{\rm Na}$, v is necessarily restricted to lie in the range $v_{\rm K} < v < v_{\rm Na}$.

If $g_{\rm Na}$ and $g_{\rm K}$ were constant, that would be the end of the story. The equilibrium at $v = 0$ would be a stable equilibrium, and following any stimulus, the potential would return exponentially to rest. But since $g_{\rm Na}$ and $g_{\rm K}$ can change, the different currents can exert their respective influences. The actual sequence of events is determined by the dynamics of m, n, and h. The most important observation for the moment is that $\tau_m(v)$ is much smaller than either $\tau_n(v)$ or $\tau_h(v)$, so that $m(t)$ responds much more quickly to changes in v than either n or h. We can now understand why the Hodgkin–Huxley system is an excitable system. As noted before, if the potential v is raised slightly by a small stimulating current, the system returns to its stable equilibrium. However, during the period of time that the potential v is elevated, the sodium activation m is tracking $m_\infty(v)$. If the stimulating current is large enough to raise the potential and therefore $m_\infty(v)$ to a high enough level (above its *threshold*), then before the system can return to rest, m will increase sufficiently to change the sign of the net current, resulting in an autocatalytic inward

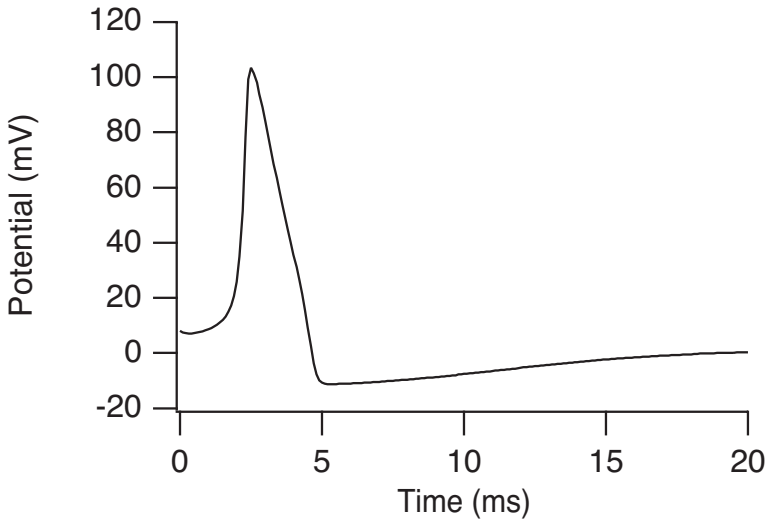

Figure 4.7: Action potential of the Hodgkin–Huxley equations.

sodium current. Now, as the potential rises, m continues to rise, and the inward sodium current is increased, further adding to the rise of the potential.

If nothing further were to happen, the potential would be driven to a new equilibrium at v_{Na}. However, here is where the difference in time constants plays an important role. When the potential is at rest, the sodium inactivation h is positive, about 0.6. As the potential increases, h_∞ decreases toward zero, and as h approaches zero, the sodium current is inactivated because g_{Na} approaches zero. However, because the time constant $\tau_h(v)$ is much larger than $\tau_m(v)$, there is a considerable delay between turning on the sodium current when m increases and turning off the sodium current when h decreases. The net effect of the two different time scales on m and h is that the sodium current is at first turned on and later turned off, and this is seen as an initial increase of the potential, followed by a decrease toward rest.

At about the same time that the sodium current is inactivated, the outward potassium current is activated. This is because of the similarity of the time constants $\tau_n(v)$ and $\tau_h(v)$. Activation of the potassium current drives the potential below rest toward v_{K}. When v is negative, n declines, and the potential eventually returns to rest, and the whole process can start again. In Fig. 4.7 is shown a plot of the potential $v(t)$ during an action potential following a superthreshold stimulus. In Fig. 4.8, $m(t)$, $n(t)$, and $h(t)$ during the same action potential are shown.

There are four recognizable phases of an action potential: the *upstroke, excited*, *refractory*, and *recovery* phases. The refractory period is the period following the excited phase when additional stimuli evoke no substantial response, even though the potential is below or close to its resting value. There can be no response, since the sodium channels are still inactivated because h is small. As h gradually returns to its resting value, further responses once again become possible.

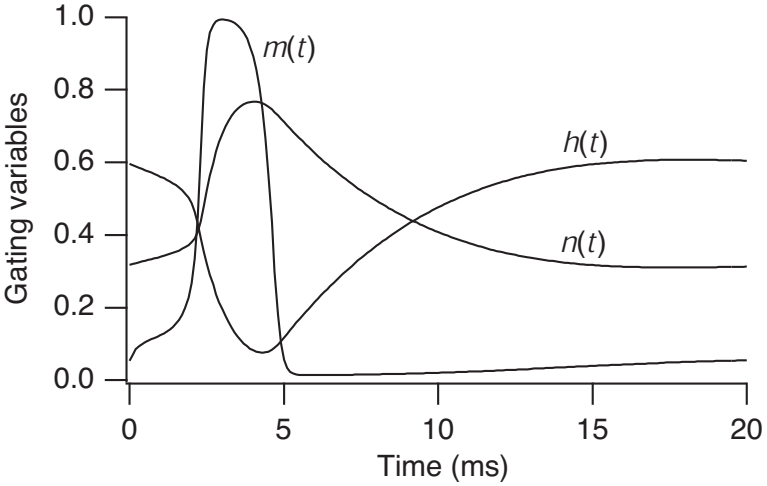

Figure 4.8: The gating variables m, n, and h during an action potential.

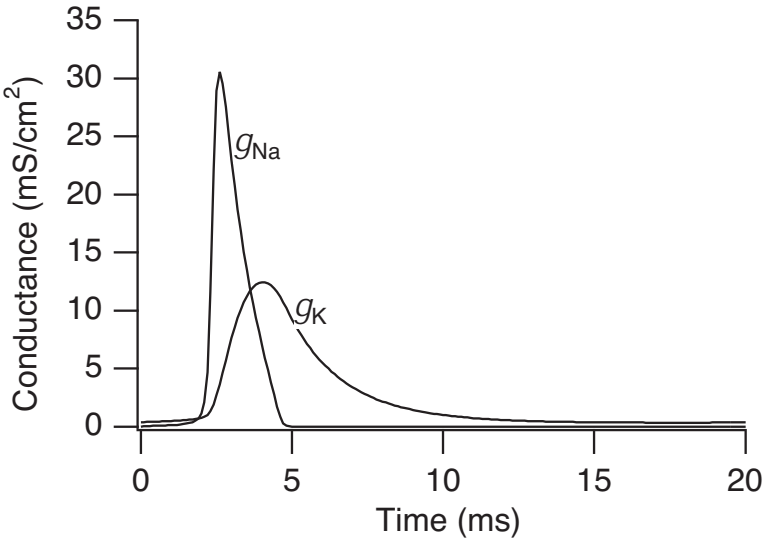

Figure 4.9: Conductances g_{Na} and g_{K} during an action potential.

There are two ways that the Hodgkin–Huxley system can be made into an autonomous oscillator. The first is to inject a steady current of sufficient strength. Such a current raises the resting potential above the threshold for an action potential, so that after the axon has recovered from an action potential, the potential rises to a superthreshold level at which another action potential is evoked.

Immersing the axon in a bath of high extracellular potassium has the same effect through a slightly different mechanism. An increase of extracellular potassium has the effect of increasing the potassium Nernst potential, effectively raising the rest potential (since the rest potential is close to the potassium Nernst potential). If this increase of the potassium Nernst potential is sufficiently large, the resting potential becomes superthreshold, and autonomous oscillations result. This mechanism of creating an autonomous oscillator out of normally excitable but nonoscillatory cells is important for certain cardiac arrhythmias.

