PHYSIOLOGY

Fifth Edition

Edited by Ewald E. Selkurt, Ph.D.

PHYSIOLOGY

Fifth Edition

Edited by Ewald E. Selkurt, Ph.D.

Distinguished Professor, Department of Physiology and Biophysics,
Indiana University School of Medicine,
Indianapolis

LITTLE, BROWN AND COMPANY BOSTON/TORONTO

Fifth Edition

Library of Congress Catalog Card No. 83-80301

ISBN 0-316-78038-3

Printed in the United States of America

DON

CONTRIBUTING AUTHORS

The contributors to *Physiology* are affiliated with the Department of Physiology and Biophysics, Indiana University School of Medicine, Indianapolis.

EWALD E. SELKURT, Ph.D., *Distinguished Professor; Editor*

WILLIAM M. ARMSTRONG, Ph.D., *Professor*

JULIUS J. FRIEDMAN, Ph.D., *Professor*

KALMAN GREENSPAN,* Ph.D., *Professor*

WILLIAM V. JUDY,† Ph.D., *Adjunct Associate Professor*

LEON K. KNOEBEL, Ph.D., *Professor*

RICHARD A. MEISS,‡ Ph.D., *Associate Professor*

WARD W. MOORE,§ Ph.D., *Associate Dean*

SIDNEY OCHS, Ph.D., *Professor*

RODNEY A. RHOADES, *Professor and Chairman*

CARL F. ROTHE, *Professor*

*Professor of Physiology and Medicine, Terre Haute (Indiana) Center for Medical Education.
†Clinical Research Division, Methodist Hospital, Indianapolis.
‡Joint Appointment in Department of Obstetrics and Gynecology.
§Director of Medical Science Program, Bloomington, Indiana.

THE CELL MEMBRANE AND BIOLOGICAL TRANSPORT

1

William M. Armstrong

Physiology is concerned with the overall functioning of the tissues and organs of which the body is composed. Essentially, these tissues and organs are *organized* assemblies of large numbers of cells, often of several different types. Through the operation of various *control mechanisms* (e.g., neural, hormonal), these complex assemblies respond to specific stimuli in an integrated fashion. One sees, then, that the cell is, in a very real sense, a fundamental functional unit in relation to the body as a whole, and that the behavior of single cells is a logical takeoff point for the study of more complex physiological systems.

Living cells are themselves complex and highly organized. In addition to a fluid phase (intracellular fluid or cytoplasm), the interior of a typical cell contains a number of *inclusions* or *organelles* (e.g., nucleus, mitochondria, lysosomes). Modern research has shown that these subcellular particles are again highly organized structures that play a vital role in the overall activity of the cell. For didactic purposes the detailed study of these structures is, nowadays, to a large extent, included in the formal disciplines of biochemistry and cell biology. Physiology proper is often considered to begin at the point where the cell as a whole interacts with its external environment or with neighboring cells, that is, with the exchanges of matter and energy that take place across its outer limiting membrane, or *plasma membrane*. In keeping with this practice this chapter will focus on these exchanges. The student should remember, however, that this division of different aspects of cell function between various academic disciplines is purely a matter of convenience. It has no real basis in the domain of biology or medicine and is, quite properly, often ignored in the classroom and completely disregarded in the clinic or research laboratory.

Cytoplasm, or intracellular fluid, differs markedly in composition from the external medium (blood or interstitial fluid) bathing the cell. In intact, normally functioning cells, some of these differences in composition remain virtually constant over long periods of time. At the same time there are continuous exchanges of material between the cell interior and the external environment, with some substances

entering the cell and others leaving it. Evidently, cells must possess a mechanism or mechanisms by which the entry and exit of dissolved solutes are regulated. Two alternative kinds of mechanism have been postulated. One invokes the idea that the cytoplasm as a whole is capable of selectively accumulating certain substances while excluding others. The other assigns the control of the transfer of materials between the cell and its environment to a special region at or near the periphery of the cell, the *cell membrane*, or *plasma membrane*. At present, although the idea of functional selectivity by the cytoplasm as a whole is vigorously championed by some workers (Ling, 1962), most physiologists accept the special role of the membrane in the regulation of cellular composition as the most satisfactory interpretation of the available experimental evidence. Therefore, in this chapter, cellular transport mechanisms will be discussed from this point of view.

THE CELL MEMBRANE

STRUCTURE. At present, the existence of the plasma membrane as a discrete entity is not seriously challenged, even by the few who dispute its importance in controlling the exchange of material between the cell interior and the external environment. Moreover, during recent years, there has been a rapidly growing awareness of the importance of internal membranous structures (e.g., the *sarcoplasmic reticulum* of muscle) as regulators of overall cellular function. It is a truism of contemporary biology that a complete understanding of the function of any system requires a detailed knowledge of its molecular architecture. The problem of the molecular structure of cell membranes, particularly as it relates to their functional properties, therefore is of crucial importance to the effective progress of physiology and medicine (Andreoli et al., 1978).

Unfortunately, we are still a long way from a complete understanding of the fine details of the structure of the cell membrane. Nevertheless, recent theoretical and experimental research has reached a stage where the arrangement of its major components can be inferred with some confidence. In this section, one current model of membrane structure, the fluid mosaic model of Singer (Singer and Nicolson, 1972), will be discussed, since it appears to represent the most convincing synthesis of theoretical concepts and experimental data presently available. However, some cautionary comments are appropriate.

First, there is by no means unanimous agreement on even the grosser aspects of membrane structure. A number of models of varying ingenuity, complexity, and plausibility have been proposed (Hendler, 1971; see also the chapter by Robertson in Andreoli et al., 1978). Second, as a result of vigorous research utilizing a variety of physical, chemical, and biological techniques, our insights into the molecular structure of membranes are expanding so rapidly that any model of the membrane that can be proposed at present must be regarded as no more than a temporary expedient. Third, in considering any generalization about membrane structure, one must bear in mind that cell membranes exhibit considerable diversity in composition. Thus, there may be—and often are—significant differences in the structural details of the membranes of different cells and between different membranous structures within the same cell. This phenomenon becomes especially apparent when one considers highly specialized membranous organelles such as the *myelin sheath* of nerve or the *inner mitochondrial membrane*. Finally, the inertia inherent in human affairs has led to the retention, in otherwise unexceptionable textbooks, of earlier models for membrane structure (e.g., that of Davson, 1970).* One supposes that time will remedy this.

CHEMICAL COMPOSITION OF CELL MEMBRANES. The principal structural components of cell membranes are lipids (mainly cholesterol and phospholipids), proteins, and oligosaccharides. In addition, membranes contain water (some of which is probably in a more or less highly *ordered* state because of its close association with the ionized and polar groups of phospholipids and proteins) and small amounts of low-molecular-weight species such as inorganic ions. By far the most abundant components are proteins and lipids. Together, these account for approximately 95 percent, by weight, of the nonaqueous components of the membrane. By contrast, oligosaccharides normally account for approximately 3 to 4 percent and are often found as integral components of membrane glycoproteins. Their importance, however, far outweighs their relative scarcity, since specific polysaccharides are frequently key elements in the cell surface *receptors* that are implicated in the recognition and binding of ligand molecules. Important interactions of this type are those involved in immunological responses.

In general, the plasma membranes of eukaryotic cells have a protein/lipid ratio, by weight, of approximately 1/1. In specialized membranous structures the ratio may be different. For example, the myelin sheath of nerve (see Chap. 2) is a membranous structure that appears to function principally as a barrier to the movement of ions (i.e., the passage of electrical current). Apart from this, the myelin is

*For a brief obituary of this long-lived model, the reader is referred to the fourth edition of this text.

