Chapter 2
Overview of Controlled Release Mechanisms

Ronald A. Siegel and Michael J. Rathbone

Abstract  Controlled release systems have been developed to improve the temporal and spatial presentation of drug in the body, to protect drug from physiological degradation or elimination, to improve patient compliance, and to enhance quality control in manufacturing of drug products. When designing controlled-release systems, it is important to identify and understand particular mechanisms involved in the release process. Often, more than one mechanism is involved at a given time or different mechanisms may dominate at different stages of the drug delivery process. This chapter begins with several vignettes, each highlighting a mode of controlled drug delivery and identifying associated mechanisms. An introductory description of several of the mechanisms follows. Details regarding these mechanisms are provided in subsequent chapters.

2.1 Introduction

Controlled-release systems are designed to enhance drug therapy. There are several motivations for developing controlled-release systems, which may depend on the drug of interest. Controlled release systems have been devised to enable superior control of drug exposure over time, to assist drug in crossing physiological barriers, to shield drug from premature elimination, and to shepherd drug to the desired site of action while minimizing drug exposure elsewhere in the body. Controlled release systems...
may also increase patient compliance by reducing frequency of administration, and may add commercial value to marketed drugs by extending patent protection. Finally, use of controlled release technology may reduce variability of performance of drug products. The latter aspect is increasingly important given the current emphasis on “quality by design” by regulatory agencies such as FDA.

The mechanisms used to achieve these goals are diverse and complex, and depend on the particular application. In fact, several mechanisms may operate simultaneously or at different stages of a delivery process. An understanding of these mechanisms is important when designing and manufacturing controlled-release systems, and in identifying potential failure modes. Delineation of mechanism is also important in intellectual property prosecution and quality assurance/quality control.

This chapter starts with a series of vignettes illustrating mechanisms and their interplay in particular controlled release systems. Essentials of individual mechanisms are then outlined. More elaborate descriptions are deferred to later chapters.

## 2.2 Vignettes

### 2.2.1 Zero Order Oral Delivery

Zero order, or constant rate release of drug is desirable in order to minimize swings in drug concentration in the blood. Such excursions, which may lead to periods of underexposure or overexposure, are particularly likely to occur for drugs that are rapidly absorbed and rapidly eliminated. Figure 2.1 illustrates the plasma concentration profile over time for such drugs when administered from rapid-release dosage forms. A rapid increase in concentration is followed by a rapid decrease, and little time is spent inside the so-called therapeutic range, which is bounded below by a minimum effective concentration (MEC) and above by a minimum toxic concentration (MTC) (see also Figs. 1.9 and 1.10). Frequent repetitive dosing is required to maintain concentration within these limits, and compliance and control are difficult.

Dosage forms that prolong release can maintain drug concentration within the therapeutic range for extended periods and minimize episodes of underexposure or toxicity. A well designed system displays a narrow, predictable residence time distribution in the gastrointestinal (GI) tract, and releases drug by a controlled mechanism. As shown in Fig. 2.1, zero order release leads, in principle, to the best control of plasma concentration. Such control leads to constant drug effect, provided the drug’s pharmacokinetic and pharmacodynamic properties, including absorption, distribution, metabolism, and excretion (ADME), and its pharmacodynamic properties relating plasma concentration to drug effect, are stationary. While this proviso is believed to apply to most drugs, there are notable exceptions, as detailed in Chap. 13.
Zero order oral drug release can be achieved, in principle, by surrounding a core tablet with a membrane that is permeable to both drug and water, as illustrated in Fig. 2.2a. After swallowing, the core becomes hydrated, and drug dissolves until it reaches its saturation concentration or solubility. The core serves as a saturated reservoir of drug. Drug release proceeds by partitioning from the reservoir into the membrane, followed by diffusion across the membrane into the gastrointestinal fluid. So long as saturation is maintained in the core, there will be a stationary concentration gradient across the membrane, and release will proceed at constant rate. Eventually, the dissolved drug’s concentration in the core falls below saturation, reducing the concentration gradient and hence the release rate, which decays to zero.

If the membrane consists of a water-soluble polymer of high molecular weight, then it will initially swell into a gel, through which drug diffuses. The thickness of the gel layer initially increases with time due to swelling, but ultimately it decreases due to disentanglement and dissolution of polymer chains. At intermediate times, the gel layer may be of approximately constant thickness, and release occurs at a relatively constant rate.

As an alternative to dissolution/partition/diffusion based devices, osmotic pumps have been developed to provide zero order release. An elementary osmotic pump, illustrated in Fig. 2.2b, is a tablet or capsule consisting of a core of drug surrounded by a membrane that is permeable to water but not to the drug. A small hole is drilled into the membrane. Upon ingestion, water is osmotically imbibed into the core through the semipermeable membrane, dissolving the drug. A constant osmotic pressure gradient is established between core and the external medium.
setting the stage for water influx, which displaces drug through the hole at a
constant rate. Eventually, drug concentration falls below its solubility, and the
rate of osmotic pumping decays.

The efficiency of osmotic devices can be improved by enriching the core with
excipients such as water soluble polymers. For example, in push-pull osmotic
systems, depicted in Fig. 2.2c, the drug formulation is layered between the water-
soluble polymer and the exit orifice. Osmotic flow into polymer pushes dissolved drug through the orifice (fat arrows)

2.2.2 Oral Delivery Directed to the Gut and Colon

Numerous drugs are susceptible to hydrolysis in the acidic environment of the
stomach. Enteric coatings, which are pH-sensitive polymers that are insoluble in
acid but dissolve in the neutral or slightly alkaline environment of the gut,
are designed to protect drug as it passes through the stomach. If the molecular weight of the coating polymer is relatively low, then it will dissolve and drug will be released rapidly. If the molecular weight of the polymer is high enough, however, it will swell into a gel layer that controls drug release as above. Passage of the dosage form through the stomach to the small intestine affects the time required following ingestion to activate swelling and diffusion.

