5

Protein Folding II:
Kinetic Pathways

Conceptual Outline

■ **5.1** ■ When kinetics limits the domain of phase space explored by a system, the scaling of the relaxation time $\tau(N)$ may be smaller than exponential. Polymers in a liquid can be in an expanded or compact form. The transition between the two—polymer collapse—is a prototype of protein folding. Using simulations, we will explore possible origins of kinetic limitations in the phase space exploration of long polymers during collapse.

■ **5.2** ■ Before we study collapse, we must understand the properties of polymers in their expanded state in good solvent. Simple arguments can tell us the scaling of polymer size, $R(N) \sim N^{\nu}$. The time scale of relaxation of a polymer from one conformation to another follows either Rouse $\tau(N) \sim N^{2\nu+1}$ or Zimm $\tau(N) \sim N^{3\nu}$ scaling, depending on the assumptions used.

■ **5.3** ■ Polymer simulations can be constructed in various forms. As long as they respect polymer connectivity and excluded volume, the behavior of long polymers is correctly reproduced. A two-space model where monomers alternate between spaces along the chain is a simple and convenient cellular automaton algorithm.

■ **5.4** ■ During polymer collapse monomers bond and aggregate. Simulations of collapse and scaling arguments suggest that the aggregation occurs primarily at the ends of the polymer because of the greater flexibility of polymer-end motion. Thus the aggregates at the end appear to diffuse along the polymer contour accreting monomers and smaller aggregates until they meet in the middle. This results in an aggregation process that is systematically ordered by the kinetics. The end-dominated collapse-time scales linearly with polymer length, which is faster than the usual polymer relaxation. The orderly formation of bonds in end-dominated collapse also suggests that kinetics may constrain the possible monomer-monomer bonds that are formed and thus limit the domain of phase space that is explored in protein folding.

5.1   Phase Space Channels as Kinetic Pathways

In Chapter 4 we introduced the problem of time scale: How can a system composed out of many elements reach a desired structure in a reasonable amount of time? The ability of proteins to fold into a well-defined compact structure exemplified this trait. For our first answer we assumed that the desired structure was the equilibrium state of the system. We then studied various energy functions that would enable relaxation to the equilibrium state. All of these energy functions embodied some variation on the idea of parallel processing. In this chapter we consider the possible influence of kinetics on the time scale for reaching a final structure. Considering kinetic pathways as a mechanism that enables a system to reach a desired structure is a qualitatively different idea from parallel processing. In this approach, a system follows a particular pathway through the phase space to the final desired structure. The pathway may not be unique, but it is severely limited compared to the space of possible paths in the whole space. As a result, there is no reason to expect that the system reaches the absolute minimum energy equilibrium conformation. It does, however, reach a conformation that is low in energy compared to any of the accessible conformations. Because the system only visits a limited set of conformations along the path from its initial to final state, our expectation is that the relaxation time—the time to reach the final conformation—will scale less than exponentially with the size of the system. There are a number of ways that such kinetic pathways can arise. We will discuss a few of these in this section and describe a strategy for considering the effect of kinetics in protein folding.

The simplest form of kinetic pathway can be illustrated using the model of two independent two-state systems introduced in Section 4.2. Each of the two-state systems (spins) has two states $s_i = \pm 1$, for $i = 1, 2$. Relaxation of each spin occurs independently from +1 to −1. Let the relaxation time of the first spin be extremely long compared to the second spin, $\tau_1 >> \tau_2$. If τ_1 is long compared to relevant times (e.g., years) then only s_2 relaxes. The system starts from the state $(1, 1)$. It makes a single transition to the state $(1, -1)$ and is stuck there. The ground state $(-1, -1)$ is never reached. This, however, is fine if $(1, -1)$ is the desired state. It actually doesn't matter whether $(1, -1)$ or $(-1, -1)$ is the ground state. The long relaxation time τ_1 corresponds to a large barrier to the transition of s_1. The accessible domain of phase space includes only states that have $s_1 = 1$. More generally, this form of kinetic pathway assumes that there are energy barriers that partition phase space so that some regions are inaccessible. These regions play no role during the relaxation. For the case of protein folding, this means that certain conformations would be completely inaccessible in a transition from the initial unfolded to the final folded conformation.