4.1.3 Qualitative Analysis

FitzHugh (1960, 1961, 1969) has given a particularly elegant qualitative description of the Hodgkin–Huxley equations that allows a better understanding of the model's behavior. More detailed analyses have also been given by Rinzel (1978), Troy (1978), Cole et al. (1955), and Sabah and Spangler (1970). FitzHugh's approach is based on the fact that some of the model variables have fast kinetics, while others are much slower. In particular, m and v are fast variables (i.e., the Na^+ channel activates quickly, and the membrane potential changes quickly), while n and h are slow variables (i.e., Na^+ channels are inactivated slowly, and the K^+ channels are activated slowly). Thus, during the initial stages of the action potential, n and h remain essentially constant while m and v vary. This allows the full 4-dimensional phase space to be broken into smaller pieces by fixing the slow variables and considering the behavior of the model as a function only of the two fast variables. Although this description is accurate only for the initial stages of the action potential, it provides a useful way to study the process of excitation.

The fast phase-plane

Thus motivated, we fix the slow variables n and h at their respective resting states, which we call n_0 and h_0, and consider how m and v behave in response to stimulation. The differential equations for the fast phase-plane are

$$C_m \frac{dv}{dt} = -\bar{g}_K n_0^4 (v - v_K) - \bar{g}_{Na} m^3 h_0 (v - v_{Na}) - \bar{g}_L (v - v_L), \tag{4.31}$$

$$\frac{dm}{dt} = \alpha_m (1 - m) - \beta_m m, \tag{4.32}$$

or, equivalently,

$$\tau_m \frac{dm}{dt} = m_\infty - m. \tag{4.33}$$

This is now a two-dimensional system and can be most easily studied in the (m, v) phase-plane, a plot of which is given in Fig. 4.10. The curves defined by $dv/dt = 0$ and $dm/dt = 0$ are the v and m nullclines, respectively. The m nullcline is the

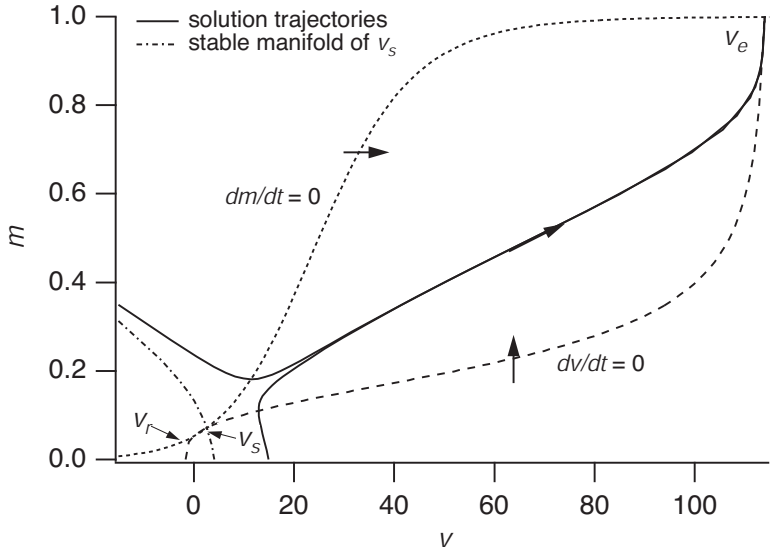

Figure 4.10: The Hodgkin–Huxley fast phase-plane, showing the nullclines $dv/dt = 0$ and $dm/dt = 0$ (with $h_0 = 0.596$, $n_0 = 0.3176$), two sample trajectories and the stable manifold of the saddle point v_s.

curve $m = m_\infty(v)$, which we have seen before (in Fig. 4.5), while the v nullcline is defined by

$$v = \frac{\bar{g}_{\mathrm{Na}} m^3 h_0 v_{\mathrm{Na}} + \bar{g}_{\mathrm{K}} n_0^4 v_{\mathrm{K}} + \bar{g}_{\mathrm{L}} v_{\mathrm{L}}}{\bar{g}_{\mathrm{Na}} m^3 h_0 + \bar{g}_{\mathrm{K}} n_0^4 + \bar{g}_{\mathrm{L}}}. \tag{4.34}$$

For the parameters of the Hodgkin–Huxley model, the m and v nullclines intersect in three places, corresponding to three steady states of the fast equations. Note that these three intersections are not steady states of the full model, only of the fast subsystem, and, to be precise, should be called pseudo-steady states. However, in the context of the fast phase-plane we continue to call them steady states. We label the three steady states v_r, v_s, and v_e (for resting, saddle, and excited).

It is left as an exercise to show that v_r and v_e are stable steady states of the fast subsystem, while v_s is a saddle point. Since v_s is a saddle point, it has a one-dimensional stable manifold. This stable manifold divides the (m, v) plane into two regions: any trajectory starting to the left of the stable manifold is prevented from reaching v_e and must eventually return to the resting state, v_r. However, any trajectory starting to the right of the stable manifold is prevented from returning to the resting state and must eventually end up at the excited state, v_e. Hence, the stable manifold, in combination with the two stable steady states, causes a threshold phenomenon. Any perturbation from the resting state that is not large enough to cross the stable manifold eventually dies away, but a perturbation that crosses the stable manifold results in a large excursion in the voltage, up to the excited state. Sample trajectories are sketched in Fig. 4.10.

If m and v were the only variables in the model, then v would stay at v_e indef-

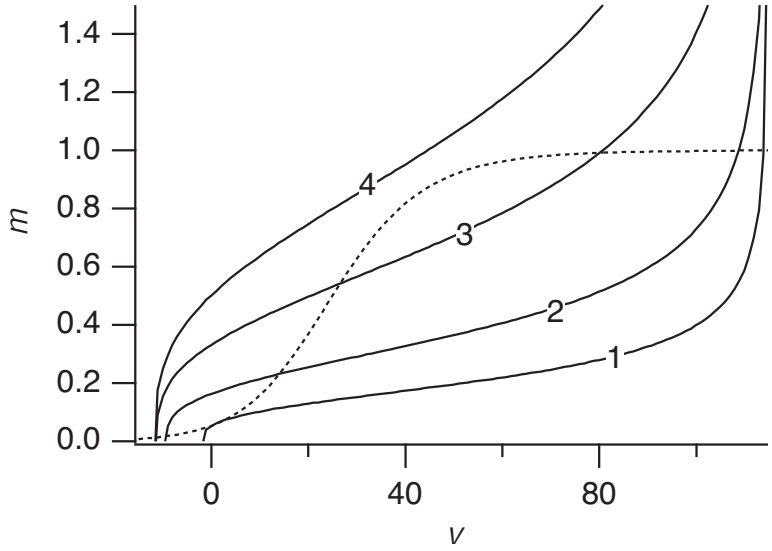

Figure 4.11: The Hodgkin–Huxley fast phase-plane as a function of the slow variables, showing the m nullcline (dashed) and the movement of the v nullcline (solid) and the disappearance of the steady states. For these curves, parameter values are (1) $h_0 = 0.596$, $n_0 = 0.3176$; (2) $h_0 = 0.4$, $n_0 = 0.5$; (3) $h_0 = 0.2$, $n_0 = 0.7$; and (4) $h_0 = 0.1$, $n_0 = 0.8$.