Certain drugs are more efficacious when released in the colon. The colon is rich in bacterial azoreductases, which cleave polymers with azoaromatic crosslinks. By encapsulating drug in such polymers, colon-specific drug delivery can be achieved. Further encapsulation by a rapidly dissolving enteric coating would permit colon-specific delivery of acid-labile drugs. The enteric coating is first stripped off upon entering the gut, but drug is released only when the internal polymer is degraded by the azoreductases in the colon.

### 2.2.3 Oral Delivery of Polypeptides

Polypeptides, including proteins, are extremely challenging to deliver orally. Problems include acid lability, susceptibility to peptidases and proteases in the stomach and gut, and limited absorption due to high molecular weight and charge. Most protein bioavailabilities, measured as fraction absorbed into the systemic circulation, hover around or below 1%. Reliable, efficient delivery of polypeptides, if possible, will have enormous payoffs.

Let us assume that acid lability can be handled by an enteric coating layer and that the polypeptide is incorporated into micro- or nanoparticles that are designed to adhere to the gut wall. The particles release their payload into the wall or are taken up by endocytosis into enterocytes. While encapsulated in the particles, the polypeptide molecules are protected from attack by enzymes. By these means, it is postulated that bioavailability will be improved.

### 2.2.4 Delivery of Drugs Through the Skin

Numerous drugs are problematic for oral delivery due to their low solubility and susceptibility to first pass metabolism in the liver. For such drugs, alternative ports of entry are of interest, and practically every available body surface and orifice has been considered. Since the skin is readily accessible and has a large surface area, transdermal drug delivery has been the subject of much research and product development.

The primary barrier layer of skin is the stratum corneum, a thin layer of dead squamous cells that are packed in a kind of brick and mortar configuration, as depicted in Fig. 2.3, with specialized lipids serving as the mortar. Lipophilic drugs can readily dissolve in this layer and diffuse through it at a rate that depends primarily on molecular size and lipophilicity. Very little drug enters the dead
cells, and the lipid pathways for diffusion are marked by numerous detours. After passing through the stratum corneum, drug encounters the more hydrophilic, viable epidermis and dermis, before being absorbed in capillaries perfusing the dermis. Drug that is absorbed through the skin is not susceptible to first pass metabolism by gut and liver, although some metabolism may occur in the skin itself.

While ointments and creams are usually used for topical delivery to the skin, patches have been developed for controlled systemic delivery. The simplest patch consists of an adhesive layer containing drug in the dissolved or in a finely divided solid form, and an impermeable backing layer, as illustrated in Fig. 2.4a. For such patches, delivery rate is controlled primarily by the permeability of the stratum corneum, which depends on the drug’s partition coefficient between the patch material and the stratum corneum, the drug’s diffusivity in the stratum corneum, and the thickness of the stratum corneum. Provided these parameters remain constant during application of the patch, and if drug activity in the patch remains constant by dissolution of solid drug into the adhesive, then zero order, constant rate delivery can be achieved.

The simple adhesive patch design is best for drugs with a large therapeutic range, since skin permeabilities may vary across patients and between sites of application in an individual patient. When more precise control of drug concentration in blood is desired, it is useful to insert a rate controlling membrane between the drug reservoir and the adhesive layer, as shown in Fig. 2.4b. The membrane’s permeability must be less than that of the skin in order to provide effective rate control.

Since the skin naturally functions as an environmental barrier, only a few drugs can penetrate it at an adequate rate by partitioning and diffusion. Generally, a drug molecule should be sufficiently lipophilic that it partitions into the stratum corneum, but sufficiently hydrophilic that it can also cross the viable layers. Its molecular weight should be low to ensure adequate mobility in the stratum corneum.
Finally, the drug’s potency and pharmacokinetic properties should be such that delivery through the skin places drug concentration in plasma within the therapeutic range. While the rate of delivery can be increased by using larger patches, there are practical size limitations.

Because the skin is so accessible, much effort has been devoted to expanding the spectrum of transdermally deliverable drugs using more complex delivery systems. For example, the skin’s barrier function can be disrupted temporarily by applying chemical permeation enhancers, microneedles, ultrasound, heat, or short, high voltage bursts of electricity (electroporation). During or immediately following disruption, drug can be administered. Alternatively, drugs can be delivered by iontophoresis, in which a steady electrical current is applied through the skin, as illustrated in Fig. 2.4c. This process relies on aqueous channels in hair follicles and

Fig. 2.4 Various transdermal patch designs (a) skin permeability control, (b) membrane control, (c) iontophoretic patch design for a cationic drug.
sweat glands, or new channels formed by the current. Charged drug molecules are driven through these channels by a like-charged electrode while uncharged drug molecules are delivered through the channels by electroosmotic convection.

Ideas discussed in this vignette may apply to drug delivery across other well perfused epithelia, including the rectum, vagina, scrotum, cornea and sclera, and the buccal and nasal mucosae.

2.2.5 Depot Delivery of Reproductive Hormones

While the introduction of daily oral steroid contraceptives in the mid-twentieth century was a breakthrough with historic medical and social consequences, it is recognized that there is substantial room for improvement. Daily oral dosing can lead to incomplete compliance and effectiveness, so other routes have been studied. For example, a transdermal, patch-based contraceptive system that delivers its payload over 1 week has appeared on the market (ORTHO EVRA®), as has an insertable vaginal ring that releases drug over three weeks (NuvaRing®).