An example where energy barriers limit the space of conformations during protein folding is the preservation of primary structure. The bonds between amino acids along the chain are strong bonds that have a low probability of breaking and reforming. Thus, during folding, the protein does not explore the possible arrangements of amino acids along the chain and all of their conformations. The breaking of the chain

is prevented by kinetics, even though there may be other orderings of the amino acids that have lower energy structures. Breaking the protein chain would enter a different domain of phase space. It is interesting that there are specific examples where the chain is broken during protein folding (proteolytic reactions). This is done by enzymes that break the amino acid chain in specific places. The subchains that result are then formed into the final folded protein structure.

The next step is to consider how kinetic pathways might affect the space of conformations of a chain with a particular amino acid sequence. Starting from an unfolded conformation, there do not appear to be any strong bonds or energy barriers that would prevent it from reaching a large number of compact polymer conformations. The number of such compact conformations grows exponentially with the size of the polymer. Thus energy barriers do not appear to be relevant in explaining the ability of a protein to reach a definite structure. However, the kinetic barriers need not exist in the initial conformation. It is enough for them to arise during the process itself. During protein folding, new bonds form. These bonds might restrict the domain of space that is explored. A pictorial illustration of the formation of barriers is shown in Fig. 5.1.1. It shows the emergence of barriers during the kinetic process. These bar-
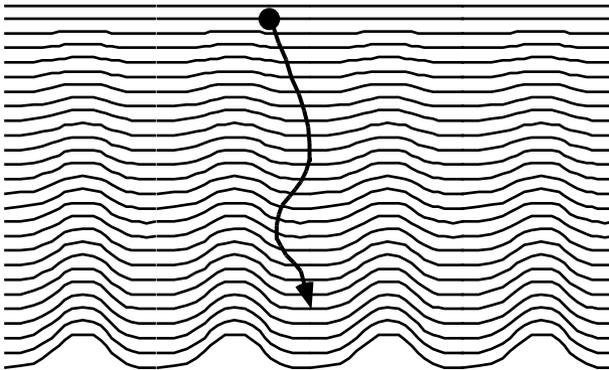
Figure 5.1.1 The simplest concept of a kinetic pathway is a path bounded by energy barriers that prevent departure from the path and thus prevent an exhaustive search of all conformations. In this figure we see that barriers may not exist initially; they may, however, develop as the relaxation proceeds. The illustration should be read as an energy landscape. Horizontal lines are plots of the energy in the horizontal direction. A vertical bias in the energy is assumed so that progressively lower lines are lower in energy. The conformation of a protein is a point in the plane of the page. A possible trajectory is illustrated. From the starting point, it appears that all three of the possible low-energy conformations at the bottom of the illustration are accessible. However, once the relaxation begins there are barriers that prevent the conformation from switching from one of the vertical channels to the other. In order for the correct final state to be chosen, the initial state must be restricted to be close to the channel that leads to the desired conformation. This conformation may or may not be the lowest-energy conformation. ∎

riers do not prevent the protein from folding into an undesirable structure. However, the conjunction of a particular initial configuration and the barriers that arise serve to limit the exploration of space and determine the ultimate conformation.

The picture of strong bonds causing large barriers that form kinetic pathways is not complete. Kinetic restrictions that limit the domain of phase space that is explored during folding arise also from entropic bottlenecks. An entropic bottleneck (Fig. 5.1.2(a)) is a narrow channel between one part of phase space and another. Because the channel is narrow, it is unlikely that the system will move from one part to the other. Thus a whole region of conformations may not arise. Another way in which entropy can be relevant is illustrated in Fig. 5.1.2(b). In this case, entropy differences in the inlets to kinetic pathways reduce the sensitivity of the final conformation to the

(a)