initely. However, as pointed out before, v_e is not a steady state of the full model. Thus, to see what happens on a longer time scale, we must consider how slow variations in n and h affect the qualitative nature of the fast phase-plane. First note that since $v_e > v_r$, it follows that $h_\infty(v_e) < h_\infty(v_r)$ and $n_\infty(v_e) > n_\infty(v_r)$. Hence, while v is at the excited state, h begins to decrease thus inactivating the Na^+ conductance, and n starts to increase thus activating the K^+ conductance. Next note that although the m nullcline in the fast phase-plane is independent of n and h, the v nullcline is not. In Fig. 4.10 the nullclines were drawn using the steady-state values for n and h: different values of n and h change the shape of the v nullcline. As n increases and h decreases, the v nullcline moves to the left and up, as illustrated in Fig. 4.11. As the v nullcline moves up and to the left, v_e and v_s move towards each other, while v_r moves to the left. During this phase the voltage is at v_e and thus decreases slowly. Eventually, v_e and v_s coalesce and disappear in a saddle-node bifurcation. When this happens v_r is the only remaining steady state, and so the solution must return to the resting state. Note that since the v nullcline has moved up and to the left, v_r is not a steady state of the full system. However, when v decreases to v_r, n and h both return to their steady states and as they do so, v_r slowly increases until the steady state of the full system is reached and the action potential is complete. A schematic diagram of a complete action potential is shown in Fig. 4.12, and the important points are labeled for comparison with the phase-plane.

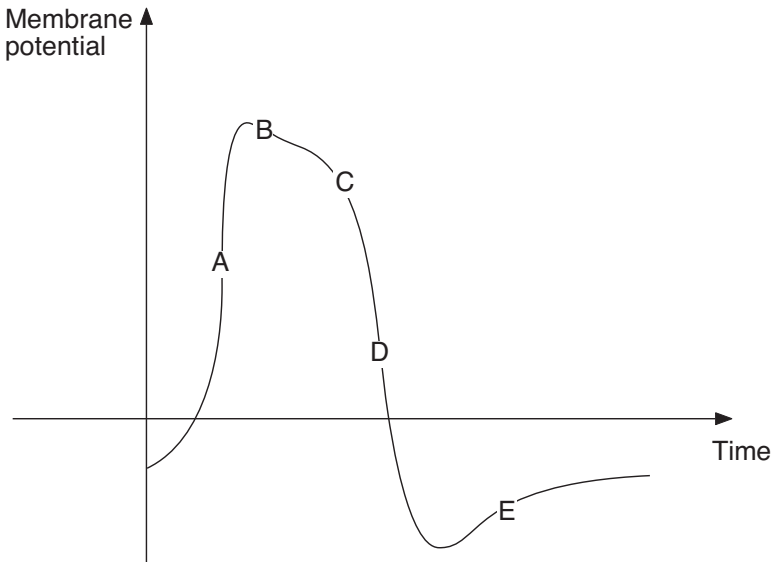

Figure 4.12: Schematic diagram of a complete action potential. A: Superthreshold stimulus causes a fast increase of v to the excited state. B: v is sitting at the excited state, v_e, decreasing slowly as n increases and h decreases, i.e., as v_e moves toward v_s. C: v_e and v_s disappear at a saddle-node bifurcation, and so, D: The solution must return to the resting state v_r. E: n and h slowly return to their resting states, and as they do so, v_r slowly increases until the steady state of the full four-dimensional system is reached.

The fast–slow phase-plane

In the above analysis, we simplified the four-dimensional phase space by taking a series of two-dimensional cross-sections, those with various fixed values of n and h. However, by taking a different cross-section we can highlight other aspects of the action potential. In particular, by taking a cross-section involving one fast variable and one slow variable we obtain a description of the Hodgkin–Huxley model that has proven to be extraordinarily useful.

We extract a single fast variable by assuming that m is an instantaneous function of v, and thus $m = m_\infty(v)$ at all times. This is equivalent to assuming that activation of the Na^+ conductance acts on a time scale even faster than that of the voltage. Next, FitzHugh noticed that during the course of an action potential, $h + n \approx 0.8$ (notice the approximate symmetry of $n(t)$ and $h(t)$ in Fig. 4.8), and thus h can be eliminated by setting $h = 0.8 - n$. With these simplifications, the Hodgkin–Huxley model contains one fast variable v and one slow variable n, and can be written as

$$-C_m \frac{dv}{dt} = \bar{g}_{\rm K} n^4 (v - v_{\rm K}) + \bar{g}_{\rm Na} m_\infty^3(v)(0.8 - n)(v - v_{\rm Na}) + \bar{g}_{\rm L}(v - v_{\rm L}), \quad (4.35)$$

$$\frac{dn}{dt} = \alpha_n(1 - n) - \beta_n n. \quad (4.36)$$

For convenience we let $f(v, n)$ denote the right-hand side of (4.35), i.e.,

$$-f(v, n) = \bar{g}_{\rm K} n^4 (v - v_{\rm K}) + \bar{g}_{\rm Na} m_\infty^3(v)(0.8 - n)(v - v_{\rm Na}) + \bar{g}_{\rm L}(v - v_{\rm L}). \quad (4.37)$$

A plot of the nullclines of the fast–slow subsystem is given in Fig. 4.13. The v nullcline is defined by $f(v, n) = 0$ and has a cubic shape, while the n nullcline is $n_\infty(v)$ and is monotonically increasing. There is a single intersection (at least for the given parameter values) and thus a single steady state. Because v is a fast variable and n is a slow one, the solution trajectories are almost horizontal except where $f(v, n) \approx 0$. The curve $f(v, n) = 0$ is called the "slow manifold." Along the slow manifold the solution moves slowly in the direction determined by the sign of dn/dt, but away from the slow manifold the solution moves quickly in a horizontal direction. From the sign of dv/dt it follows that the solution trajectories move away from the middle branch of the slow manifold and toward the left and right branches. Thus, the middle branch is termed the unstable branch of the slow manifold. This unstable branch acts as a threshold. Suppose a perturbation from the steady state is small enough so that v does not cross the unstable manifold. Then, the trajectory moves horizontally towards the left and returns to the steady state. However, if the perturbation is large enough so that v crosses the unstable manifold, then the trajectory moves to the right until it reaches the right branch of the slow manifold, which corresponds to the excited state. On this right branch $dn/dt > 0$, and so the solution moves slowly up the slow manifold until the turning point is reached. At the turning point, n cannot increase any further, as the right branch of the slow manifold ceases to exist, and so the solution moves over to the left branch of the slow manifold. On this left branch $dn/dt < 0$, and so the solution moves down the left branch until the steady state is reached, completing the action

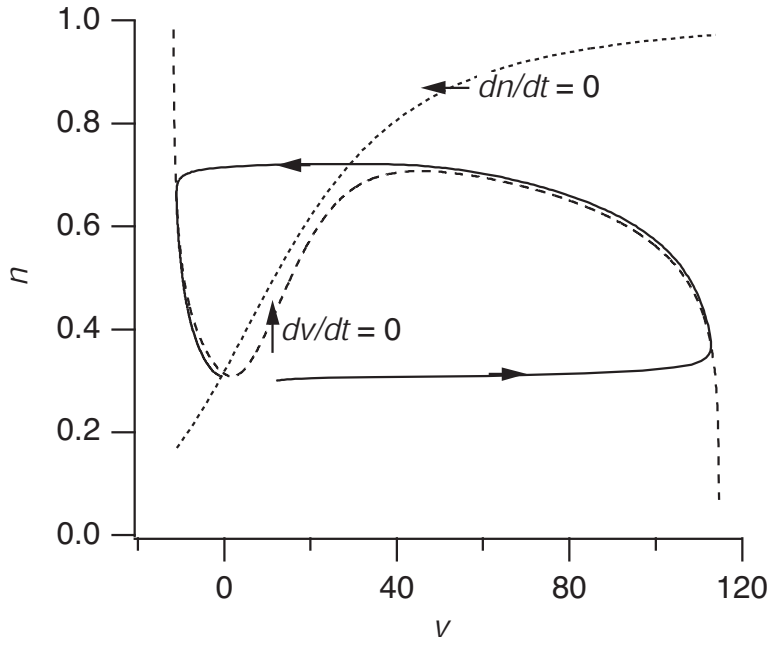

Figure 4.13: Fast–slow phase-plane of the Hodgkin–Huxley model.