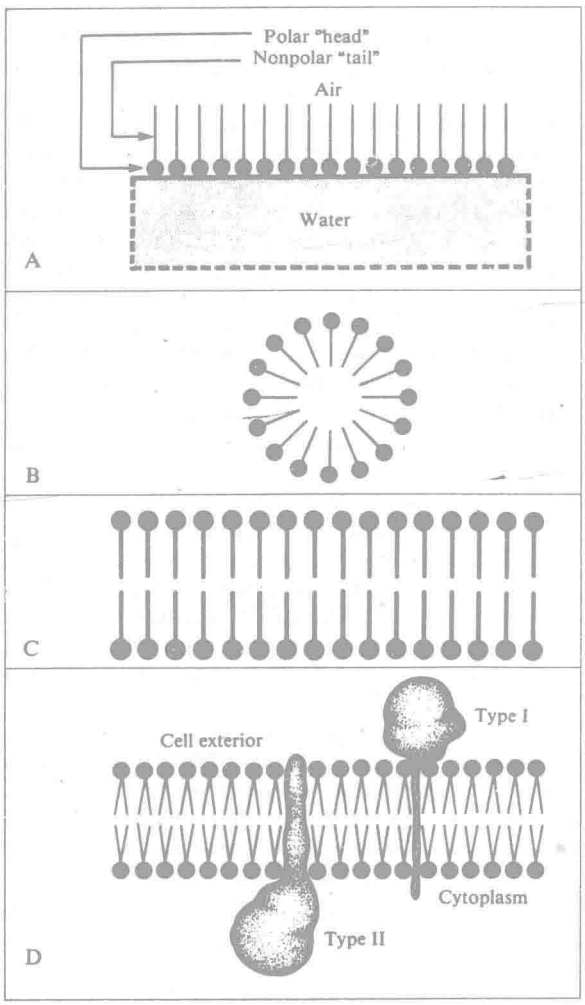

Fig. 1-1. (A) Orientation of a monolayer of amphiphilic lipid at the interface between a polar and a nonpolar phase. (B) Phospholipid micelle in a polar liquid. (C) Phospholipid bilayer. (D) Insertion of intrinsic proteins in the cell membrane.

relatively inert. The protein/lipid weight ratio in myelin is about 1/4. By contrast, the inner mitochondrial membrane, which is the site of a variety of enzymatic reactions, has a protein/lipid weight ratio of about 3/1.

ORGANIZATION OF MEMBRANE LIPIDS AND PROTEINS. In considering the problem of cell membrane ultrastructure, the basic question is the way in which the lipid and protein moieties of the membrane are arranged. These substances contain both polar and nonpolar groups and are therefore *amphiphilic*, that is, display both *hydrophilic* and *lipophilic* or *hydrophobic* properties. Their polar groups tend to interact strongly with water or other polar solvents (hydrophilic reactions), and at the same time their nonpolar residues tend to interact with nonpolar solvents or other nonpolar entities (hydrophobic reactions). Thus, if proteins or lipids are spread as a layer one molecule thick (*monolayer*) at the interface between water and a nonpolar phase (e.g., oil or air), they tend to orient themselves so that their polar groups are in contact with the polar phase while their nonpolar portions are in contact with the nonpolar phase (Fig. 1-1A). In aqueous solutions, polar-nonpolar lipids are likely to clump together to form *micelles* (Fig. 1-1B), in which the polar "heads" are oriented outward toward the water, while the nonpolar "tails" are tucked inside the micelle, where their interactions with the polar water molecules are minimized and their interactions with each other are maximized. Similarly, proteins in aqueous solution are oriented so that as many as possible of their ionic residues are in contact with water (mostly at the outside of the protein molecule) and as many as possible of their hydrophobic residues are folded inside the molecule, away from the external aqueous environment.

This tendency of amphiphilic substances to orient themselves so that, as far as possible, both hydrophilic and hydrophobic (van der Waals) interactions are maximized is a reflection of a very general principle embodied in the second law of thermodynamics: The state of maximum stability for any system is that state in which the free energy of the system is at a minimum. Looked at in another way, the state of minimum free energy is the most probable state for the system to achieve as a result of spontaneous changes; or, conversely, one can say that spontaneous changes in any system always proceed in the direction of decreasing free energy. Applying this principle to the problem of the orientation of lipids and proteins in cell membranes, which may be regarded essentially as a thin layer, about 75 to 100 Å in thickness, bounded on both sides by an aqueous solution, one can predict that, within the constraints imposed by the overall geometry of the membrane, both these molecular species will dispose themselves in such a way that their capacity for hydrophilic and hydrophobic interaction with neighboring molecules or groups is maximized. A conformation that permits phospholipids to achieve this condition within the cell membrane is the *bilayer* (Fig. 1-1C). In this structure, the polar head groups of the phospholipids make contact with the aqueous media on both sides of the membrane, while the predominantly hydrocarbon tails are sequestered inside the bilayer.

The suggestion that the lipid moiety of the cell membrane

is organized as a bilayer was initially put forward about 50 years ago on the basis of studies with lipids extracted from red blood cell "ghosts," i.e., red blood cells from which the internal contents have escaped following *hemolysis* (see Appendix). Since then, this concept has not only remained viable but has been strongly reinforced by recent experimental work (Singer and Nicolson, 1972). Thus, despite the fact that a number of models for cell membrane structure have been proposed in which most membrane lipids are supposed to be organized as micelles, either alone or as mixed lipoprotein micelles (Lucy, 1968; Hendler, 1971), the existence of a lipid bilayer as an essential structural component of the cell membrane now seems well established. In particular, artificially produced lipid bilayers have been shown to possess many properties that are similar to those of naturally occurring cell membranes (Christensen, 1975). One may therefore conclude (Singer and Nicolson, 1972) that the greater part (at least 70 percent) of the membrane lipid is in the bilayer form, though some fraction (about 30 percent) may be organized in a different form, either alone or with proteins (i.e., as lipoproteins).

In the model proposed by Singer and Nicolson (1972), membrane proteins are divided into two classes. Singer and Nicolson originally called these classes *integral* and *peripheral* proteins, but the preferred terminology at present seems to be *intrinsic* and *extrinsic* proteins. (Such abrupt changes in fashion are somewhat baffling, but one must keep up to date!) Intrinsic proteins account for about 70 percent of the total amount of protein associated with the membrane. They are rather tightly bound to the membrane and can only be removed from it by relatively harsh treatment (e.g., extraction with detergents). Thus, they are considered to be integral components of the membrane matrix. In Singer and Nicolson's model, intrinsic proteins are considered to be inserted into the lipid bilayer as discrete entities at somewhat irregular intervals. Looked at from above, the membrane would present (on a microscale, of course) something of the aspect of a mosaic structure—hence the use of *mosaic* by Singer and Nicolson.

The insertion of intrinsic proteins into the lipid bilayer is illustrated in simple schematic form in Figure 1-1D. Two points about this schema may be noted. First, although it was not a part of Singer and Nicolson's original proposal, it is now believed that intrinsic proteins nearly always pass completely through the bilayer and are partly exposed to the predominantly aqueous media on both sides of it. Second, the disposition of these proteins with respect to the bilayer is not symmetrical. *Type I* proteins have most of their mass in the aqueous medium outside the membrane. *Type II* proteins reside mainly on the inner (cytoplasmic) side. As discussed later, this is not a trivial difference. A third important point (not shown in Fig. 1-1D) is that many intrinsic proteins, especially those involved in membrane transport (see p. 10), are not *monomers*. They are *oligomers*, i.e., they are composed of several subunits. For example, the ubiquitous membrane-bound Na^+-K^+ adenosine triphosphatase (ATPase) has two kinds of subunits, an alpha unit (90,000 daltons) and a beta unit (40,000 daltons), the latter being a glycoprotein. The complete enzyme has a total of four subunits and has the composition $\alpha_2\beta_2$. In the acetylcholine receptor protein that is present in the postsynaptic membrane of the neuromuscular junction (see Chap. 3), five different kinds of protein subunits have been identified. (The debilitating disease myasthenia gravis is characterized by a deficiency in this receptor protein—a good example of a physiological condition associated with a specific membrane defect.)