The Norplant® system was introduced in the 1980s to provide five years continuous release of levonorgestrel. Drug is incorporated into silicone capsules that are placed under the skin in a routine clinical procedure. Release is mediated by slow diffusion through the silicone matrix. Because the silicone capsules do not degrade, they must be retrieved after they are spent. An alternative biodegradable implant called Capronor was investigated but was not marketed.

Besides steroid hormones, analogs of luteinizing hormone-releasing hormone (LHRH) have been developed. LHRH is the master hormone that is secreted rhythmically in the hypothalamus, and activates numerous hormones on the reproductive axis. Both LHRH agonists and antagonists have been developed as contraceptives, and they also have been used to treat disorders, such as endometriosis, vaginal bleeding due to fibroids, precocious puberty, and prostate cancer. When these analogs are delivered continuously, they interfere with the rhythmic signaling by endogenous LHRH. Because they are extremely potent, they can be injected as a slow-release depot. In one system, Leupron Depot®, leuprolide acetate is formulated into biodegradable polymer microspheres, which degrade and release drug over three months. In this system, drug release is controlled by diffusion through a pore network whose structure evolves as the polymer degrades.

Osmotic pumping provides another potential approach to long-term contraceptive delivery. One example is a narrow metal cylinder containing two compartments that are separated by a movable piston, as shown in Fig. 2.5. The drug formulation is introduced into one compartment, which is capped on the end, except for a small exit orifice. The other compartment contains an osmotically active agent, and is capped by a membrane that is permeable to water but not to that agent. Osmotic water flow across the membrane displaces the piston, and drug is pushed out through the exit orifice. By proper selection of the semipermeable membrane, the pumping rate and hence duration of release can be precisely controlled.
Complementary to contraception is fertility therapy. Patients with lesions that suppress LHRH secretion can be treated with rhythmic intravenous injections of LHRH, delivered from an externally worn, programmed pump through a catheter. This mode is best for short term needs, such as induction of fertility, but it is less desirable when the need is long term, as in the treatment of arrested puberty. Since LHRH is exceptionally potent, each dose is very small, so the possibility of an implantable rhythmic dosing device is intriguing. Such devices may ameliorate the inconvenience associated with intravenous delivery. One approach under consideration is a controlled-release microchip, into which thousands of microwells are machined. Each well is filled with a single dose of LHRH and sealed by a thin gold membrane that is addressably connected to a current source. Under the control of a microprocessor, individual membranes are ruptured with a current pulse and their encapsulated doses are released. By proper programming, any sequence of release pulses can be programmed into the system.

2.2.6 Regional Drug Delivery

Thus far, we have discussed scenarios in which drug enters the systemic circulation after release. Drug then distributes according to its relative affinities to all tissues, and only a small fraction is present at or near the target site. Drug toxicity and side effects are often associated with accumulation in tissues not associated with the target. In regional (sometimes called local or topical) delivery, drug is administered directly to the target tissues. Under proper conditions, regional delivery should permit substantially reduced drug dosing to reach the desired effect, with reduced exposure of other tissues to the drug.

Regional delivery is potentially most effective when drug is not transferred substantially from the target tissue to the systemic circulation due to anatomic or physiological barriers, or when systemic drug is rapidly eliminated. Traditional examples include topical drugs, inhalation based asthma therapies, and chemotherapies directed by drug pumps to tumors. The release of chemotherapeutic agents from polymer disks implanted next to brain tumors provides another example, as does insulin delivery to the peritoneal cavity, which drains through the hepatic portal vein into the liver, a primary target for insulin.
In this vignette, we first consider local delivery to the eye, noting different strategies that must be applied to delivery into the aqueous humor and the retina. We then discuss drug-eluting stents, which provide local delivery of drugs to arteries following injury.

Anatomical features and routes for drug delivery to the eye are shown in Fig. 2.6. The eye cavity is a useful port of entry for antibiotics and drugs meant to treat disorders in tissues perfused by tears and aqueous humor. To reach the aqueous humor, which lies under the cornea and houses the lens and iris, drug must cross the cornea, which contains both lipophilic and hydrophilic layers. Conventional eye drops are notoriously inefficient, since much of the drop is lost by overflow and drainage into the nasolacrimal duct. To increase drug retention in the eye cavity and hence bioavailability, drug can be formulated in gels that spread over and adhere to the ocular surface. Alternatively, drug-soaked contact lenses have been considered for topical delivery. In addition to increasing bioavailability, these formulations may prolong the release process, reducing the required frequency of administration.

An early drug delivery product was Ocusert®, in which a saturated pilocarpine reservoir was placed between two membranes which could control release, by the partition/diffusion mechanism, of the drug for up to 1 week. This product was placed under the lower eyelid and released drug at constant rate into the tear fluid, with subsequent absorption through the cornea.

Drug administered into the eye cavity is generally not available to the retina. To reach the retina, drug can be delivered to the vitreous humor, which lies behind...
the lens. The vitreous is a viscous gel that slowly circulates, providing convective transport of drug to the retinal surface. Several schemes have been investigated. The ocular sclera (white) provides a large surface area, and patches have been devised for transscleral delivery into the vitreous. A problem arises because the sclera is heavily perfused by choroidal blood vessels, which remove drug before it reaches the vitreous by diffusion. Alternatively, solid implants that slowly release the drug can be injected or placed surgically into the vitreous. Presentation of drug to the retina, then, depends both on the rate of release by diffusion and rate of convection to the retinal surface.