(b)
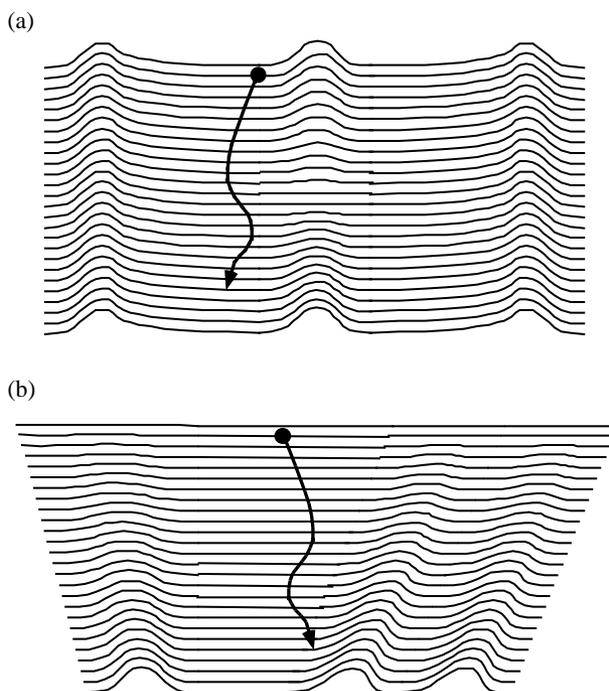
Figure 5.1.2 Entropy can play more than one role in the properties of kinetic pathways, as shown in (a) and (b). (a) illustrates an entropy bottleneck that prevents exploration of all conformations. Two regions of conformations with different energy minima are connected by a channel that is very narrow; like two valleys connected by a narrow mountain pass. It is unlikely that the system will go through the channel, because it will rarely be found by random motion, even if there is no energy barrier in the channel. A different effect of entropy is shown in (b). A wide inlet to a particular kinetic pathway causes it to be preferentially selected over a channel with a narrow inlet. This picture explains how kinetic pathways may lead to a predictable final conformation independent of the initial conformation. Such predictability would be necessary for kinetic pathways to be relevant to protein folding. ∎

initial conformation. Compare this picture with the picture illustrated in Fig. 5.1.1. For our study of protein folding, Fig. 5.1.2(b) will turn out to be relevant. Certain bonds are likely to form during the initial stages of folding. These bonds then inhibit the exploration of all conformations.

Considering energy barriers or entropic bottlenecks to the exploration of phase space are both part of a thermodynamic approach. They are relevant when diffusion dominates the kinetics. Diffusion is a random-walk process that occurs when a system is coupled to a thermal reservoir (Section 1.4.4 and Section 7.2.3). Diffusion is not important in a system that is far from equilibrium and not coupled to a thermal reservoir. The system then follows ballistic motion along a path that is determined by Newtonian equations of motion. This can give rise to other kinetic effects because a system follows a specific trajectory rather than an exploration process. Proteins are, however, embedded in a liquid that serves as a thermal reservoir. The kinetic energy is dissipated, and stochastic diffusive motion dominates. Thus, for proteins, we are amply justified in limiting ourselves to consider diffusive motion. More generally, in order for a stable final conformation to be reached, there must be dissipation of kinetic energy. This suggests that diffusion is important, but does not imply that ballistic motion plays no role.

In the previous chapter we adopted a series of models that ignored the spatial structure of polymers in favor of abstract representations in terms of Ising spin variables. This approach was helpful in developing an understanding of the issues and concepts of parallel processing for protein folding. However, in order to address kinetic limitations that may select pathways for folding, we must build a model of the polymer in space and its dynamics. Our objective is to establish the possible existence of a specific sequence of events in protein folding. Such a sequence of events would result from a particular order in which amino acids encounter each other. The encounters can then result in specific bonds being formed. To determine if a particular sequence of encounters occurs, we can consider a simplified polymer model that retains its spatial structure but does not represent the details of amino acid structure. This approach implies that we are interested in the very first part of the folding process, which might extend no more than a microsecond out of the typically one-second folding time. The potential impact of this initial time is to set the stage for later processes by forming bonds that limit the subsequent exploration of conformations, and by placing the system conformation in the vicinity of its eventual stable state.

In contrast to our studies in Chapter 4 which were essentially analytic, in this chapter we will focus on simulations (Section 1.7) as a tool for investigating the behavior of complex systems. Nevertheless, we begin in Section 5.2 by describing an analytic theory of the kinetics of long polymers. This analytic theory sets the tone for our investigations. It shows that many of the properties that are of interest do not depend on the specific microscopic structure of the polymer, but rather on the general behavior of a long chain and its many conformations in space.