Within the lipid bilayer the polypeptide chains of intrinsic proteins are so arranged that the greatest possible number of hydrophobic (lipophilic) groups is externalized. Thus, van der Waals interactions between the "submerged" portions of the intrinsic proteins and the hydrophobic "tails" of the lipid bilayer are mainly responsible for "anchoring" the protein components in the complex shown schematically in Figure 1-1D.

The second class of proteins associated with the membrane are the so-called extrinsic proteins. These lie on the surface of the membrane, outside the matrix structure illustrated in Figure 1-1D and are not shown in this figure. They are attached to the membrane proper mainly by polar interactions between their polar groups and the polar "heads" of the membrane lipids and can be detached by relatively gentle methods, e.g., hypotonic solutions, strong salt solutions, or removal of divalent cations during membrane fractionation. In principle, one would expect to find extrinsic proteins associated with both the outer and the cytoplasmic surface of the membrane. In practice, they are nearly always found on the cytoplasmic side. This may, however, be an artifact, since extrinsic proteins at the outer surface could easily become detached and "lost" during early stages of membrane fractionation.

MEMBRANE DYNAMICS. In recent years a number of powerful techniques have become available for the investigation of molecular motion within the membrane. These include electron spin resonance, nuclear magnetic resonance, Raman spectroscopy, and the use of fluorescent probes. As a result, much is now known about the freedom of individual molecules to move within the membrane and about the constraints that restrict such movement.

Turning first to lipids, the study of intramolecular motion (the relative movement of different parts of a single molecule) and of the rotational motion of whole molecules indicates that the interior of the bilayer is a fluid structure (the basis of Singer and Nicolson's "fluid mosaic").* Investigation of the translational movement of lipids in membranes has revealed the important fact that whereas translational movement within the plane of the bilayer, that is, within one-half of the bilayer (monolayer), occurs readily (at least above the transition temperature*), migration across the bilayer (from one monolayer to the other—the so-called flip-flop) is difficult and probably occurs very rarely. This is easily understood on theoretical grounds. Lateral movement within one monolayer does not necessitate any change in the relative orientations of polar and nonpolar portions of the lipid molecule. On the other hand, flip-flop requires that the polar head group of the moving molecule be dragged through the nonpolar interior of the bilayer—an inherently improbable occurrence on energetic grounds.

It is easily seen that the same arguments apply, with even more force, to the large intrinsic protein molecules. These may have some freedom of movement in the plane of the bilayer; indeed, they are often compared to icebergs floating in a "sea" of lipid. Without wishing to push the analogy too far, one might add that, like real icebergs, they may drift but are very unlikely to turn cartwheels!

An interesting consequence of the relative absence of flip-flop movements on the part of membrane lipids is as follows: By suitable techniques, e.g., freeze fracture, it is possible to "peel" apart the two halves of the bilayer. The lipid composition of each half can then be determined. When this is done, it is usually found that individual lipids are asymmetrically distributed between the two halves. For example, with human red blood cells, it was found that, although both halves of the bilayer had identical amounts (per unit weight) of total phospholipid, the outside monolayer was relatively rich in sphingomyelin and phosphatidylcholine. The inner monolayer had relatively small proportions of these substances but had high proportions of phosphatidylethanolamine and phosphatidylserine.

*This fluidity is very dependent on the lipid composition of the membrane. Certain lipids, notably cholesterol, exert a "condensing" effect and reduce the bilayer fluidity. Furthermore, highly interesting "phase transitions" have been reported both for artificial bilayers and natural membranes. These are characterized by a sudden marked transition from a highly fluid to a viscous, gel-like state (and vice versa) over a very narrow range of temperature. The transition temperatures for these changes are, again, highly composition dependent (e.g., cholesterol can markedly increase the transition temperature of a mixed lipid bilayer). Membrane fluidity and phase transitions may well turn out to be of great significance in physiology and medicine. It has been suggested, for example, that altered membrane fluidity may be an important factor in the action of certain anesthetics and that its composition dependence might be a significant variable in temperature adaptation. However, such suggestions are still largely speculative.

FUNCTIONS OF MEMBRANE PROTEINS AND LIPIDS. Clearly (Fig. 1-1C), the lipid bilayer forms the structural core or matrix of the membrane. In addition, it forms an important barrier to the passage of charged particles, or *ions*, across the membrane and effectively restricts their diffusive movement to a limited number of channels that are associated with specific membrane proteins. Since ions are the charge carriers involved in the passage of *electrical currents* across cell membranes, the lipid bilayer is responsible for the high electrical resistance possessed by these structures. Measured membrane resistances are of the order of 1000 to 2000 ohm·cm^{-2}. A film of extracellular fluid or cytoplasm with the same dimensions (1 cm^2 × 100 Å thick) would have an electrical resistance of approximately 1×10^{-4} ohm, i.e., about 10 million times less.

The high electrical resistance of the cell membrane is important for the following reason: The normal function of cells requires that they maintain an electrical potential difference (the *membrane potential*, or *resting potential*) across their membranes. This is an essential part of the property of excitability in nerve and muscle cells (see Chaps. 2 and 3) and plays a key role in many other cell and tissue functions, e.g., absorption and secretion. As will be discussed later, the existence of cell membrane potentials is associated with the asymmetrical distribution of ions (notably K^+ and Na^+ ions) between the cell interior and its environment. The high electrical resistance of the cell membrane permits these ionic concentration differences to be preserved without prohibitive demands on the cell's energy resources. Finally, the lipid bilayer plays a role in the *permeability* of the membrane to certain substances. Molecules that have a certain degree of lipid solubility (see Factors Influencing Diffusive Transport) appear to penetrate cells by solution in, and passage through, the lipid bilayer. The significance of this fact to the practicing physician is that many drugs belong to this class of substance.

Type I intrinsic proteins have all their functional properties (as well as most of their mass) in the aqueous environment outside the lipid bilayer. The *intramembranous* part of these proteins is usually a relatively short amino acid sequence (about 20 or so amino acid residues) sufficient to span the bilayer in a single strand as an alpha-helix and "anchor" the protein to it. Proteins of this type are frequently involved in immunological reactions. Examples are the major sialoglycoprotein of the red blood cell membrane and the histocompatibility antigens of human cells. The gly-

coprotein nature of type I intrinsic proteins is noteworthy; e.g., in the most intensively studied example, the major sialoglycoprotein of the red blood cell, two-thirds of the total molecular mass consists of carbohydrate residues.

Type II intrinsic proteins (Fig. 1-1D) are characterized by the fact that most of their mass lies in the cytoplasm and that only a very small fraction of the molecule (10 percent or less) is exposed to the external environment on the noncytoplasmic side of the membrane. They appear to fall into two major categories. The first category consists of proteins in which all *functional properties* are located on the cytoplasmic side of the membrane. These proteins appear to be, in the main, membrane-bound enzymes (e.g., cytochrome b_5, cytochrome b_5 reductase). The second category includes proteins that subserve *transmembrane* functions, e.g., transport processes. Examples are the ubiquitous Na^+-K^+ ATPase, associated with Na^+ "pumping" in many cells; Ca^{2+} ATPase; the acetylcholine receptor protein; and the so-called (because of the position in which it appears during sodium dodecyl sulfate [SDS] gel electrophoresis) band 3 protein of the human red cell. Band 3 protein is involved in anion transport across the red cell membrane. All these proteins are involved in the transfer of specific substances across the cell membrane. It seems reasonable to suppose that proteins engaged in various kinds of transmembrane signaling (e.g., hormone receptors) will also be found to belong to this category. These proteins are, again, glycoproteins. They are also oligomers, and their oligomeric structure extends to the intramembranous moiety. It seems highly probable that the transport sites in these proteins (channels or pores) are located at the interfaces between the oligomeric subunits.

The precise function of many extrinsic proteins remains to be clarified. In some cases they may be components of enzyme complexes; e.g., the band 6 protein of the human red blood cell is a subunit of the enzyme glyceraldehyde phosphate dehydrogenase. Spectrin, a large and complex extrinsic protein, has been identified in the red blood cell membrane. It has been suggested that spectrin provides a protein network (cytoskeleton) that stabilizes the lipid bilayer. So far, this idea remains speculative. However, it is noteworthy that large polypeptides, comparable in size to spectrin, have been found in other kinds of cells, e.g., smooth muscle cells and macrophages.