Drug eluting stents, discussed in Chap. 14, have recently been developed to prevent restenosis or reclosing of coronary arteries following angioplasty and stenting procedures in response to heart attacks. Restenosis is an inflammatory response to these procedures, and involves the growth of arterial smooth muscle cells over the stents. To arrest such growth, small amounts of anti-inflammatory and antiproliferative drugs are coated onto the stents and are released directly into the adjacent arterial tissue by dissolution, partitioning, and diffusion. Because the dose is so small and targeting is so precise, it is possible to prevent restenosis without releasing detectable amounts of drug into the systemic circulation and other tissues.

### 2.2.7 Nanoparticulate Targeting of Drugs to Specific Tissues

Besides improving systemic bioavailability and the temporal and regional patterns of drug release and absorption, controlled release systems have been developed to alter the residence time of circulating drug. In these systems, drug is incorporated in nanocarriers that have access to the whole systemic circulation, but are cleared less rapidly than free drug. The nanocarriers can be regarded as circulating drug depots. Nanocarriers may also have favorable distribution properties into target tissues and away from tissues associated with toxic side effects. Examples of nanocarriers include microemulsions, liposomes, dendrimers, block polymer micelles, solid lipid and polymer nanoparticles, and soluble polymers with drug attached on side chains by biodegradable linkages.

At the nano level, it is also possible to incorporate targeting ligands that permit particles to bind preferentially to specific cell types and promote the uptake and drug release into those cells. It has been suggested that cellular processes that rely on multivalent attachment, including particle uptake, can be modulated by drug/nanoparticle composites by suitable placement of multiple-targeting ligands on particle surfaces.

Design of nanoparticulate drug delivery systems must take into account normal physiological scavenging processes that remove small foreign objects from the blood. Special coatings, such as poly(ethylene oxide)s, are used for this purpose. Suitably coated nanoparticulates exhibit reduced opsonization and clearance by the reticuloendothelial system. Renal clearance is avoided when nanoparticulates are larger than glomerular pores. Hence, circulating half-lives of nanoparticulates and
their associated drugs are prolonged. Furthermore, coated nanoparticles and their associated drug are largely restricted to the vascular space, in contrast to free drug which may have much a larger volume of distribution. It should be noted, however, that if drug is released from the nanoparticle into systemic circulation, as opposed to a specific target site, it will possess the same pharmacokinetic properties as otherwise administered free drug.

It is believed that nanoparticulate delivery systems may be very useful in treating some cancers due to the enhanced permeation and retention (EPR) effect. Compared to normal tissues, tumors have leaky capillaries with large fenestrations in the capillary walls that permit the passage of nanoparticulates. Drug loaded into the nanoparticulates is, therefore, relatively more accessible to tumor tissues compared to tissues associated with toxic side effects.

### 2.3 Survey of Mechanisms

The previous vignettes highlighted several controlled-release mechanisms, including dissolution, partitioning, diffusion, osmosis, swelling, erosion, and targeting. Basic principles associated with these mechanisms are presented in this section.

#### 2.3.1 Dissolution

Most drug molecules form crystals at room temperature. In fact, they may take on various crystal forms (polymorphs) or form crystal hydrates, depending on their processing conditions. In some cases drug particles can be processed into an amorphous, glassy form. These forms have differing thermodynamic stabilities, and interconversion between solid forms can occur during storage and after administration. Dissolution involves transfer of drug from its solid phase to the surrounding medium, which may be water, polymer, or tissue. The solubility of drug in a medium, \( C_{S,\text{medium}} \), is defined as the concentration of drug in the medium at saturation, i.e., in equilibrium with the solid form. Higher concentrations of drug are thermodynamically unstable, and with time drug crystallizes out of solution until its concentration equals \( C_{S,\text{medium}} \). Useful rules of thumb are that \( C_{S,\text{medium}} \) decreases with increasing melting point of the drug and increases with increasing chemical compatibility of drug with the surrounding medium.

While solubility is a thermodynamic property of a drug and a medium, the dissolution rate is a kinetic property. Dissolution rate increases with solubility and decreases with drug particle size. As discussed below, dissolution rate is commonly controlled by diffusion.


2.3.2 Partitioning

During drug delivery, drug molecules often encounter an interface between two materials or phases. The partition coefficient is a measure of the relative affinity for drug between the two phases, and is roughly given by the ratio of drug solubilities in the two phases. At the interface, the partition coefficient prescribes the relative frequency that a molecule moves into one medium compared to the other.

As an example, recall that drugs of high lipid solubility are suitable for entry into the stratum corneum. However, if the drug is not sufficiently water soluble, i.e., its lipid/water partition coefficient is too high, it will not partition efficiently into the viable epidermis, and drug will be detained in the stratum corneum. Absorption into capillaries might then occur at an unacceptably low rate.

As a second example, block copolymer micelles are formulated with hydrophobic cores and hydrophilic coronas, hence they are soluble in blood. Hydrophobic drugs preferentially partition into the core, where they are retained for extended periods of time. Pharmacokinetic characteristics of such drugs, i.e., clearance and volume of distribution, reflect those of the micelles, leading to longer retention in the circulation and preferred distribution into tumors due to the EPR effect.

2.3.3 Diffusion

Diffusion is a very important component of many controlled-release systems, hence we devote considerable space in this chapter to it. More details about diffusion-controlled drug delivery systems are provided in Chaps. 6 and 9.

2.3.3.1 Molecular Basis

All molecules constantly undergo random collisions with other molecules. As a result, molecules execute thermal or Brownian motion. At any step, the direction of motion of a molecule is random, and it repeatedly changes due to collisions with other molecules. Over time, the displacement of the molecule from its point of origin is the result of a multitude of such random steps. Macroscopically, the independent random walks taken by large number of drug molecules lead them from regions of higher concentration to regions of lower concentration. Thus diffusion of a substance occurs down its concentration gradient.