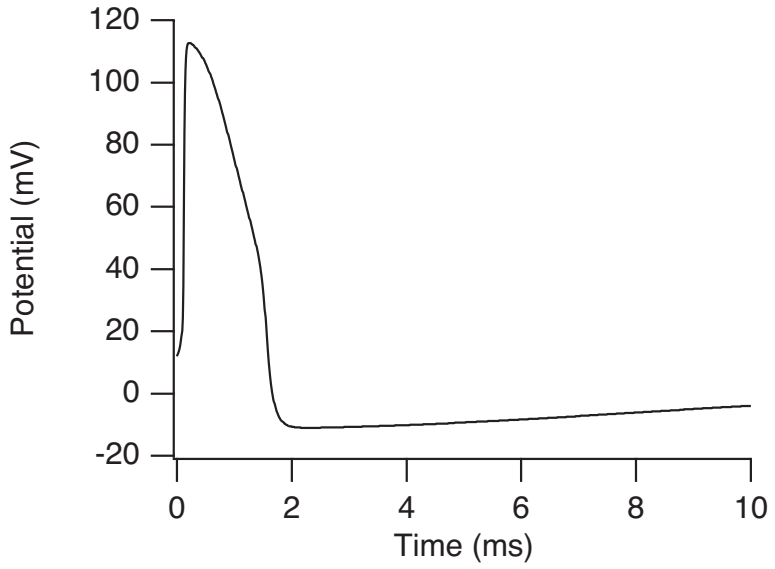

Figure 4.14: Action potential for thereduced fast–slow Hodgkin–Huxleymodel.

potential (Fig. 4.13). A plot of the potential as a function of time is shown in Fig. 4.14.

The variables v and n are usually called the excitation and recovery variables, respectively: excitation because it governs the rise to the excited state, and recovery because it causes the return to the steady state. In the absence of n the solution would stay at the excited state indefinitely.

There is a close relationship between the fast phase-plane and the fast–slow phase-plane. Recall that in the fast phase-plane, the v and m nullclines have three intersection points when $n = n_0$ and $h = h_0$. These three intersections correspond to the three branches of the curve $f(v, n_0) = 0$. In other words, when n is fixed at n_0, the equation $f(v, n_0) = 0$ has three possible solutions, corresponding to v_r, v_s and v_e in the fast phase-plane. However, consideration of Fig. 4.13 shows that, as n increases, the two rightmost branches of the slow manifold coalesce and disappear. This is analogous to the merging and disappearance of v_e and v_s seen in the fast phase-plane. The fast–slow phase-plane is a convenient way of summarizing how v_r, v_s, and v_e depend on the slow variables.

This representation of the Hodgkin–Huxley model in terms of two variables, one fast and one slow, is the basis of the FitzHugh–Nagumo model for excitability, and we discuss models of this generic type in some detail throughout this book.

4.2 Two-Variable Models

There is considerable value in studying systems of equations that are simpler than the Hodgkin–Huxley equations but that retain many of their qualitative features. This is the motivation for the FitzHugh–Nagumo equations and their variants. Basically, the FitzHugh–Nagumo model extracts the essential behavior of the Hodgkin–Huxley fast–slow phase-plane and presents it in a simplified form. Thus, the FitzHugh–Nagumo model has two variables, one fast (v) and one slow (w). The fast variable has a cubic nullcline and is called the excitation variable, while the slow variable is called the recovery variable and has a nullcline that is monotonically increasing. The nullclines have a single intersection point, which, without loss of generality, may be assumed to be at the origin. A schematic diagram of the phase-plane is given in Fig. 4.15, where we introduce some of the notation used later in this section.

The FitzHugh–Nagumo model can be derived from a simplified model of the cell membrane (Fig. 4.16). Here the cell (or membrane patch) consists of three components, a capacitor representing the membrane capacitance, a nonlinear current–voltage device for the fast current, and a resistor, inductor, and battery in series for the recovery current. In the 1960s Nagumo, a Japanese electrical engineer, built this circuit using a tunnel diode as the nonlinear element (Nagumo et al., 1964), thereby attaching his name to this system.

Using Kirchhoff's laws, we can write down equations for the behavior of this membrane circuit diagram. We find

$$C_m \frac{dV}{d\tau} + F(V) + i = -I_0, \tag{4.38}$$

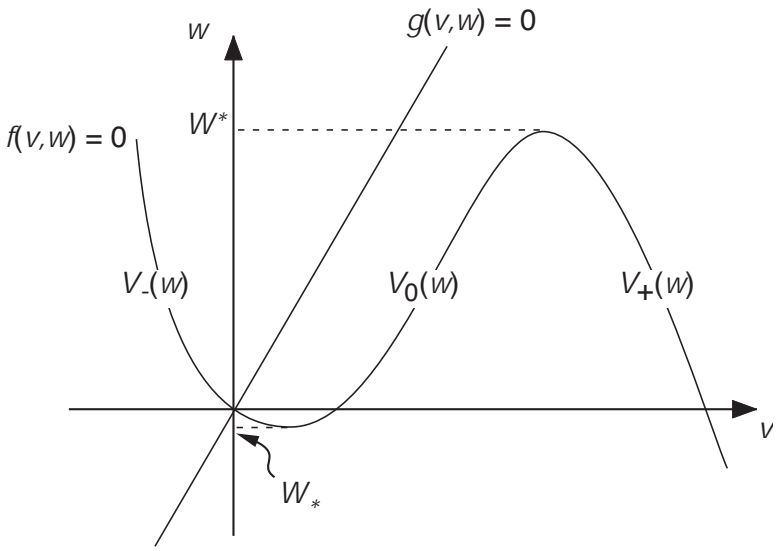

Figure 4.15: Schematic diagram of the generalized FitzHugh–Nagumo phase-plane.

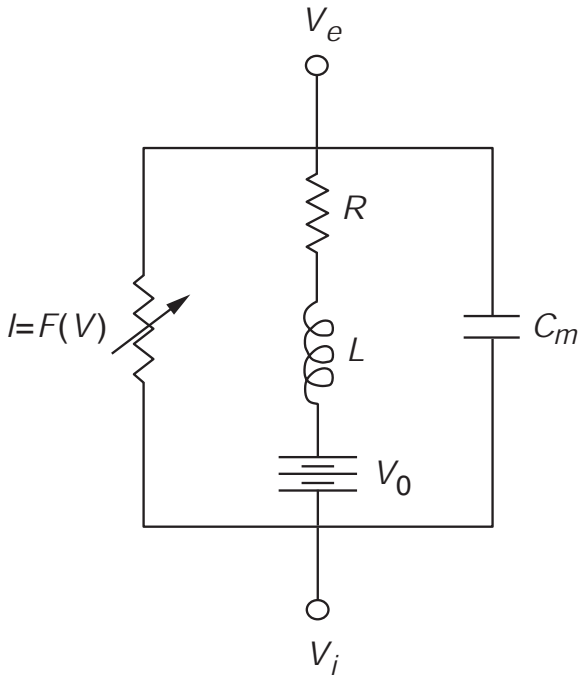

Figure 4.16: Circuit diagram for the FitzHugh–Nagumo equations.