THE TRANSPORT OF SMALL MOLECULES AND IONS ACROSS THE CELL MEMBRANE

MECHANISMS OF MEMBRANE TRANSPORT: GENERAL PRINCIPLES. The term *membrane transport*, as generally employed in physiology, has come to have a rather arbitrarily restricted meaning. Ordinarily, any general discussion, such as that presented here, is largely confined to transmembrane movements of water and low-molecular-weight solutes, the latter being assumed to be in the monomeric state. However, many cells of the body subserve important transport functions that are not subject to the rules to be discussed in detail subsequently. These transport functions may be absorptive, secretory, or protective. For example, the epithelial cells lining the lumen of the small intestine absorb fats. Multimolecular aggregates such as micelles (Fig. 1-1B) play an important role in this process. Again, the transmembrane transport of large molecules such as proteins is of vital importance to the normal functioning of the body. One may cite as examples the intestinal absorption of immunoglobulins by the newborn and the secretion of salivary and pancreatic enzymes from the cells in which they are synthesized to the lumen of the alimentary canal. Scavenging by the cells of the reticuloendothelial system of macromolecular or particulate matter of exogenous (e.g., bacterial) origin is an important element in the body's defense against infection or other toxic invasion.

It is generally believed that the process of *exocytosis* and its converse, *endocytosis*, are notable factors in the transport of relatively large molecules and aggregates. In this process a segment of the cell membrane folds around the transported species to form a vesicle that then separates either externally (exocytosis) or internally (endocytosis) from the rest of the membrane. The membranous "wall" of the vesicle is then degraded enzymatically, thus releasing the transported species to the external or internal medium. These processes, though they are of great physiological importance and will in some instances be discussed elsewhere in this book, will not be considered further in this chapter.

Even with the restrictions that have been noted, the characterization of membrane transport processes is complex, particularly because certain substances (e.g., water, Na^+ and K^+ ions) may be transported across the cell membrane at the same time by at least two different mechanisms. However, most attempts to categorize the transmembrane transport of simple solutes are structured around two basic considerations. These are the *kinetics* and the *energetics* of membrane transport. By a happy coincidence, the "rule of twos" applies in turn to each of these considerations. Thus, *kinetically*, membrane transport processes can be categorized as either *diffusive* or *saturable*. *Energetically*, they can be classified as *passive* (downhill) or *active* (uphill). Passive (downhill) transport processes are those in which net solute transfer ceases when the solute is at equilibrium across the cell membrane; i.e., when the *driv-*

ing force or gradient that effects a net transmembrane flow or flux of solute becomes zero. Thus, passive transport tends to dissipate those forces or gradients that can give rise to net flows of energy or matter across the cell membrane.

Clearly, systems in which passive flows alone occurred would be inconsistent with life. Therefore, in living cell membranes, one finds, as one would expect, a complex interplay and balance between passive and active transport processes. In the latter, the end point (zero net flow) is a steady state that differs from equilibrium in that finite driving forces (and the gradients that give rise to them) are conserved. It is clear, on thermodynamic grounds, that such processes, if they are to operate continuously, require an energy input from an external source. In the final analysis, this input can only be achieved at the expense of the metabolic energy of the cell. Indeed, a large part of the total metabolism is coupled to active transport in certain cells, such as those of the renal cortex, where the transport of large amounts of water and solutes is a major cellular activity.

DIFFUSIVE TRANSPORT. Diffusive membrane transport processes are so called because they bear a striking resemblance to diffusion in homogeneous physical systems, i.e., gases and liquids. Diffusion results from the spontaneous tendency for any substance (e.g., a gas or a solute in aqueous solution) to distribute itself uniformly throughout the whole space available to it. For example, if a solute is present initially at a higher concentration in one region of a solution than in another, with the passage of time the concentration difference disappears, and the solute concentration becomes uniform throughout the solution. As long as a solute concentration difference, or *concentration gradient*, exists between two different regions of the solution, there will be a spontaneous net movement of solute from the region of higher to the region of lower concentration. When all concentration gradients have been dissipated, *net* solute movement between different regions of the solution will cease. The solution is then said to be in a state of equilibrium with respect to the solute.

The rate of net solute diffusion between two regions of a solution containing unequal concentrations of solute is obviously equal to $-dS/dt$, the amount of solute passing from the region of higher to that of lower solute concentration per unit time. At constant temperature, this quantity is given by Fick's equation, as follows:

$$J = -\frac{dS}{dt} = DA(S_1 - S_2) = DA\,\Delta S \qquad (1)$$

In this equation, S_1 and S_2 are the solute concentrations in the regions of higher and lower concentration respectively; A is the cross-sectional area of the boundary between these regions (the area across which diffusion is taking place); and D is the *diffusion coefficient*, or *diffusivity*, of the solute. It is apparent from equation 1 that D is numerically equal to the rate of diffusion across the unit area (1 cm^2) when the concentration gradient, ΔS, across the boundary layer is unity. Thus, D is a measure of the inherent ability of the solute molecules to move through the solution. It is also clear from equation 1 that the net rate of diffusion at constant temperature depends on the concentration gradient, ΔS, and on the magnitude of D; D has been found to depend on temperature, becoming larger as the temperature is increased, and also varies with the molecular weight of the dissolved substance, becoming smaller as the molecular weight increases. It likewise depends to some extent on molecular shape as well as on molecular size. With large molecules in particular, the extent and strength of their interactions with other molecules, and hence the overall resistance to their diffusion in aqueous solution, are governed by shape as well as size. In these circumstances the relationship between D and molecular weight can be complex.

In thermodynamic terms the tendency of a solution to move spontaneously in the direction of equilibrium can be viewed as a reflection of the general tendency of natural systems to strive to attain the condition of minimum free energy, or maximum *entropy*. It can be shown by the methods of statistical thermodynamics that this condition corresponds to the most probable state of the system, which is, in turn, the state of *maximum disorder*, or *randomness*.

When a concentration gradient, ΔS, for a solute exists across a cell membrane, the rate of solute diffusion across the membrane is given by

$$J = \frac{PA}{x} \cdot \Delta S \qquad (2)$$

Equation 2 is analogous to equation 1 except for the term x, which represents the thickness of the membrane, and the term P, the *permeability coefficient* for the solute, which replaces the diffusion coefficient (D) in equation 1. The substitution of P for D merely reflects the fact that in these conditions it is the rate of diffusion of the solute within the membrane, rather than its diffusivity in free solution, that is the rate-determining step in the penetration process.

Equation 2 states that, for a diffusive process, the net rate of penetration, J, is a linear function of ΔS. This is illustrated graphically in Figure 1-2A. It is evident that the slope of the line relating J to ΔS in the graph is equal to PA/x. Thus, if one knows A and x, P is readily determined. In practice, A, the area of the membrane, and x may be difficult to determine with accuracy. However, if these vari-