The theory of random walks shows that the average (actually, root mean squared) distance that molecules travel by diffusion is proportional to the square root of time, i.e., average distance traveled \( \sim \sqrt{Dt} \), where \( D \) (cm\(^2\)/s) is the diffusion coefficient, or diffusivity, and \( t \) is time (s). The diffusion coefficient is a measure of the molecule’s mobility in the medium. Conversely, the typical time required to diffuse over a particular distance is proportional to the square of that distance and
inversely proportional to the diffusion coefficient. Thus, while diffusion is an efficient means of mass transport over short distances, its effectiveness decreases over longer distances.

Figure 2.7 illustrates and contrasts partitioning and diffusion. Two media are placed next to each other. Within each medium, symmetric arrows depict the magnitude of the diffusion coefficient, which characterizes the motion of the molecule exclusively inside that medium. A molecule moves in either direction with equal probability. At the interface between the media, however, the molecule must make a choice. The partition coefficient determines the relative frequencies that this molecule “jumps” into either medium. The two different frequencies are depicted by arrows of different thicknesses. The lengths of the arrows correspond to the respective diffusion coefficients.

2.3.3.2 Reservoir Versus Monolithic Systems

We have already introduced systems in which a membrane mediates diffusion from a reservoir. In reservoir systems, drug first partitions into the membrane from the reservoir and then diffuses to the other side of the membrane, where it is taken up by the receiving medium. While the reservoir is saturated, a constant concentration gradient of drug is maintained in the membrane, the rate of drug flux is constant, and zero order release is achieved. Eventually, drug concentration in the reservoir falls below saturation, and the gradient across the membrane and release rate both decay.

In reservoir systems, the purpose of the membrane is to mediate diffusion of drug. Because of their simplicity of mechanism and their ability to produce zero order
release, reservoir systems would seem to be highly advantageous. However, reservoir systems can be difficult to fabricate reliably. Pinhole defects and cracks in the membrane can lead to dose dumping. These problems are avoided in monolithic systems, in which drug is loaded directly into a polymer, which now acts as both a storage medium and a mediator of diffusion.

Drug is typically loaded uniformly into monolithic devices, and release is controlled by diffusion through the monolith’s matrix material or through aqueous pores. Monolithic devices typically exhibit an initial burst of release from the surface. With passing time, release rate decreases as drug that is deeper inside the monolith must diffuse to the surface, since it has farther to travel, and the quadratic relation between distance and time becomes important. This effect occurs in planar monoliths, but it is even more prominent with cylinders or spheres, as the amount of drug available decreases with distance from the surface. This geometric factor can be substantially reversed using specially coated wedge, cone, or hemisphere monoliths to provide near-zero-order release, but such devices are not easy to fabricate.

### 2.3.3.3 Factors Affecting Diffusivity

The diffusivity, \( D \), depends on the molecule and the medium. For a hard spherical molecule in a liquid solvent, the Stokes-Einstein equation prescribes

\[
D = \frac{k_B T}{6 \pi \eta a},
\]

where \( a \) is the molecule’s radius, \( \eta \) is the solvent’s viscosity, \( k_B \) is Boltzmann’s constant, and \( T \) is absolute (Kelvin) temperature. This relation confirms the intuition that large molecules should diffuse more slowly than small ones and that diffusion should be slowed in viscous liquids. The factor \( k_B T \) accounts for the intensity of thermal agitation, which drives Brownian motion.

In typical polymeric controlled release systems, the polymer matrix does not flow like a liquid, and bulk viscosity is not the correct parameter to use in predicting mobility of drug. The matrix may possess, however, a “microviscosity” that is related to molecular mobility. Free volume theory provides a useful picture that accounts for both bulk and microviscosity. While it may be natural to think of a polymer matrix as a static solid, it is actually a dynamic fluctuating structure, and \( D \) may be thought of as a measure of the degree that these fluctuations accommodate random motion of the diffusing molecule. In free volume theory, each drug, solvent, and polymer molecule contains an impenetrable core that is surrounded by nanovoids, called free volume. Thermal motions cause the size of voids to fluctuate. Occasionally, a void becomes large enough for a diffusing molecule to move into or through it. Clearly, if this mechanism is operative, then the diffusion coefficient will decrease sharply with increasing molecular radius and when the matrix’s density increases upon cooling. At a critical density, often associated with the medium’s glass transition temperature, \( T_g \), free volume becomes so sparse that the diffusion coefficient drops by several orders of magnitude.

In addition to temperature, the free volume of a polymer matrix depends on its composition. For homogeneous materials, free volume increases as the difference
between ambient temperature and $T_g$ increases. Copolymerization and blending can lead to matrices with suitably averaged free volumes and mobility properties. Free volume can also be increased substantially by sorption of small molecules, such as water. Thus, a glassy dry polymer can be converted to the rubbery state by sorption of a small amount of water, substantially increasing the mobility of drug molecules in the polymer.

Besides the glass transition, polymers can form crystalline domains which exclude drug molecules and obstruct diffusion. The propensity to crystallize depends on the polymer’s melting point and its stereoregularity. Random copolymers generally do not form crystalline domains. Crystallization can be mediated by the polymer backbone or by the side chains, especially when the latter are long.

For a molecule diffusing through a water-swollen hydrogel, diffusivity of drug is affected by the viscosity of the water space and also by obstructions placed in the drug molecule’s path by the hydrogel chains. Many models of diffusion in hydrogels, therefore, combine elements of Stokes–Einstein and free volume theories. In this case, the size of water-filled spaces between hydrogel chains is assumed to fluctuate, making room for movement of the diffusing drug molecule. The characteristic distance between points of chain crossings in the hydrogel is called the correlation length, and the ratio of molecular radius of drug to the correlation length is considered to be the primary structural parameter governing the drug’s diffusion coefficient in the hydrogel.