$$L\frac{di}{d\tau} + Ri \quad = \quad V - V_0, \tag{4.39}$$

where I_0 is the applied external current, i is the current through the resistor–inductor, $V = V_i - V_e$ is the membrane potential, and V_0 is the potential gain across the battery. Here τ is used to represent dimensional time because we will shortly introduce t as a dimensionless time variable. The function $F(V)$ is assumed to be of "cubic" shape, having three zeros, of which the smallest $V = 0$ and largest $V = V_1$ are stable solutions of the differential equation $dV/d\tau = -F(V)$. We take R_1 to be the "passive" resistance of the nonlinear element, $R_1 = 1/F'(0)$. Now we introduce the dimensionless variables $v = V/V_1, w = R_1 i/V_1, f(v) = -R_1 F(V_1 v)/V_1$, and $t = L\tau/R_1$. Then (4.38) and (4.39) become

$$\epsilon\frac{dv}{dt} \quad = \quad f(v) - w - w_0, \tag{4.40}$$

$$\frac{dw}{dt} \quad = \quad v - \gamma w - v_0, \tag{4.41}$$

where $\epsilon = R_1^2 C_m/L, w_0 = R_1 I_0/V_1, v_0 = V_0/V_1$, and $\gamma = R/R_1$.

At this point we must specify $f(v)$. As we discussed above, the only requirement is that it have the general shape shown in Fig. 4.15. The classic choice is the cubic polynomial

$$f(v) = Av(v - \alpha)(1 - v) \qquad \text{with} \qquad 0 < \alpha < 1, \tag{4.42}$$

which gives the FitzHugh–Nagumo model. Other choices include the McKean model (McKean, 1970), for which

$$f(v) = H(v - \alpha) - v, \tag{4.43}$$

where H is the Heaviside function. This choice recommends itself because then the model is piecewise linear, allowing explicit solutions of many interesting problems. Another piecewise linear model (also proposed by McKean, 1970) has

$$f(v) = \begin{cases} -v, & \text{for } v < \alpha/2, \\ v - \alpha, & \text{for } \frac{\alpha}{2} < v < \frac{1+\alpha}{2}, \\ 1 - v, & \text{for } v > \frac{1+\alpha}{2}. \end{cases} \tag{4.44}$$

A third piecewise linear model that has found widespread usage is the "Pushchino" model, so named because of its development in Pushchino (about 70 miles south of Moscow), by Krinksy, Panfilov, Pertsov, Zykov, and their coworkers. The details of the Pushchino model are described in Exercise 13.

An important variant of the FitzHugh–Nagumo equations is the *van der Pol oscillator*. An electrical engineer, van der Pol built the circuit using triodes because it exhibits stable oscillations. As there was little interest in oscillatory circuits at the time, he proposed his circuit as a model of an oscillatory cardiac pacemaker (van der Pol and van der Mark, 1928). Since then it has become a classic example of a system with limit cycle behavior and relaxation oscillations, included in almost every textbook on oscillations (see, for example, Stoker, 1950, or Minorsky, 1962).

If we eliminate the resistor R from the circuit (Fig. 4.16), differentiate (4.38), and eliminate the current i, we get the second-order differential equation

$$C_m \frac{d^2V}{d\tau^2} + F'(V)\frac{dV}{d\tau} + \frac{V}{L} = \frac{V_0}{L}. \tag{4.45}$$

Following rescaling, and setting $F(v) = A(v^3/3 - v)$, we arrive at the *van der Pol equation*

$$v'' + a(v^2 - 1)v' + v = 0. \tag{4.46}$$

From now on, by the *generalized FitzHugh–Nagumo equations* we mean the system of equations

$$\epsilon \frac{dv}{dt} = f(v, w) + I, \tag{4.47}$$

$$\frac{dw}{dt} = g(v, w), \tag{4.48}$$

where the nullcline $f(v, w) = 0$ is of "cubic" shape. By this we mean that for a finite range of values of w, there are three solutions $v = v(w)$ of the equation $f(v, w) = 0$. These we will denote by $v = V_-(w), v = V_0(w)$, and $v = V_+(w)$, and where comparison is possible (since these functions need not all exist for the same range of w),

$$V_-(w) \le V_0(w) \le V_+(w). \tag{4.49}$$

We denote the minimal value of w for which $V_-(w)$ exists by W_*, and the maximal value of w for which $V_+(w)$ exists by W^*. For values of w above the nullcline $f(v, w) = 0, f(v, w) < 0$, and below the nullcline, $f(v, w) > 0$ (in other words, $f_w(v, w) < 0$).

The nullcline $g(v, w) = 0$ is assumed to have precisely one intersection with the curve $f(v, w) = 0$. Increasing v beyond the curve $g(v, w) = 0$ makes $g(v, w)$ positive (i.e., $g_v(v, w) > 0$), and decreasing w below the curve $g(v, w) = 0$ increases $g(v, w)$ (hence $g_w(v, w) < 0$). The nullclines f and g are illustrated in Fig. 4.15.

4.2.1 Phase-Plane Behavior

One attractive feature of the FitzHugh–Nagumo model is that because it is a two-variable system, it can be studied using phase-plane techniques. (For an example of a different approach, see Troy, 1976.) There are two characteristic phase portraits possible (shown in Figs. 4.17 and 4.19). By assumption, there is only one steady state, at $v = v^*, w = w^*$, with $f(v^*, w^*) = g(v^*, w^*) = 0$. Without loss of generality, we may assume that this steady state occurs at the origin, as this involves only a shift of the variables. Furthermore, it is typical that the parameter ϵ is a small number. For small ϵ, if the steady state lies on either the left or right solution branch of $f(v, w) = 0$, i.e., the curves $v = V_\pm(w)$, it is linearly stable. Somewhere on the middle solution branch $v = V_0(w)$, near the extremal values of the curve $f(v, w) = 0$, there is a Hopf bifurcation point. That is, if parameters are varied so that the steady state passes through this point, a periodic orbit arises as a continuous solution branch and bifurcates into a stable limit cycle oscillation.

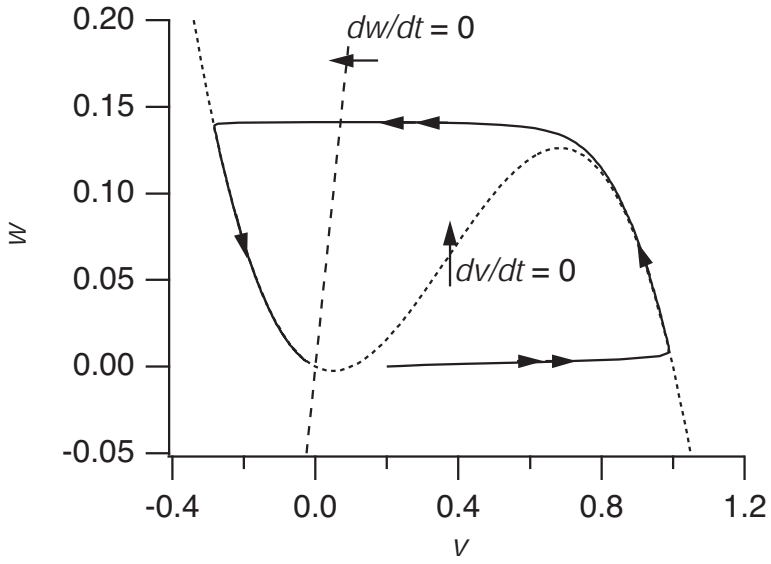

Figure 4.17: Phase portrait for a FitzHugh–Nagumo system with $f(v,w) = v(v - 0.1)(1-v) - w$, $g(v,w) = v - 0.5w$, $\epsilon = 0.01$. For these parameter values the system has a unique globally stable rest point, but is excitable.