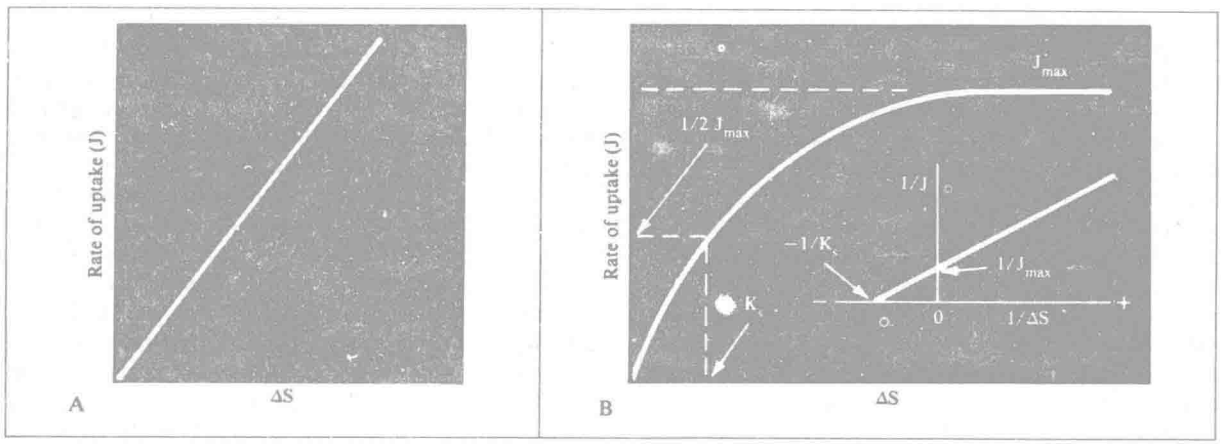

Fig. 1-2. Kinetic characteristics of transport processes. (A) Diffusion type. (B) Carrier-mediated transport. (From H. Stern and D. L. Nancy. *The Biology of Cells*. New York: Wiley, 1965. P. 326.)

ables are assumed to remain constant, the *relative* permeabilities of a given cell species for a number of solutes can be derived from uptake studies such as the one illustrated in Figure 1-2A.

It is important to note that in the present discussion and elsewhere in this chapter, *net* rates of transport are considered. If the concentration of a permeable solute is finite on both sides of a membrane, there will be, simultaneously, an *influx* of solute into the cell and an *efflux* of solute out of the cell. The *net flux* is obviously the difference between these two *unidirectional fluxes*. Now, in general, a flow or flux of material in any system implies a conjugate *driving force*. In thermodynamic terms the concentration gradient (ΔS of equation 2) represents the force driving the flow, J, in a diffusive process. The term PA/x is an example of a coupling coefficient that expresses the relationship between ΔS and J. A familiar example of an equation describing a linear relationship between a flow and its conjugate driving force is Ohm's law, $E = IR$, which relates the current or flow of electricity (I) between two points in a conductor to the driving force, E, the difference in electrical potential between these two points. The electrical resistance, R, is an expression of the coupling coefficient between E and I. Systems of linear equations of this kind form the basis of Onsager's analysis of the thermodynamics of nonequilibrium systems. In recent years, Onsager's theory has been extensively applied to the theoretical analysis of membrane transport (Katchalsky and Curran, 1965).

It is plain from equation 2 that the net flow in diffusive transport is always in the direction of the concentration gradient; i.e., net flow is *downhill* in an energetic sense. Hence, diffusion is an example of what is frequently called *passive* transport. When ΔS becomes zero, net flux is also zero. Clearly, this is the condition for *equilibrium* (zero net driving force) and occurs when the concentrations of the solute inside and outside the cell are equal. Note, however, that this is always true only for uncharged solutes. With charged species (i.e., ions), the condition that equilibrium corresponds to equal concentrations inside and outside the cell holds only if there is no electrical potential difference across the cell membrane, a situation that virtually never exists with living cells (except very transiently as in the *action potentials* of nerve and muscle cells—see Chaps. 2 and 3). The implications of this fact are explored in detail under Ion Transport and Bioelectric Potentials.

FACTORS INFLUENCING DIFFUSIVE TRANSPORT. As a result of extensive studies on the penetration of cell membranes by a variety of compounds, the German physiologist E. Overton proposed two generalizations that apply to substances that cross the membrane by diffusive mechanisms. Although they were formulated in the latter part of the nineteenth century, these generalizations (which have become familiar to generations of physiology students as "Overton's rules") have stood the test of time remarkably well.

Overton's first rule is that for predominantly nonpolar molecules *permeability is directly proportional to lipid solubility*. *Lipid solubility* is defined as the *partition* or *distribution coefficient* (K) for a given substance between water and a lipid or fat solvent (e.g., chloroform, benzene) that is immiscible with water. That is, $K = C_1/C_2$, where C_1 and C_2

are the equilibrium concentrations of the solute in question in the nonaqueous medium and in water, respectively. This relationship implies that, for predominantly nonpolar molecules, high lipid solubility is associated with high permeability, and vice versa.*

The role of lipid solubility in cell penetration is further underlined by the behavior of weak electrolytes, i.e., electrolytes that undergo *reversible dissociation* of the type

$$HA \rightleftharpoons H^+ + A^- \quad (3)$$

or

$$BH^+ \rightleftharpoons H^+ + B \quad (4)$$

within the physiological pH range. It is found that, although weak electrolytes (e.g., acetic acid) penetrate cells readily in the nonionized state, they fail to do so when *ionized*. This finding is in keeping with the fact that K values for these compounds are much greater in the nonionized than in the ionized state.

The implications for drug therapy are of considerable practical importance, since many drugs are weak electrolytes. Consider the absorption from the alimentary canal into the bloodstream of two orally administered drugs, one of which (A) dissociates according to equation 3, while the other (B) dissociates in conformity with equation 4. In any given situation the degree of ionization of both these drugs will be determined by the relationship between the pH of the surrounding medium and the pK (pH of half-dissociation) of the drug. For drug A the degree of ionization will decrease as the ambient pH decreases; the opposite will be true for drug B. Let us suppose that both our hypothetical drugs have a pK of 6 (and that neither is chemically altered by the enzymes of gastrointestinal juice). In the stomach, where the pH of the gastric juice is ordinarily about 2 to 3, virtually all of drug A will be in the nonionized form and thus will pass freely through the membranes of the gastric mucosa, whereas drug B will be almost completely ionized and will not be absorbed to any great extent during its passage through the stomach. On the other hand, further down the intestine, where the alkaline secretions of the pancreas increase the luminal pH to 7 or 8, drug B will be largely converted to the nonionized form and readily absorbed. In consequence, if both A and B are administered together, one would expect to find an effective level of A in the bloodstream before B appeared in quantity. Such considerations (in addition to intrinsic therapeutic potency, toxicity, etc.) underlie the many chemical variants of a basic therapeutic agent that are frequently available to the physician.

Overton's second rule is concerned with predominantly polar or hydrophilic molecules. It states that, for these molecules, permeability is *inversely related to molecular size;* i.e., the smaller the molecule, the more readily it penetrates the cell, and vice versa. A much fuller discussion of the role of lipid solubility, molecular size, and dissociation in the permeability of nonelectrolytes and weak electrolytes is given by Davson (1970).

The model of the cell membrane shown in Figure 1-1D is easily reconciled with Overton's rules. Since, in this model, the fluid matrix of the membrane is lipid in nature, one can easily see how lipid solubility can be the major determinant of permeability for nonpolar molecules. Similarly, the model readily permits the existence of discrete aqueous pores through which small, water-soluble molecules and ions (in addition to water itself) might be expected to move. As already mentioned, the most likely location of these pores is between the subunits of oligomeric intrinsic proteins. Current evidence indicates that the apparent average pore diameter in the membranes of a number of cell species is about 8 Å.

CARRIER-MEDIATED TRANSPORT. With many substances of physiological importance (e.g., sugars, amino acids), the experimental curve relating transport rate to the transmembrane concentration difference (ΔS) is frequently of the type shown in Figure 1-2B. At first, as ΔS is increased, the transport rate also increases. Finally, however, a value of ΔS is reached at which the transport rate becomes maximal and independent of further increases in ΔS (J_{max} in Fig. 1-2B). Nonlinear kinetic behavior of this type is referred to as *saturation kinetics* and is often exhibited by systems in which the kinetic process involves the reversible combination of the substrate with a receptor site (e.g., enzyme reactions).

In the case of membrane transport it is postulated that the transported solute or *substrate* enters the cell by combining reversibly with a specific membrane component or *carrier* at the outer surface of the membrane. The carrier-substrate complex so formed moves across to the inner surface of the membrane, where it dissociates, releasing free substrate to the cell interior. The carrier then moves back across the membrane to the outer boundary, where it combines with a second molecule of substrate, and the cycle begins again.