### 2.3.3.4 Heterogeneous Systems

Thus far, we have discussed diffusion mediated systems in which the medium is a uniform polymer matrix or hydrogel. Local matrix fluctuations were assumed to control the rate of diffusion. In more heterogeneous media, other factors also become important.

We have already noted that the presence of dead cell bodies in the stratum corneum increases the effective path length for drugs diffusing through skin lipids. We have also seen that crystalline domains in a polymer can obstruct and retard diffusion. More generally, diffusion of drug through a heterogeneous medium depends on the solubility and diffusivity of drug in the different material domains of the medium, and the geometric manner in which the domains are dispersed.

For example, consider a polymer blend or block structure, where one component has a much higher drug solubility than the other. If the “drug-philic” domains comprise a discrete phase dispersed in a “drug-phobic” continuous phase, then the disconnected phases will retain drug and retard its release, by analogy to affinity chromatography. If on the other hand the drug-philic domain is continuous, then release will be controlled by diffusion through the continuous phase, but will be retarded by detours around the drug-phobic domains.
Porous systems are often encountered in controlled release. Empty pores can be introduced into a matrix during fabrication to serve as pathways for drug diffusion through water that enters the pores. Alternatively, solid drug or excipient particles can be introduced into a polymer, and pores form around the particles. Also drug and excipient may precipitate from a polymer solution during solvent removal, again resulting in a porous amalgam of drug and polymer. The pores then act as both depots for drug storage and as conduits for diffusion. Pore structure and connectivity may have a profound effect on release by diffusion, as is discussed in Chap. 9.

2.3.3.5 Diffusion Affects Dissolution

We conclude this section with a discussion of dissolution and diffusion in drug delivery. Dissolution occurs when the solvating medium surrounding a solid drug particle is not saturated. This process involves two steps. First, drug must dissociate from the surface of the particle and surround itself with solvent. Second, the newly solvated drug must diffuse away from the surface. The first process is usually more rapid than the second, unless the drug is extremely insoluble. Thus, the drug is very close to its saturation concentration in the immediate vicinity of the particle. A concentration gradient is, therefore, established between the particle/medium interface and the “bulk” of the medium, and diffusion controls the rate that drug flows down this gradient. In drug delivery systems containing solid drug particles, both $C_{S,\text{medium}}$ and $D$ are therefore important determinants of release rate.

In an important class of drug delivery systems discussed in Chap. 6, solid drug particles are incorporated into a monolithic matrix. Release of drug occurs by dissolution followed by diffusion through the matrix. Particles at the surface dissolve quickly, leading to a burst. Particles further inside dissolve more slowly, since dissolution rate is controlled by diffusion through the matrix. At intermediate times, a moving front is observed, separating a central core containing solid drug from a periphery containing completely dissolved drug. Because the diffusion distance from the front to the monolith’s surface increases with time, the march of this front slows down as the release process proceeds, and the rate of release decreases with time.

2.3.4 Osmosis

Osmosis is a dramatic phenomenon that occurs when a membrane that is permeable to water but not to particular solutes, called osmolytes, separates aqueous solutions of the osmolytes. Water flows through the semipermeable membrane in an effort to equalize concentrations of the impermeable solutes on both sides of the membrane. In most cases of interest, water flow occurs by diffusion through the semipermeable membrane. However, the nature of water transport may differ from that discussed
above for drugs. First, it should be emphasized that there tends to be a lot of water on both sides of the membrane, and flux of water through the membrane is determined by the difference in chemical potentials of water on the two sides, not simply the concentration gradient of water. These chemical potentials may depend on both concentrations of the osmolytes and the thermodynamic compatibility of water with the osmolytes. When the osmolytes are small molecules, such as salts, osmotic pressure is reasonably accounted for osmolyte concentrations according to van’t Hoff’s law, but when the osmolytes are polymers, osmotic pressure is determined jointly by polymer concentration and polymer/water compatibility. Second, when the membrane is adequately hydrated, water molecules are in contact with each other and neighboring molecules’ motions are correlated. The Brownian mode of diffusion discussed for drug molecules is then replaced by the so-called collective mode. The rate of osmotic flow across a unit area of the membrane is determined by the concentration and nature of osmolytes on both sides of the membrane, temperature, and the hydraulic permeability of the membrane, which can be determined by measuring water flow when a hydrostatic pressure is applied across the membrane. Osmotic flow is reduced when the membrane is partially permeable to the osmolytes. As water flows into a device containing osmolytes, it dilutes the osmolytes, lowering the osmotic pressure, unless new osmolytes are introduced, for example, by dissolution.

We have already described osmotic pumps in which water invasion across a membrane displaces drug through an orifice. Another way to use osmosis is to coat individual drug particles with semipermeable polymers. After release from a capsule, these particles are exposed to gastric fluid. Water crosses the polymer coatings and dissolves the drug, leading to a gradient in solute concentration that drives even more water inside. To accommodate, the coating must expand, and wall stresses are developed. With sufficient osmotic driving force, the coating ruptures, releasing the drug. Using different coating thicknesses, particles can be programmed to burst at different times. The original time release capsules were based on this principle.

A variation of the elementary osmotic pump theme involves particles or tablets that are coated with a semipermeable polymer membrane which includes sparsely but well-distributed aqueous pores. These pores can be created by excipients blended into the membrane, which dissolve upon exposure to water. Here, water flows across the semipermeable parts of the membrane and displaces dissolved drug inside through the aqueous pores into the release medium.