To further develop an understanding of polymer kinetics, particularly the kinetics of polymer collapse, we will turn to Monte Carlo simulations. In Section 5.3 we construct a lattice Monte Carlo model of polymer dynamics. This simple model of

polymer dynamics is constructed to be in the form of a cellular automaton (Section 1.5). The model dynamics serves as a simulation tool. Over many steps, the motion of the model polymer is consistent with the expected behavior of long polymers for relaxation or diffusion. This is ensured by the Monte Carlo method because only local steps in the space of polymer conformations are taken. In Section 5.4, using this model for polymer dynamics, we simulate polymer collapse and find evidence for specific kinetic pathways dictated by a preferential ordering of encounters between parts of the polymer chain. Motivated by the results of the simulations, we develop an analytic scaling theory that describes the kinetics of the transition of a polymer from an expanded to a collapsed state. This generalizes and reinforces the conclusions from the simulations. Finally, we consider a number of variations of the simulations to test the scaling theory and explore the domain of its applicability to physical polymers.

Before we proceed, we point out an example of relaxation that does not illustrate kinetic pathways, even though it appears to. The concept of a kinetic pathway implies a well-defined sequence of intermediate protein structures between the initial and the final conformation. However, the converse is not true. The existence of a well-defined sequence of intermediate protein conformations does not necessarily mean that the system is kinetically limited to this pathway. Another mechanism that may cause a well-defined pathway is the inhomogeneous decoupled model discussed in Section 4.5.1. In that model, certain degrees of freedom relax before others. If the degrees of freedom can be grouped into sets with well-separated relaxation time, then after each set of degrees of freedom relaxes, a well-defined structure arises. Kinetics does play a role because of the degrees of freedom that are frozen at any time. However, by the end of the relaxation, all degrees of freedom can and do relax. This is counter to the assumption of kinetic pathways.

5.2   Polymer Dynamics: Scaling Theory

Polymers are molecules formed out of long chains of atoms that are generally recognizable as a sequence of units (monomers) like amino acids. Biological polymers include proteins, DNA, and polysaccharides. Artificial polymers include polystyrene and polyethylene. Polymers whose monomers are all the same are known as homopolymers. Polymers that have more than one kind of monomer are called heteropolymers. Homopolymers are simpler to model, and we will concentrate on describing their dynamics, though many of the results also apply to heteropolymers. Polymers are found dissolved in liquids or embedded in composite materials. When they are dissolved in liquids, there are essentially two possible structures. Either the polymer collapses into a compact structure or the polymer is expanded. Which of these occurs depends on whether the effective interaction between monomers is attractive or repulsive. The effective interaction includes both the hard core repulsion between atoms and the longer-range attraction or repulsion. Entropy favors the expanded polymer structure over the compact structure because of the greater number of expanded conformations. However, this is generally a weaker effect than that of the

energy of attraction or repulsion. In a solution with many polymers, compact poly-mers often aggregate to each other and precipitate. We will focus on the dynamics of a single polymer, not on polymer-polymer interactions.

When a protein is folded and unfolded in solution, it is crossing the line between compact and expanded structures. The transition may be driven by changes in tem-perature. However, the more usual approach is to add a chemical to the solution that affects the monomer-monomer attraction. What is relevant is the affinity of the monomers for each other compared to their affinity for the solvent. Because of the importance of the solvent for the transition, a polymer in its expanded state is said to be in a good solvent. A compact polymer is said to be in a poor solvent. The transi-tion is called the -point. In this section we are concerned with the properties of a polymer in a good solvent. It is essential to understand the structure and dynamics of this state before we can study the dynamics of transition from the expanded to the compact state. We will be concerned with the scaling behavior of the properties of long polymers as a function of polymer length N. This is similar to Chapter 4 where we considered the scaling of relaxation with system size.

The scaling of the structure and dynamics of long polymers should not depend greatly on their chemical composition. The scaling theory of polymers is one of the great successes of simple concepts in understanding complex systems. A book by de Gennes, who received the 1991 Nobel Prize in physics, contains many of the elegant arguments that describe polymers simply and successfully.

A long polymer in a liquid has a local structure that is more or less flexible. The bonding of adjacent monomers controls the local polymer structure. For a specific pair of monomers, there may be several possible bonding configurations or there may be only one allowed configuration. However, a long enough polymer is always flexi-ble, so we can start by considering it to be a random walk in space with N steps. When a polymer is modeled as a random walk, the size of a step is understood to depend on the polymer flexibility, with stiff polymers having many monomers per step and flex-ible polymers having few monomers per step. For convenience, we can redefine our monomers so that each step is between the new effective monomers.