*Some discussion of the meaning of the term *high lipid solubility* in the context of cell physiology seems appropriate. Clearly, it must mean that K exceeds a certain critical value. A survey of the pertinent literature shows that Overton and his successors set this value at about 0.02. Hence, the word *high* in the present context is a very relative one indeed! Conversely, we may note that the term *low lipid solubility* is used by membrane physiologists to denote K values that are two to three orders of magnitude less than this.

Thus, a relatively small number of carrier molecules operating cyclically can transport large amounts of substrate. On this basis the occurrence of a maximum in the rate of transport with increasing substrate concentration is readily explained as being due to saturation of the available carrier sites when the external substrate concentration reaches a sufficiently high level. Clearly, an analogous explanation applies to net outward transport across the cell membrane.

With one simplifying assumption, the kinetic behavior of carrier-mediated transport systems can be quantitatively described by an equation identical to the well-known Michaelis-Menten equation of enzyme kinetics, i.e.,

$$J/J_{max} = \frac{[S]}{[S] + K_s} \tag{5}$$

In this equation, J is the transport rate at a given substrate concentration, [S], and K_s is analogous to the Michaelis-Menten constant, K_m; i.e., K_s is numerically equal to that substrate concentration for which $J = \frac{1}{2} J_{max}$.

The assumption previously mentioned is that, of the three steps involved in carrier-mediated transport, i.e., an *association* between substrate and carrier at one side of the membrane, a *translocation* of the carrier-substrate complex across the membrane, and a *dissociation* of this complex at the other side of the membrane, the *translocation* step is by far the slowest. With this assumption, $1/K_s$ gives a measure of the *affinity* of the carrier for the transported substance. Note that this affinity is *reciprocally* related to the magnitude of K_s.

In practice, this assumption is well justified for many membrane transport processes.* This has proved useful because, as is well known from enzyme studies, equation 5 can be used as the basis for a kinetic analysis of the action of specific inhibitors. Unlike diffusive processes, carrier-mediated transport is highly sensitive to these inhibitors. In terms of this analysis the action of competitive inhibitors is characterized by an increase in the apparent value of K_s with no change in J_{max}. Noncompetitive inhibitors lower J_{max} without affecting K_s. As shown in the insert of Figure 1-2B, K_s and J_{max} are easily determined from a plot of 1/J against 1/[S] for any given set of conditions.

A further application of this kind of kinetic analysis has been to show that substances that are closely related in chemical structure (e.g., hexose sugars, some amino acids, alkali metal cations) often compete for a common carrier in the cell membrane. This approach has been extremely useful in analyzing the different carrier-mediated transport pathways that exist in various cell types and in constructing *affinity* or *selectivity sequences* for the transport of chemically related species by specific membrane carriers. These are simply obtained by ranking, in descending order, the reciprocals of the K_s values for the transported species involved.†

TRANSPORT ENERGETICS: THE COUPLING OF TRANSPORT PROCESSES. In terms of energetics, we may note that while diffusive transport is *always* passive, carrier-mediated transport processes may be passive or active. The corollary to this is that *active transport always involves a membrane carrier*. As previously noted for diffusive transport, net passive (downhill) carrier-mediated transport occurs in the direction of the substrate concentration gradient and ceases when the substrate is at equilibrium across the membrane. For uncharged solutes this corresponds to the condition $\Delta S = 0$. At this point the substrate is transported at the same rate in both directions across the membrane (*exchange diffusion*). Equilibrating processes of this kind are often referred to as *facilitated diffusion*. They do not depend on metabolism and can, in fact, be studied in systems such as isolated membrane vesicles in which metabolic energy production is entirely absent.

Uphill carrier-mediated transport processes are of two kinds: those that directly utilize metabolic energy to effect solute transfer (primary active transport) and those in which uphill transfer of one solute species is achieved at the expense of the transmembrane energy gradient for another (secondary active transport).

Primary active transport processes ("pumps") appear to be less common than was formerly thought. In general, they appear to be restricted to systems that operate through specific membrane-bound ATPase, e.g., the Na^+-K^+ ATPase (the Na^+ pump) or the Ca^{2+} ATPase that is found in

*There are, of course, exceptions. Many of the complexities associated with enzyme kinetics and receptor binding studies have been noted also in transport kinetics. A number of these complexities are discussed in the article by Wilbrandt and Rosenberg (1961) cited at the end of this chapter.

†It is important to note that such characteristics as the existence of competition between structurally related species and of sensitivity to specific inhibitors may, in some cases, be the only criteria whereby carrier-mediated transport can be distinguished from diffusive movement. This occurs when one is constrained to work with concentrations of the transported solute that are far below its K_s. It will be clear from an inspection of Figure 1-2 that, under these conditions, the kinetics of carrier-mediated transport become virtually linear. A slightly embarrassing example of this is the transport of urea across the membrane of the human red blood cell. In the lecture hall and in the laboratory, this has been exhibited to generations of admiring students as the example, par excellence, of diffusive transport. There is now increasing evidence (Macey, 1979) that it is a carrier-mediated process! Charity compels one to assume that the reason why this escaped the notice of so many perspicacious professors for so many years lies in the enormous (in physiological terms) K_s involved (about 2 *M*).

the membranes of many cell species. Though many of the detailed molecular mechanisms involved are still unknown, these transport systems can directly utilize the energy liberated by the hydrolysis of the terminal phosphate bond of ATP to transport ions (Na^+, K^+, Ca^{2+}) against a net energy gradient.

The Na^+-K^+–ATPase pump is an example of a *coupled transport* process. To operate successfully, it requires that transmembrane transport of Na^+ and K^+ occur at the same time. Normally, Na^+ is moved (pumped) outward across the cell membrane in an energetically uphill direction (see p. 12). The movement of K^+ is inward and may or may not be uphill, depending on the cell species involved and the specific experimental situation. Coupled movement of two dissolved solutes in opposite directions is called *antiport*. Analogous movement, but in the same direction, is called *symport*.

A further point of interest about the Na^+-K^+ pump is the stoichiometry or *coupling ratio* between Na^+ and K^+. This is normally not unity and in most cases appears to be 3 Na^+/2 K^+. Since the moving species are electrically charged, it is clear that the operation of this pump results in *net charge transfer*, in an outward direction, across the cell membrane. This will tend to generate a transmembrane electrical potential difference, oriented so that the exterior of the membrane is positive with respect to the inside. Transport processes of this kind are called *electrogenic*, and there is indeed evidence that the Na^+-K^+ pump contributes to the resting membrane potential of some cell species. Alternatively, since a continuous transfer of electrical charge across a membrane is equivalent to a flow of current, transport processes of this kind are sometimes called *rheogenic*.

Secondary active transport is another example of *coupled transport*. In processes of this kind, *uphill* transport of one species is obligatorily coupled to the simultaneous downhill transport of another. The transmembrane energy gradient for the downhill part of the overall process serves as the energy source for its uphill component. Coupled processes of this kind are perhaps most easily understood by considering some specific examples, such as the four diagrammed in Figure 1-3. These are chosen from the wide variety of such processes that are known, or suspected, to occur in certain epithelial tissues, e.g., the renal proximal tubules, gallbladder, and small intestine. More specifically, they represent coupled transport systems that have been identified in the *apical* membranes of the absorptive cells in these tissues.

The first transapical coupled transport process shown in Figure 1-3 (reading from top to bottom on the left-hand side) is Na^+-Cl^-–coupled symport. This is the major entry route across the apical cell membrane for these ions in a number of epithelia. There is compelling evidence that the stoichiometry of this process is 1/1; i.e., no net charge is transported across the membrane by this mechanism, which is therefore *electroneutral*. It would not be expected to have any effect on the apical membrane potential. This prediction has been confirmed by direct experimentation (Garcia-Diaz and Armstrong, 1980).