When the steady state is on the leftmost branch, but close to the minimum (Fig. 4.17), the system is excitable. This is because even though the steady state is linearly stable, a sufficiently large perturbation from the steady state sends the state variable on a trajectory that runs away from the steady state before eventually returning to rest. Such a trajectory goes rapidly to the rightmost branch, which it hugs as it gradually creeps upward, whence upon reaching the maximum, it goes rapidly to the leftmost branch and then gradually returns to rest, staying close to this branch as it does. Plots of the variables v and w are shown as functions of time in Fig. 4.18.

The mathematical description of these events follows from singular perturbation theory. With $\epsilon \ll 1$, the variable v is a fast variable and the variable w is a slow variable. This means that if possible, v is adjusted rapidly to maintain a pseudo-equilibrium at $f(v,w) = 0$. In other words, if possible, v clings to the stable branches of $f(v,w) = 0$, namely $v = V_{\pm}(w)$. Along these branches the dynamics of w are governed by the reduced dynamics

$$\frac{dw}{dt} = g(V_{\pm}(w), w) = G_{\pm}(w). \tag{4.50}$$

When it is not possible for v to be in quasi-equilibrium, the motion is governed approximately by the differential equations,

$$\frac{dv}{d\tau} = f(v,w), \qquad \frac{dw}{d\tau} = 0, \tag{4.51}$$

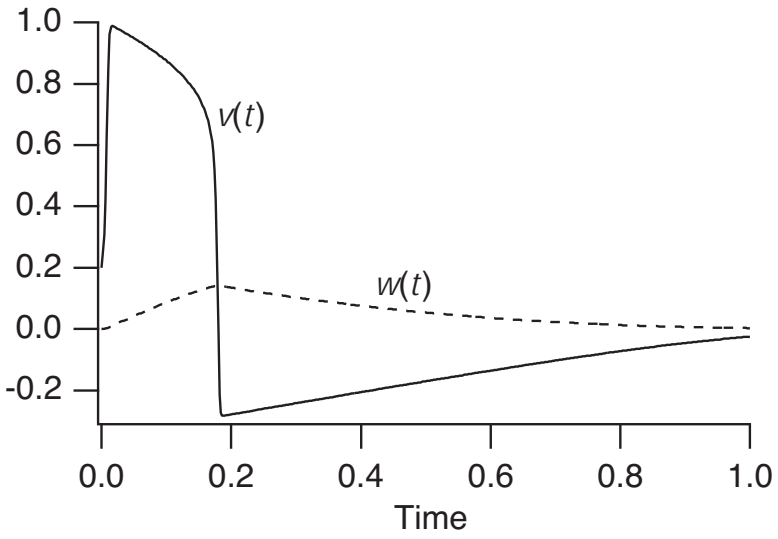

Figure 4.18: Solutions of a FitzHugh–Nagumo system with $f(v,w) = v(v-0.1)(1-v) - w$, $g(v,w) = v - 0.5w$, $\epsilon = 0.01$.

found by making the change of variables to the fast time scale $t = \epsilon\tau$, and then setting $\epsilon = 0$. On this time scale, w is constant, while v equilibrates to a stable solution of $f(v,w) = 0$.

The evolution of v and w starting from specified initial conditions v_0 and w_0 can now be described. Suppose v_0 is greater than the rest value v^*. If $v_0 < V_0(w)$, then v returns directly to the steady state. If $v_0 > V_0(w)$, then v goes rapidly to the upper branch $V_+(w)$ with w remaining nearly constant at w_0. The curve $v = V_0(w)$ is a *threshold curve*. While v remains on the upper branch, w increases according to

$$\frac{dw}{dt} = G_+(w) \tag{4.52}$$

as long as possible. However, in the finite time

$$T_e = \int_{w_0}^{W^*} \frac{dw}{G_+(w)}, \tag{4.53}$$

w reaches the "knee" of the nullcline $f(v,w) = 0$. This period of time constitutes the *excited phase* of the action potential.

When w reaches W^* it is no longer possible for v to stay on the excited branch, so it must return to the lower branch $V_-(w)$. Once on this branch, w decreases following the dynamics

$$\frac{dw}{dt} = G_-(w). \tag{4.54}$$

If the rest point lies on the lower branch, then $G_-(w^*) = 0$, and w gradually returns to rest on the lower branch.

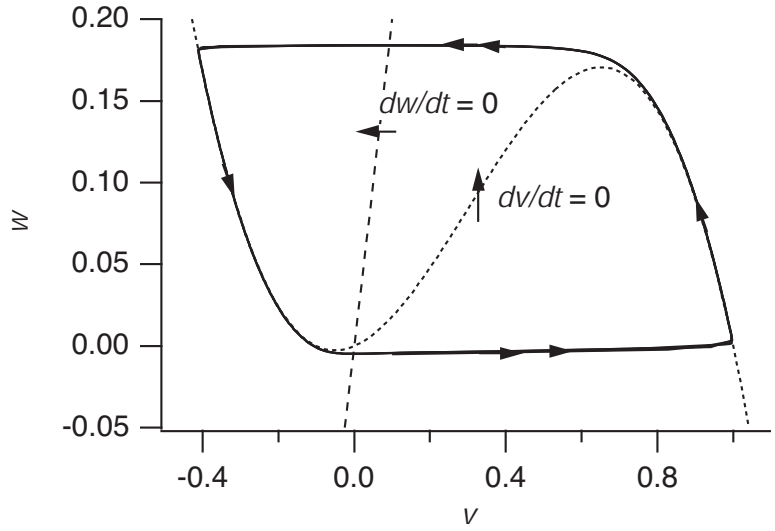

Figure 4.19: Phase portrait for a FitzHugh–Nagumo system with $f(v, w) = v(v + 0.1)\ (1 - v) - w$, $g(v, w) = v - 0.5w$, $\epsilon = 0.01$. For these parameter values, the unique rest point is unstable and there is a globally stable periodic orbit.

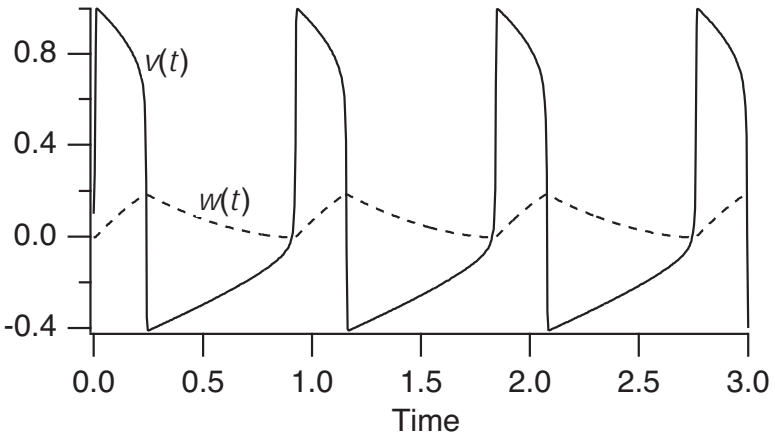

Figure 4.20: Solutions of a FitzHugh–Nagumo system with $f(v, w) = v(v+0.1)(1-v) - w$, $g(v, w) = v - 0.5w$, $\epsilon = 0.01$.