### 2.3.5 Swelling

Swelling refers to the uptake of water by a polymer system, with increase in volume. Swelling is often a prelude to polymer dissolution. However, swelling may occur without dissolution if water and the polymer are insufficiently compatible, if polymer chain length is sufficiently large, or if crosslinks are introduced to form
a polymer network. Swollen polymer networks or hydrogels reviewed in Chap. 4 may imbibe many times their weight in water.

The swelling process is analogous to osmosis, since water enters the polymer relatively rapidly, while dissolution of polymer into water, if it occurs, is comparatively slow because of the need for polymer chains to disentangle. The extent of swelling depends on the compatibility of water with the polymer material, i.e. the polymer’s hydrophilicity, and on the density of crosslinks between polymer chains, if present. Hydrophobic polymers, reviewed in Chap. 3, imbibe very little water and hence do not swell significantly.

Swelling is a mechanism by which release of otherwise confined drug is activated. If swelling is rapid, then drug diffusion through the swollen polymer is the controlling process for drug release. If swelling is relatively slow, then it can be the process controlling the rate of drug release. A more detailed description of swelling controlled systems is given in Chap. 7.

Swelling controlled release systems are typically glassy polymers at room and body temperatures. Water uptake is initially resisted by the glass, but eventually it makes its way into the free volume at the surface. The glassy polymer at the surface relaxes to a configuration that is more compatible with water, and swells. This permits water to intrude even further, and a moving front is often observed separating a swollen outer layer from a dry inner core. Usually, swelling is accompanied by a glass-to-rubber transition. If drug is trapped inside the glass, it will be liberated when the polymer swells, and if it can diffuse through the softened matrix faster than water can invade, then the release process is swelling controlled. Swelling dynamics are often complex, and a variety of temporal release patterns are observed under swelling control. Under proper conditions, swelling, dissolution of polymer chains, and drug release may occur simultaneously, further contributing to complexity.

Swelling in a polymer may be induced or accelerated by drugs or other additives, which act as effective osmolytes, drawing water into the polymer. By proper selection of polymer, it is also possible to induce swelling by changes in external parameters, such as temperature and pH, which may occur, for example, upon ingestion. Reversible swelling and shrinking of hydrogels can also be induced by alternating these parameters with concomitant on/off patterns of drug release.

2.3.6 Erosion and Degradation

Erodible and degradable drug delivery systems are popular, particularly for implantable or injectable therapies, since they do not require retrieval after drug is fully released. Presently, the most common erodible systems are based on poly(lactic acid) or poly(lactic acid-co-glycolic acid), although systems based on poly(ε-vinyl caprolactone), poly(ortho esters), polyanhydrides, polyphosphates, poly(phosphazenes), and pseudo-poly(amino acids) have also been utilized or studied. Important characteristics of erodible systems are their mechanism and kinetics of erosion. Erosion products must be nontoxic and excretable or resorbable. Principles and applications of erodible
systems are elaborated in Chaps. 5, 8, and 10. In this section, we call attention to two limits of behavior in erodible systems, namely, bulk erosion and surface erosion.

Erosion of polymer monoliths occurs when components of the release medium, especially water, attack covalent bonds in the polymer matrix. For hydrolytically labile bonds, availability of water is an important determinant of local erosion rate. Hydrolysis of bonds may also be acid or base catalyzed, and if so depends on local concentration of proton donors and acceptors. For PLA and PLGA and other polyesters or polyamides, acidic protons are provided by chain ends; hence, concentration of acid protons is inversely proportional to chain length.

Bulk erosion, depicted in Fig. 2.8, occurs when water invades the polymer more rapidly than hydrolysis can occur. In this case, water establishes its presence throughout the matrix, and chain scission processes are initiated everywhere. Hydrolysis may initially be very slow, however, especially if the polymer chains are long. Moreover, initial scissions may endow chains with sufficient mobility that they migrate and form crystallites, which are less susceptible to hydrolysis. However, once a certain degree of hydrolysis has occurred, the process may accelerate. For example, formation of short chains may lead to overall loss of polymer and increase in water concentration by diffusion and/or osmosis. If chain scission results in the formation of acidic end groups and the scission process is acid

![Diagram of the three stages of release of drug from bulk eroding polymers.](image)

Fig. 2.8 Diagrammatic representation of the three stages of release of drug from bulk eroding polymers. The first stage (a) corresponds to drug that is released from the device surface or from pores that are connected to the surface. A second, latent stage follows, during which there is little degradation of polymer and the remaining drug is trapped (b). In the third stage, the trapped drug is released rapidly when the polymer autocatalytically disintegrates (c).
catalyzed, then erosion will be autocatalytic. Thus, bulk erosion may exhibit a sustained quiescent phase, followed by rapid disintegration of the matrix. Prior to disintegration, the dimensions of the device remain relatively constant.

Release of drug from bulk eroding polymers typically exhibits three stages. The first stage corresponds to drug that is released from the device surface or from pores that are connected to the surface. A second, latent stage follows, during which there is little degradation of polymer, and the remaining drug is trapped. In the third stage, the trapped drug is released rapidly when the polymer disintegrates.

Surface erosion, illustrated in Fig. 2.9, occurs either when water invasion is slow or hydrolysis is rapid. For example, polyanhydrides are exceptionally hydrophobic, and the hydrolytically labile anhydride bonds are protected from exposure to water in the interior of the polymer matrix. Thus, hydrolysis with accompanying drug release only occurs at or close to the surface.

A hallmark of surface erosion is that device dimensions decrease with time. If the device is formulated as a slab, then release will be approximately zero order, since each time interval will correspond to the erosion of a layer of polymer and release of drug incorporated in that layer. Erosion rate of cylinders and spheres decreases with time, however, due to reduction in exposed surface area. In principle, drug release correlates with erosion.