Polymers are generally found in three-dimensional space. However, we general-ize our discussion to d-dimensions. In d-dimensions a random walk is performed in-dependently in each dimension. The average distance traveled along the polymer from one end to the other is called the polymer end-to-end distance R_0. We use σ to represent the root mean square distance traveled in a random walk in a single di-mension. The random walk in one dimension satisfies (Section 1.2):

$$\sigma = \sqrt{<(x_N - x_1)^2>} = \sqrt{<\left(\sum_i (x_{i+1} - x_i)\right)^2>} = \sqrt{\sum_i <(x_{i+1} - x_i)^2>} = \sqrt{N}a$$

(5.2.1)

where x_N and x_1 are the x coordinates of the last and first monomers respectively. More generally, x_i is the x coordinate of the ith monomer. The third equality follows from the independence of the steps. a is the distance of an elementary step in one di-

mension—the average distance between coordinates of adjacent monomers. Since the random walk in each dimension is independent, the end-to-end polymer length, R_0, is given by a similar expression:

$$R_0 = \sqrt{< (\mathbf{r}_N - \mathbf{r}_1)^2 >} = \sqrt{< \left(\sum_i (\mathbf{r}_{i+1} - \mathbf{r}_i) \right)^2 >} = \sqrt{\sum_i < (\mathbf{r}_{i+1} - \mathbf{r}_i)^2 >} = \sqrt{dN}a$$

(5.2.2)

where $\mathbf{r}_i$ is the d-dimensional vector position of monomer i. The result follows from the existence of dN independent terms in the sum. It could also be written as $R_0 = \overline{N}\overline{a}$, where $\overline{a} = \sqrt{d}a$ is the monomer-monomer distance.

A polymer is a thermodynamic system whose equilibrium size is determined by its free energy. When we consider the polymer as a random walk, we assume that all of the possible configurations have the same energy. The size is then determined by the entropic part of the free energy. As discussed in Section 1.4, the probability of a particular end-to-end polymer distance R can be found using the relationship:

$$P(R, N) = e^{-F(R, N)/kT}/Z = e^{-(E(R, N) - TS(R, N))/kT}/Z = e^{S(R, N)/k}/Z \qquad (5.2.3)$$

$F(R, N)$ is the free energy, $E(R, N)$ and $S(R, N)$ are the energy and entropy respectively, and Z is a normalization constant. The final expression only applies when the energy can be neglected. We have already essentially calculated the probability $P(R, N)$ when we counted the number of random walks of a particular length in Section 1.2. From Eq. (1.2.39), the probability distribution for the distance traveled in a one-dimensional random walk is a Gaussian. In d-dimensions we take a product of the Gaussian probability in each dimension to obtain the probability for a particular end-to-end vector $\mathbf{R}$:

$$P(\mathbf{R}, N) = \frac{1}{(2\pi\sigma)^{d/2}} e^{-R^2/2\sigma^2} = \frac{1}{\left(2\pi R_0/d^{1/2}\right)^{d/2}} e^{-dR^2/2R_0^2} \qquad (5.2.4)$$

where R is the magnitude of $\mathbf{R}$. To obtain the probability of a particular end-to-end distance, R, this must be multiplied by the $d - 1$ dimensional surface area of a sphere. It turns out that none of this detail is important for what follows; however, for completeness we can write the surface area as a constant Ξ^{d-1} times R^{d-1}.

$$P(R, N) = \frac{\Xi^{d-1} R^{d-1}}{\left(2\pi R_0/d^{1/2}\right)^{d/2}} e^{-dR^2/2R_0^2} \qquad (5.2.5)$$

From Eq. (5.2.3), the free energy of a particular end-to-end distance is given by the logarithm of the probability of a particular length times the normalization constant Z:

$$F_{random-walk}(R, N) = -kT \ln\left(ZP(R, N)\right) = F_0 + \frac{dkTR^2}{2R_0^2} \qquad (5.2.6)$$

where we have neglected the logarithmic terms in R and N. Since only free-energy differences matter, the constant F_0 could also be neglected.