If the rest point lies on the middle branch $V_0(w)$ (depicted in Fig. 4.19), it is unstable. Instead of returning to rest after one excursion on the excited branch, the trajectory alternates periodically between the upper and lower branches, with w varying between W_* and W^*. This periodic limit cycle behavior (with the solution plotted as a function of time shown in Fig. 4.20) is called a *relaxation oscillation.* The period of the oscillation is approximately

$$T = \int_{W_*}^{W^*} \left(\frac{1}{G_+(w)} - \frac{1}{G_-(w)} \right) dw. \tag{4.55}$$

This number is finite because $G_+(w) > 0$, and $G_-(w) < 0$ for all appropriate w.

Exercises

1. Show that, if $k > 1$, then $(1 - e^{-x})^k$ has an inflection point, but $(e^{-x})^k$ does not.

2. Explain why replacing the extracellular sodium with choline has little effect on the resting potential of an axon. Calculate the new resting potential with 90% of the extracellular sodium removed. Why is the same not true if potassium is replaced?

3. Plot the nullclines of the Hodgkin–Huxley fast subsystem. Show that v_r and v_e in the Hodgkin–Huxley fast subsystem are stable steady states, while v_s is a saddle point. Compute the stable manifold of the saddle point and compute sample trajectories in the fast phase-plane, demonstrating the threshold effect.

4. Show how the Hodgkin–Huxley fast subsystem depends on the slow variables; i.e., show how the v nullcline moves as n and h are changed, and demonstrate the saddle-node bifurcation in which v_e and v_s disappear.

5. Plot the nullclines of the fast–slow Hodgkin–Huxley phase-plane and compute a complete action potential.

6. How does the phase plane of the fast-slow Hodgkin-Huxley equations change with applied current? How much applied current in the fast-slow Hodgkin-Huxley equations is needed to generate oscillations? Plot a typical oscillation in the phase plane. Plot the maximum of the oscillation against the applied current to construct a bifurcation diagram.

7. Suppose that in the Hodgkin–Huxley fast–slow phase-plane, v is slowly decreased to $v^* < v_0$ (where v_0 is the steady state), held there for a considerable time, and then released. Describe what happens in qualitative terms, i.e., without actually computing the solution. This is called *anode break excitation* (Hodgkin and Huxley, 1952d. Also see Peskin, 1991). What happens if v is instantaneously decreased to v^* and then released immediately? Why do these two solutions differ?

8. In the text, the Hodgkin-Huxley equations are written in terms of $v = V - V_{\rm eq}$. Show that in terms of V the equations become

$$\begin{aligned} C_m \frac{dV}{dt} &= -\bar{g}_{\rm K} n^4 (V - V_{\rm K}) - \bar{g}_{\rm Na} m^3 h (V - V_{\rm Na}) \\ &\qquad -\bar{g}_{\rm L}(V - V_{\rm L}) + I_{\rm app}, \qquad (4.56) \\ \frac{dm}{dt} &= \alpha_m (1 - m) - \beta_m m, \qquad (4.57) \\ \frac{dn}{dt} &= \alpha_n (1 - n) - \beta_n n, \qquad (4.58) \\ \frac{dh}{dt} &= \alpha_h (1 - h) - \beta_h h, \qquad (4.59) \end{aligned}$$

where (in units of $(\text{ms})^{-1}$),

$$\begin{aligned} \alpha_m &= 0.1 \frac{-40 - V}{\exp\left(\frac{-40-V}{10}\right) - 1}, \qquad (4.60) \\ \beta_m &= 4 \exp\left(\frac{-V - 65}{18}\right), \qquad (4.61) \\ \alpha_h &= 0.07 \exp\left(\frac{-V - 65}{20}\right), \qquad (4.62) \\ \beta_h &= \frac{1}{\exp(\frac{-35-V}{10}) + 1}, \qquad (4.63) \\ \alpha_n &= 0.01 \frac{-55 - V}{\exp(\frac{-55-V}{10}) - 1}, \qquad (4.64) \\ \beta_n &= 0.125 \exp\left(\frac{-V - 65}{80}\right), \qquad (4.65) \end{aligned}$$

and

$$\bar{g}_{\rm Na} = 120, \qquad \bar{g}_{\rm K} = 36, \qquad \bar{g}_{\rm L} = 0.3, \qquad (4.66)$$

$$V_{Na} = 55, \qquad V_K = -77, \qquad V_L = -54.4, \qquad V_{\rm eq} = -65. \qquad (4.67)$$

9. Solve the full Hodgkin–Huxley system numerically with a variety of constant current inputs. For what range of inputs are there self-sustained oscillations? Construct the bifurcation diagram as in Exercise 6.

10. The Hodgkin–Huxley equations are for the squid axon at 6.3°C. Using that the absolute temperature enters the equations through the Nernst equation, determine how changes in temperature affect the behavior of the equations. In particular, simulate the equations at 0°C and 30°C to determine whether the equations become more or less excitable with an increase in temperature.

11. Show that a Hopf bifurcation occurs in the generalized FitzHugh–Nagumo model when $f_v(v^*, w^*) = -\epsilon g_w(v^*, w^*)$, assuming that

$$f_v(v^*, w^*) g_w(v^*, w^*) - g_v(v^*, w^*) f_w(v^*, w^*) > 0.$$

$C_m = 20\ \mu\text{F/cm}^2$	$I_{\text{app}} = 0.06\ \text{mA/cm}^2$
$g_{\text{Ca}} = 4.4\ \text{mS/cm}^2$	$g_{\text{K}} = 8\ \text{mS/cm}^2$
$g_{\text{L}} = 2\ \text{mS/cm}^2$	$\phi = 0.04\ (\text{ms})^{-1}$
$V_1 = -1.2\ \text{mV}$	$V_2 = 18\ \text{mV}$
$V_3 = 2$	$V_4 = 30\ \text{mV}$
$V_{\text{Ca}}^0 = 120\ \text{mV}$	$V_{\text{K}}^0 = -84\ \text{mV}$
$V_{\text{L}} = -60\ \text{mV}$	

Table 4.1: Typical parameter values for the Morris–Lecar model.

12. Morris and Lecar (1981) proposed the following two-variable model of membrane potential for a barnacle muscle fiber:

$$C_m \frac{dV}{dT} + I_{\text{ion}}(V, W) = I_{\text{app}}, \tag{4.68}$$

$$\frac{dW}{dT} = \phi \Lambda(V)[W_\infty(V) - W], \tag{4.69}$$

where V = membrane potential, W = fraction of open K^+ channels, T = time, C_m = membrane capacitance, I_{app} = externally applied current, ϕ = maximum rate for closing K^+ channels, and

$$I_{\text{ion}}(V, W) = g_{\text{Ca}} M_\infty(V)(V - V_{\text{Ca}}^0) + g_{\text{K}} W (V - V_{\text{K}}^0) + g_{\text{L}}(V - V_{\text{L}}^0) \tag{4.70}$$

$$M_\infty(V) = \frac{1}{2}\left(1 + \tanh\left(\frac{V - V_1}{V_2}\right)\right), \tag{4.71}$$

$$W_\infty(V) = \frac{1}{2}\left(1 + \tanh\left(\frac{V - V_3}{V_4}\right)\right), \tag{4.72}$$

$$\Lambda(V) = \cosh\left(\frac{V - V_3}{2V_4}\right). \tag{4.73}$$

Typical rate constants in these equations are shown in Table 12.

(a) Compute the phase portrait of the Morris–Lecar equations. Plot the nullclines and show some typical trajectories, demonstrating how the model is excitable.