The second process shown, Na^+-glucose symport, is an example of Na^+-coupled transport of a nonelectrolyte. Processes like this are of key importance in the reabsorption (or absorption) of organic metabolites by the kidney and intestine (see Chaps. 22 and 26) and are also found in other tissues (Schultz and Curran, 1970). Since one of the transported solutes is electrically charged and the other is not, it will be evident, regardless of the Na^+/glucose coupling ratio, that this transport process must be rheogenic. In fact, since the cell membrane potential is normally oriented so that the inside is negative with respect to the outside, one would predict that the inward flow of Na^+ ions induced by Na^+-coupled glucose entry should *depolarize* the cell membrane (i.e., reduce the magnitude of the electrical potential difference across it). This prediction was experimentally confirmed for the small intestine by White and Armstrong (1971).

The third process illustrated in Figure 1-3 is a Na^+-H^+ antiport or exchange. This has been shown to be electroneutral (i.e., one Na^+ ion exchanges for one H^+ ion). However, because H^+ ions produced as a result of cell metabolism can be "extracted" from the cell interior by Na^+-H^+ exchange, this mechanism is important, in some cells at least, in the regulation of intracellular pH (Roos and Boron, 1981).

Finally in Figure 1-3, a Cl^--HCO_3^- antiport is indicated. This again is usually considered to be a 1/1 electroneutral process. Since HCO_3^- ions play an important regulatory role with respect to intracellular pH (see Chap. 24), Cl^--HCO_3^- exchange is important in this context. This exchange, by removing HCO_3^- ions from the cell interior, would tend to acidify this region. Clearly, the Na^+-H^+ exchange already discussed would have the opposite effect.

Specific membrane inhibitors have been found for all the four coupled processes just discussed. The loop diuretics furosemide and bumetanide (so called because their primary target in the kidney is the loop of Henle) are potent inhibitors of Na^+-Cl^- symport. The diuretic amiloride inhibits Na^+-H^+ exchange (though this is *not* the primary physiological basis of its diuretic action). The glycoside phlorhizin combines avidly with the sugar-binding site of the Na^+-glucose symport but is not transported across the cell membrane. Finally, Cl^--HCO_3^- exchange is specifically in-

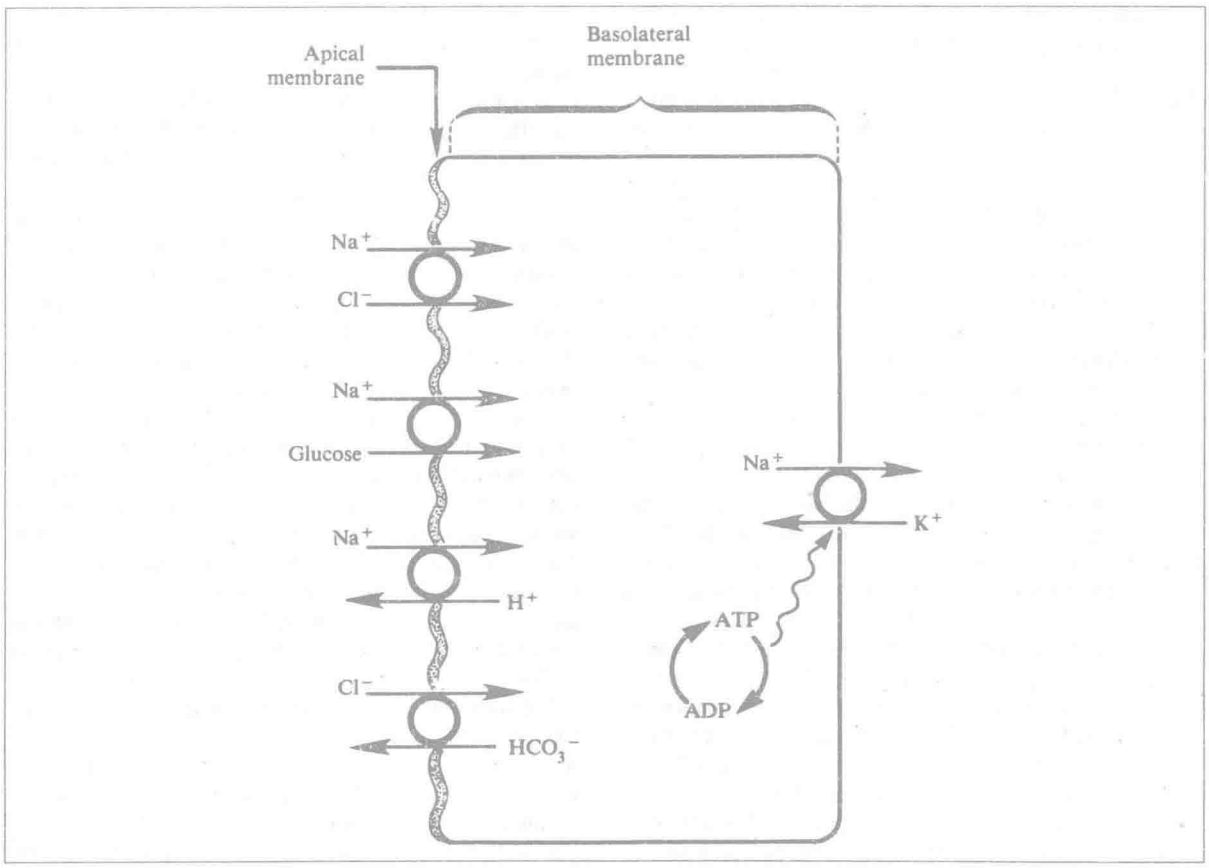

Fig. 1-3. Coupled transport processes in epithelial cells. *ATP* = adenosine triphosphate; *ADP* = adenosine diphosphate.

hibited by a series of substituted stilbene derivatives (known "in the trade" by such evocative acronyms as SITS and DIDS). Such inhibitors have proved extremely useful in identifying and characterizing the mechanisms involved in coupled transport.

In principle, none of these transport processes depends directly on metabolism. This is elegantly demonstrated by the fact that they can be shown to occur in isolated vesicles prepared from the apical membranes of cells from the kidney and intestine, among others. However, in vesicular systems, after their initial induction, the transport processes tend to run down; i.e., the net rate of transport declines with time. This is because the energy gradients that drive them are progressively dissipated. In living cells, metabolic processes serve to *conserve* these gradients and thus contribute indirectly to the continuous operation of gradient-dependent transport. As a reminder of this, the basolateral Na^+-K^+ pump is included in the schematic representation of an epithelial cell shown in Figure 1-3. By removing Na^+ continuously from the cell interior, this pump conserves the transapical Na^+ gradient, that is, the energy source for three of the coupled transfer processes shown.

The examples in Figure 1-3 are all representative of highly specific coupled transport mechanisms; that is, they involve membrane carriers that possess special combining sites for both transported solutes. Normally, these sites are highly selective. For example, in coupled processes that are sodium dependent, other cations, including alkali metal cations, cannot, as a rule, substitute effectively for Na^+. Similarly, the Na^+-glucose co-transport system illustrated in Figure 1-3 will accept only other hexose sugars (e.g., galactose) that share certain structural features with D-glucose (i.e., a pyranose ring with a carbon attached to C_5 and the OH on C_2 in the same stereochemical orientation).

In addition to these highly selective coupled transport mechanisms, there are other forms of coupling that are much less specific, well exemplified by the net osmotic movement of water that is induced by a transmembrane solute gradient (see p. 14). Such movement can occur even when water is, in chemical terms, at equilibrium across the membrane and is clearly driven by the solute gradient. In a purely mathematical and phenomenological context, this kind of coupling has been successfully analyzed by the methods of nonequilibrium thermodynamics (Katchalsky and Curran, 1965). In terms of specific molecular models it remains something of a puzzle. When, as frequently happens, water is one of the species transported, this kind of coupling is often called *frictional coupling*, the general idea being that energy transfer between the different molecular species that are moving across the membrane occurs by mechanisms analogous to mechanical friction. Although this has proved to be a useful approach for mathematical analysis, it is, for descriptive purposes, not entirely satisfactory.