There is something unusual about this expression for the free energy. The free energy minimum does not occur at the end-to-end distance R_0 that was found before. It occurs instead at $R = 0$. Part of this problem arises because we neglected the logarithmic term $(d - 1)\ln R$. However, even when we include this term, the minimum occurs at $\overline{((d-1)/d)}R_0$ rather than R_0. This system does not satisfy the usual property of a macroscopic system, that the probability distribution becomes sharp as the system becomes large. In the usual case we can identify the expected value of a system property as the value that maximizes the free energy. For the random walk, the free energy has the more general interpretation, discussed in Section 1.4, as the logarithm of the probability. If we were to use the free energy to evaluate the value of the average radius, we would still have to calculate the average over the probability distribution. After calculating the average we would recover the value R_0. This discussion shows the connection between the entropy, free energy and the characteristic size of a polymer. When we need to add additional terms to the free energy, such as the monomer-monomer interactions, we can recalculate the polymer size using the same expressions.

Our discussion of the random walk in Section 1.2 included a proof of the central limit theorem that allows us with some confidence to consider various random walks to have a Gaussian probability distribution. However, the polymer differs in an essential way from the random walks that were discussed there. The difference is that the steps are not uncorrelated. The stiffness of a polymer would tend to make a polymer continue in the same direction. More generally, the bonding and interactions between monomers near each other along the contour cause constraints between neighboring steps. This turns out not to be an essential problem, because the coupling between steps along a polymer decays exponentially. There is a characteristic distance along the chain after which the constraints become negligible. This means we can choose to label our polymer with random steps as long as the steps are larger than the correlation distance (persistence length). The number of these steps becomes our effective monomer number N. When we have many of them, then, and only then, can we consider the polymer to be long. Proteins in their expanded form turn out to be quite flexible and so the correlation length is short, approximately a single amino acid. On the other hand, DNA is quite stiff. For a single strand of DNA the persistence length is roughly 200 to 400 monomers. For a double strand helix, the persistence length is approximately 2,000 monomers (base pairs).

We are, however, still missing an important aspect of the interaction between monomers. This is the contact interaction between any two monomers that encounter each other. We have argued that we do not need to consider the interactions of monomers near to each other along the chain, because the correlations disappear for long enough polymers. However, the interactions that occur between any two monomers still must be considered. Since the monomers are repelling each other when the polymer is in a good solvent, we can represent the interaction between them as an excluded volume—a volume around each monomer that other monomers can-

not enter. In Section 1.10 we discussed the use of renormalization theory to understand the relevance of parameters in the macroscopic limit. It is possible to show that for isolated polymers in good solvent, the only relevant parameters in the long length limit are the length of the polymer and the excluded volume. This simplifies our considerations and guarantees that our results apply to any polymer, if it is long enough. Our task is to identify the properties of a polymer that has an excluded volume. Such abstract polymers are called self-avoiding random walks. Self avoiding walks should be larger than random walks because of the repulsive interactions between the monomers.

There is a scaling argument for the size of a self-avoiding random walk constructed by Flory. The argument is based on the competition between the repulsive energy of the excluded volume that tries to expand the polymer, and the tension due to entropy reduction when the polymer chain is stretched. Assume that we have a polymer that occupies a volume R^d. The density of the monomers in this volume is given by:

$$c = \frac{N}{R^d} \qquad (5.2.7)$$

In this expression, and throughout, we avoid numerical coefficients, since the objective is only to understand the scaling. The energy of the monomer-monomer interactions is given by the probability that two monomers encounter each other. On average, an encounter costs an amount of energy characteristic of the thermal kinetic energy of the monomers, kT. Once two monomers approach each other close enough to cost this amount of energy, they cannot approach any closer. If we neglect the structure of the polymer, then we can calculate the probability of an encounter. We think of the monomers as distributed with uniform probability in the volume R^d. We imagine placing each of the monomers at random in this volume. Each monomer has an excluded volume V. The number of monomer-monomer overlaps (interactions) is then proportional to the square of the concentration. More specifically, it is given by the number of monomers times the fraction of the volume occupied by monomers. The energy associated with the excluded volume is thus:

$$F_{excluded-volume}(R,N) = kTN \frac{NV}{R^d} = kTV \frac{N^2}{R^d} \qquad (5.2.8)$$

where kT gives the units of energy. The neglect of the polymer structure is a neglect of correlations between monomer positions. This is characteristic of a mean field approach (see Section 1.6). Thus this equation is a kind of mean field treatment of monomer interactions.