(b) Does the Morris–Lecar model exhibit anode break excitation (see Exercise 7)? If not, why not?

13. The Pushchino model is a piecewise linear model of FitzHugh–Nagumo type proposed as a model for the ventricular action potential. The model has

$$f(v, w) = F(v) - w, \tag{4.74}$$

$$g(v, w) = \frac{1}{\tau(v)}(v - w), \tag{4.75}$$

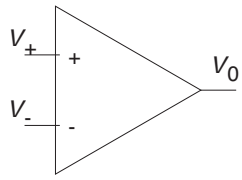

Figure 4.21: Diagram for an operational amplifier (op-amp).

where

$$F(v) = \begin{cases} -30v, & \text{for } v < v_1, \\ \gamma v - 0.12, & \text{for } v_1 < v < v_2, \\ -30(v-1), & \text{for } v > v_2, \end{cases} \tag{4.76}$$

$$\tau(v) = \begin{cases} 2, & \text{for } v < v_1, \\ 16.6, & \text{for } v > v_1, \end{cases} \tag{4.77}$$

with $v_1 = 0.12/(30+\gamma)$ and $v_2 = 30.12/(30+\gamma)$.

Simulate the action potential for this model. What is the effect on the action potential of changing $\tau(v)$?

14. Perhaps the most important example of a nonphysiological excitable system is the Belousov–Zhabotinsky reaction. This reaction denotes the oxidation of malonic acid by bromate in acidic solution in the presence of a transition metal ion catalyst. Kinetic equations describing this reaction are (Tyson and Fife, 1980)

$$\epsilon \frac{du}{dt} = -fv\frac{u-q}{u+q} + u - u^2, \tag{4.78}$$

$$\frac{dv}{dt} = u - v, \tag{4.79}$$

where u denotes the concentration of bromous acid and v denotes the concentration of the oxidized catalyst metal. Typical values for parameters are $\epsilon \approx 0.01, f = 1, q \approx 10^{-4}$. Describe the phase portrait for this system of equations.

15. It is not particularly difficult to build an electrical analogue of the FitzHugh–Nagumo equations with inexpensive and easily obtained electronic components. The parts list for one "cell" (shown in Fig. 4.23) includes two op-amps (operational amplifiers), two power supplies, a few resistors, and two capacitors, all readily available from any consumer electronics store (Keener, 1983).

The key component is an operational amplifier (Fig. 4.21). An op-amp is denoted in a circuit diagram by a triangle with two inputs on the left and a single output from the vertex on the right. Only three circuit connections are shown on a diagram, but two more are assumed, being necessary to connect

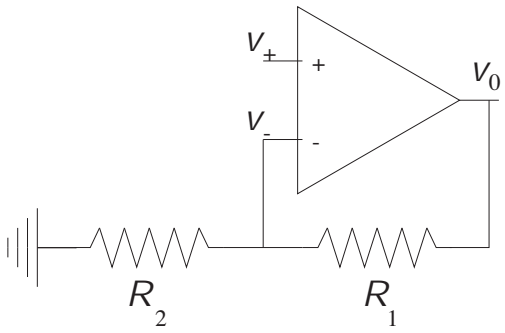

Figure 4.22: Linear amplifier using an op-amp.

with the power supply to operate the op-amp. Corresponding to the supply voltages V_{s-} and V_{s+}, there are voltages V_{r-} and V_{r+}, called the *rail voltages*, which determine the operational range for the output of an op-amp. The job of an op-amp is to compare the two input voltages v_+ and v_-, and if $v_+ > v_-$, to set (if possible) the output voltage v_0 to the high rail voltage V_{r+}, whereas if $v_+ < v_-$, then v_0 is set to V_{r-}. With reliable electronic components it is a good first approximation to assume that the input draws no current, while the output v_0 can supply whatever current is necessary to maintain the required voltage level.

The response of an op-amp to changes in input is not instantaneous, but is described reasonably well by the differential equation

$$\epsilon_s \frac{dv_0}{dt} = g(v_+ - v_-) - v_0. \tag{4.80}$$

The function $g(v)$ is continuous, but quite close to the piecewise constant function

$$g(v) = V_{r+}H(v) + V_{r-}H(-v), \tag{4.81}$$

with $H(v)$ the Heaviside function. The number ϵ_s is small, and is the inverse of the *slew-rate*, which is typically on the order of 10^6–10^7 V/sec. For all of the following circuit analysis, take $\epsilon_s \to 0$.

(a) Show that the simple circuit shown in Fig. 4.22 is a linear amplifier, with

$$v_0 = \frac{R_1 + R_2}{R_2} v_+, \tag{4.82}$$

provided that v_0 is within the range of the rail voltages.

(b) Show that if $R_1 = 0, R_2 = \infty$, then the device in Fig. 4.22 becomes a *voltage follower* with $v_0 = v_+$.

(c) Find the governing equations for the circuit in Fig. 4.23, assuming that the rail voltages for op-amp 2 are well within the range of the rail voltages for op-amp 1.

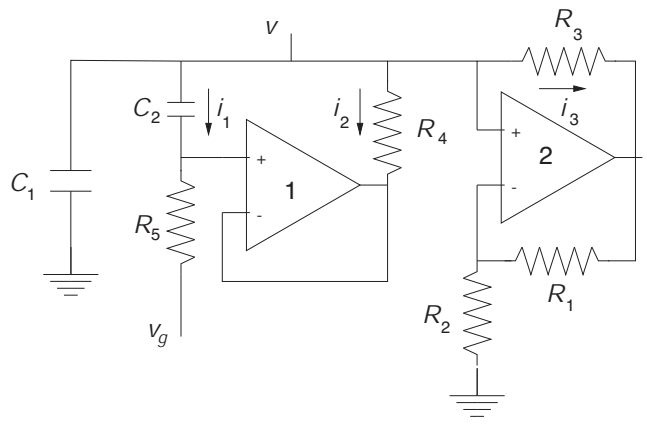

Figure 4.23: FitzHugh–Nagumo circuit using op-amps.

Table 4.2: Parts list for the FitzHugh–Nagumo analog circuit.

2 LM 741 op-amps (National Semiconductor)	
$R_1 = R_2 = 100\text{k}\Omega$	$R_3 = 2.4\Omega$
$R_4 = 1\text{k}\Omega$	$R_5 = 10\text{k}\Omega$
$C_1 = 0.01\mu\text{F}$	$C_2 = 0.5\mu\text{F}$
Power supplies:	
± 15V for op-amp #1	± 12V for op-amp #2

Show that

$$C_1 \frac{dv}{dt} + i_2 \left(1 - \frac{R_4}{R_5}\right) + \frac{F(v)}{R_3} + \frac{v - v_g}{R_5} = 0, \tag{4.83}$$

$$C_2 R_5 \frac{di_2}{dt} + R_4 i_2 = v - v_g, \tag{4.84}$$

where $F(v)$ is the piecewise linear function

$$F(v) = \begin{cases} v - V_{r_+}, & \text{for } v > \alpha V_{r_+}, \\ -\frac{R_1}{R_2} v, & \text{for } \alpha V_{r_-} \le v \le \alpha V_{r_+}, \\ v - V_{r_-}, & \text{for } v < \alpha V_{r_-}, \end{cases} \tag{4.85}$$

and $\alpha = \frac{R_2}{R_1 + R_2}$.

(d) Sketch the phase portrait for these circuit equations. Show that this is a piecewise linear FitzHugh–Nagumo system.

(e) Use the singular perturbation approximation (4.55) to estimate the period of oscillation for the piecewise linear analog FitzHugh–Nagumo circuit in Fig. 4.23.