WATER MOVEMENT: OSMOSIS AND OSMOTIC PRESSURE. Virtually all cells are freely permeable to water. The regulation of the water content of cells and tissues is an important factor in the maintenance of the overall *fluid and electrolyte balance* of the body (see Chap. 23), which is crucial in human health and disease. Indeed, one of the most striking contributions of physiology to modern clinical medicine lies in the sophisticated system of fluid and electrolyte therapy that has been and is still being developed to deal with imbalances in body water and electrolytes resulting from pathological conditions or from accidental or surgically induced trauma. It is therefore not surprising that the movement of water across cell membranes and the forces governing this movement have been of prime interest to physiologists for many years.

Osmotic Phenomena in Dilute Solutions. Consider a vessel divided into two parts by a membrane that is permeable to water but impermeable to dissolved solutes (a *semipermeable* membrane). If the compartment on one side of the membrane contains pure water and the other compartment contains an aqueous solution, water will flow spontaneously from the side containing pure water to the side containing the solution. This spontaneous net flow of water is called *osmosis*, or *osmotic flow*. Its origin may be explained as follows: As noted at the beginning of the chapter, all natural processes tend to proceed spontaneously in the direction of equilibrium. In the system under consideration, there are, initially, two concentration gradients across the membrane: (1) a solute concentration gradient from the side containing the solution to the side containing pure water and (2) because of the diluting effect of the solute on the water in the side containing the solution, a concentration gradient for water in the opposite direction. In the absence of external restraints, both water and solute would diffuse freely in the direction of their respective concentration gradients until mixing was complete. Because of the restraint imposed on the system by the semipermeable membrane, the solute cannot diffuse into the side containing pure water. The only net movement of material that can take place across the membrane is a flow of water into the side containing the solution. If, instead of pure water and a solution, two solutions containing unequal concentrations of solutes were used, a similar osmotic flow of water would take place. In this case, osmosis would occur from the more dilute to the more concentrated solution.

Clearly, unless some additional restraint is imposed, water will continue to move in the direction of its concentration gradient* as long as that gradient exists; i.e., until the concentrations of water on both sides of the membrane become equal. In the case of two solutions, equilibrium would be achieved when the solute concentrations on both sides of the membrane became equal. In a system containing pure water and an aqueous solution, osmotic equilibrium cannot, in principle at least, be realized in this way at any finite solute concentration. However, it is evident that if a force or pressure equal and opposite to the force generated by the concentration gradient for water across the membrane could be applied to the side containing the solution, osmotic flow could be prevented. The hydrostatic pressure (ΔP) required to prevent osmotic flow of water into a given solution is called the *osmotic pressure* ($\Delta\pi$) of that solution.

There are many ways in which osmotic flow can be prevented by the application of an external force. One of the simplest—and it also permits the osmotic pressure to be determined directly—is utilization of the force of gravity. The solution whose osmotic pressure is to be measured is placed inside a thin, semipermeable bag, which is then immersed in a vessel containing water. One end of a fine capillary tube is inserted into the bag, and the apparatus (called an *osmometer*) is adjusted so that the capillary tube is vertical. Initially, the apparatus is adjusted so that the liquid levels in the capillary and in the outer vessel, apart from the slight rise due to capillary action, are the same. As water

*This concentration gradient can then be said to be the driving force for osmotic water flow. More exactly, this force is related to the chemical potential gradient for water.

enters the bag by osmosis, the level of the liquid in the capillary rises until the hydrostatic pressure developed is just sufficient to balance the osmotic driving force across the wall of the bag. If the volume of the column of solution in the capillary is very small compared with the total volume enclosed in the bag (so that the total amount of water that enters the bag and its diluting effect on the solution within the bag are negligible), the osmotic pressure of the original solution is given by

$$\Delta\pi = \Delta P = h\rho g \tag{6}$$

where $\Delta\pi$ is the osmotic pressure difference across the membrane, h is the height of the column of solution in the capillary necessary to balance the osmotic driving force, ρ is the density of the solution, and g is the acceleration due to gravity. For moderately dilute solutions, ρ does not differ significantly from the density of water; hence, the height in centimeters of the column in the capillary will give the osmotic pressure directly in centimeters of water. Since 1034 cm of water is equivalent to 76 cm of mercury or 1 standard atmosphere, $\Delta\pi$ is readily obtained in either of these units.

Osmotic Pressure and Solute Concentration: Units of Osmotic Concentration. It is clear from the preceding discussion that the magnitude of the osmotic pressure in any given solution depends only on the difference between the concentration of water in that solution and its concentration in the pure liquid. This dependence, however, is of little practical use in physiology. Useful and meaningful relationships can be obtained if osmotic pressure is expressed in terms of solute concentration. Fortunately, despite the fact that the solute, aside from its diluting effect on the concentration of water in the solution, is not a fundamental factor in the generation of osmotic pressure, a simple relationsip between osmotic pressure and solute concentration does exist. The reason is that, in moderately dilute solutions, there is a simple complementary relationship between the concentration of water and that of solute.

Consider a solution containing n_1 moles of water and n_2 moles of solute per unit volume. The *molar fractions* of water and solute are respectively $X_1 = n_1/(n_1 + n_2)$ and $X_2 = n_2/(n_1 + n_2)$. Hence, $(X_1 + X_2) = 1$, and $X_2 = (1 - X_1)$. Evidently, it is the factor $(1 - X_1)$ that determines the osmotic pressure of a solution in contact with pure water. Therefore, the osmotic pressure is directly proportional to X_2 or solute concentration. For a dilute solution the quantitative relationship between osmotic pressure and solute concentration is given by the van't Hoff equation

$$\Delta\pi = CRT \tag{7}$$

where C is the solute concentration, R is the gas constant, and T is the absolute temperature.

It is apparent from equation 7 that, if the concentration of solute in a given solution is known, its osmotic pressure can readily be calculated. Conversely, given the osmotic pressure, the solute concentration can be obtained from this equation. It is important to realize that there is a difference between concentrations as they relate to osmotic activity and ordinary chemical concentrations. Since, as has already been pointed out, the solute has no intrinsic effect on osmotic pressure, the osmotic pressure of a solution is independent of the nature of the solute particles and depends only on their number per unit volume. In other words, in a given volume of solution, equal numbers of dissolved particles will contribute equally to osmotic pressure whether the particles are large molecules, small molecules, or ions. For this reason, the osmotic activities of solutions containing equal *chemical* concentrations of different solutes will not necessarily be identical. Consider, for example, two solutions, one containing 0.1 *M* sucrose and the other containing 0.1 *M* NaCl. Although the concentrations of these two solutions are equal in chemical terms (in moles/L), their osmotic pressures are not the same, because NaCl exists in solutions as Na^+ and Cl^- ions. Consequently, the osmotic pressure of a 0.1 *M* solution of NaCl is approximately twice that of a 0.1 *M* sucrose solution. Thus, it is apparent that, with electrolyte solutions, if one wishes to relate the osmotic pressure or effective osmotic concentration of solute to its chemical concentration, one must multiply the term C in equation 7 by n, where n is the number of ions produced by one molecule of electrolyte and is sometimes referred to as the van't Hoff coefficient.

The situation is further complicated by the fact that this simple relationship between the number of ions formed by an electrolyte and its osmotic activity holds only for very dilute solutions. Because of the attractive forces between ions of opposite charge and between individual ions and water molecules, the apparent value of n for a given electrolyte varies with concentration. Also, the concentration dependence of n is different for different electrolytes. At physiological concentrations the divergence of n from its limiting value in dilute solution is sufficiently great to affect appreciably the accuracy of results calculated on the basis of equation 7. Further, the physiologist is frequently confronted with solutions, such as blood or urine, that contain complex mixtures of solutes, both electrolytes and nonelectrolytes, in widely different concentrations. In this situation the need for a practical unit of osmotic concentrations that is independent of the apparent variation of n with concentration for different individual solutes will readily be ap-