**Fig. 2.9** Diagrammatic representation of surface erosion. Drug trapped in the outer layers of the delivery system is released into the surrounding media following erosion of the surface of the polymer. Remaining drug is trapped in the delivery system; however, as time progresses, the polymeric devise erodes from the surface inward and reduces in size, eventually resulting in all drug being released.
While the idealized mechanisms underlying bulk and surface erosion-controlled release are simple, practical systems exhibit extra complexity. Pure surface erosion is almost impossible to achieve, and diffusion of drug out of a matrix may occur ahead of erosion. The drug itself may draw in water, and osmotic stresses (due also to small chain fragments) in the polymer can lead to fracture and uneven penetration. In bulk eroding systems, degradation may even occur more rapidly in the interior of the device due to accumulation of autocatalytic erosion products while leaching of these products leads to slower erosion at the surface. When this is true, thicker matrices may erode more rapidly than thinner matrices.

This section has focused on erosion as a means for controlling drug release. However, it is also possible to program polymer degradation to occur after drug release is more or less complete. For example, hydrogels with degradable crosslinks have been prepared for release of proteins. As these crosslinks degrade, the hydrogel first swells, and then it eventually disintegrates when too few crosslinks are left to maintain the polymer network. Eventually, only primary polymer chains remain, and these are either excreted or resorbed. If degradation is slow, then release is controlled by protein diffusion through the swollen hydrogel network. If degradation of crosslinks is relatively rapid, then the swelling state of the network may change during the release process, and a complex interplay between swelling and diffusion will determine release kinetics.

Finally, we note that water need not be the only agent causing polymer degradation. Incorporating enzyme-labile chains or crosslinkers into a polymer network renders it susceptible to enzymatic degradation. For example, collagen and fibrin gels are specifically degraded in the presence of collagenase and plasmin, respectively. Enzyme-labile peptide fragments of collagen and fibrin can be incorporated into other hydrogels, yielding similar, enzyme specific degradation patterns. Enzyme-mediated degradation exhibits either surface- or bulk mediated erosion features, depending on the ability of enzyme to diffuse into the network and the reactivity of enzyme with the labile components of the network. Such enzyme-degradable systems may be useful in tissue engineering applications, reviewed in Chap. 17, as degradation of a hydrogel may be desirable with growth of tissue, which is signaled by local release of enzyme by cells.

Besides enzymes, small molecules can trigger erosion by cleaving polymer chains or crosslinks. For example, reducing agents can degrade polymers that include disulfide bonds. Since small molecules readily diffuse in even moderately swollen networks, bulk erosion is expected to predominate.

2.3.7 Regional Delivery and Targeting

The benefit of a drug can be greatly enhanced if it can be targeted to its preferred site of action and kept away from sites associated with toxicity. Localization can occur at the organ, tissue, cellular, and subcellular compartment or organelle level. Direct administration at or near the site of action has already been discussed, with
examples provided by systems designed for drug delivery to the eye and coronary arteries. Direct injection of drug carriers into solid tumors or wound sites provides another example. As a third example, the growth, integration, and vascularization of surgically implanted tissue engineered constructs (Chap. 17) may require the localized and well-timed release of growth and angiogenesis factors.

We have also discussed nanocarriers that distribute preferentially in tumors by the EPR effect. To further specify delivery at the cellular level, it is necessary to coat the carrier surface with ligands that bind to specific cell surface features, such as polysaccharides or receptor proteins. Antibodies raised against antigens expressed at the cell surface are the most obvious targeting ligands, but in recent years peptide ligands have been designed based on other known interactions between cell surface receptors and both soluble and extracellular matrix proteins.

Since tumor cells express multiple drug resistance transporters, release of drug from the carrier at the cell surface may not result in increased drug uptake in target cells. The drug/nanocarrier combination is likely to be more effective if it can be brought into the cell by active processes, such as coated pit-mediated endocytosis. Once in the cell, the drug needs to dissociate from the carrier and exit the endosome, in either order. Further targeting of drug to an organelle may require that an organelle-specific “address label” be conjugated to the drug. For example, gene and protein delivery to the nucleus may require that a nuclear localization sequence be conjugated to the active biomolecule in order for the latter to be able to penetrate through nuclear pores.

Targeting systems are the subject of Chaps. 10–12 and further examples are provided in Chaps. 14–16.

2.4 Concluding Remarks

This chapter has illustrated a variety of controlled release strategies and underlying mechanisms. We emphasize that several mechanisms may be at play in a particular controlled release system, especially when more than one stage is involved. We also have reviewed methods to achieve the various goals of controlled release, including improved temporal presentation, drug protection, and localization of drug at the preferred site of action.

This chapter and this book are written from the perspective that controlled release adds substantial value to a drug. However, it should be recognized that development of a controlled-release product can be expensive. For many drugs, the extra expense may not be warranted on purely therapeutic grounds, although developers may pursue controlled release formulations for marketing, quality control, and regulatory reasons. Drugs with a relatively narrow therapeutic range, drugs that are eliminated rapidly from the body, drugs whose efficacy would be enhanced by targeting, and drugs that are susceptible to physiological degradation before absorption are probably the best candidates for controlled release.
Paradoxically, molecular entities that possess these attributes are often screened out early in the discovery and development stages. With improved understanding of controlled-release mechanisms and improved development of technologies, it may be possible to increase the number of bioactive molecules that can be developed fully into drug products [1–35].

References

Fundamentals and Applications of Controlled Release
Drug Delivery
Siepmann, J.; Siegel, R.A.; Rathbone, M.J. (Eds.)
2012, XIV, 594 p., Hardcover