The excluded volume energy in Eq. (5.2.8) is smaller for larger R. To this energy we add the free energy for the random walk, Eq. (5.2.6), that neglected the excluded volume. We obtain the free-energy expression:

$$F(R,N) = F_0 + \frac{dkTR^2}{2R_0^2} + kTV \frac{N^2}{R^d} \qquad (5.2.9)$$

We minimize this expression to obtain the typical size of the polymer. We can do this as long as the result is not zero. Neglecting all coefficients we obtain:

$$\frac{R_g}{R_0^2} = \frac{N^2}{R_g^{d+1}} \tag{5.2.10}$$

From Eq. (5.2.2) we have $R_0^2 \sim N$, so:

$$R_g \sim N^\nu \tag{5.2.11}$$

where the exponent ν is given by the expression

$$\nu = \frac{3}{d+2} \tag{5.2.12}$$

The tilde, $\sim$, is used to indicate that the result only holds in the asymptotic regime, i.e., for long enough polymers. The result we have obtained is remarkable. It is exact in one dimension where $\nu = 1$, because the excluded volume walk is a straight line. It has been shown to be exact in two dimensions where $\nu = 0.75$. In three dimensions, where $\nu = 0.6$, it is in reasonable agreement with both experiment and numerical simulation. In four dimensions, it gives the random walk result $\nu = 0.5$. In higher than four dimensions, this must continue to be the result, since it indicates that the excluded volume is irrelevant. This has also been proven to be correct. The reason that the random walk becomes correct in four or higher dimensions is that the free energy due to monomer-monomer interactions for a random walk decreases with the length of the chain (see Question 5.2.1). Thus this simple mean field scaling argument appears to give the exact result in all dimensions. Why does the mean field approach give an exact result in all dimensions? Unlike the mean field treatment of the Ising model, which was exact only in four or higher dimensions, the mean field treatment of polymers appears to benefit from a cancellation of errors.

The alert reader may note that we actually made what might seem an unreasonable step in combining the free-energy expressions in Eq. (5.2.6) and Eq. (5.2.8) to obtain Eq. (5.2.9). The definition of R used to obtain Eq. (5.2.6) was the end-to-end distance of the polymer. The definition of R used to derive the form of the excluded volume energy in Eq. (5.2.8) was the characteristic spatial size of the polymer. In effect, we assumed that all characteristic linear dimensions of the polymer behave in the same way. This is a simplification that is a reasonable first assumption to be made in constructing a scaling theory.

Question 5.2.1 Show in more than four dimensions that the monomer-monomer interactions have decreasing importance and are therefore irrelevant in the long polymer limit. However, for fewer than four dimensions they are not irrelevant.

Solution 5.2.1 In order to see whether the excluded volume is relevant, we evaluate its effect on the polymer free energy. We do so assuming the polymer has a volume given by the random walk without excluded volume. This

is the maximum effect the excluded volume can have. Using the value of the polymer end-to-end distance for a random walk, the excluded volume term in the free energy scales as:

$$F_{excluded-volume} = kTV \frac{N^2}{R_0^d} \quad N^{2-d/2} \tag{5.2.13}$$

The random-walk term in the free energy is independent of polymer length. Thus for any dimension greater than four, the excluded volume interaction has decreasing relative importance with length of the polymer and will not significantly affect the asymptotic behavior of the polymer size. For fewer than four dimensions, the excluded volume term in the free energy increases with the size of the polymer and therefore is relevant. ∎

The dynamics of an isolated polymer consists of diffusion of the whole polymer and internal relaxation of its conformation. Diffusion describes the motion of the polymer center of mass. The internal relaxation of a polymer describes how the polymer changes from one conformation to another. We think of this as a relaxation process because if we know the conformation of a polymer at one time, then the ensemble of this polymer consists of many replicas of the same conformation. However, the random motions of the liquid will cause the ensemble over time to forget the initial conformation and become indistinguishable from an ensemble that started from any other conformation. This is the equilibrium ensemble. The process of relaxation to the equilibrium ensemble resembles the exponential relaxation in a two-state system in Section 1.4. The characteristic relaxation time $\tau(N)$ depends on the length of the polymer. Our objective is to determine the scaling of the relaxation time with polymer length. Dynamic scaling is generally more difficult and less universal than scaling of static quantities like the size of a polymer. Of particular significance when we consider the dynamics of polymers is our treatment of the fluid. This was not relevant when we considered the static structure of the polymer.

There are two established estimates for the scaling of the relaxation time with polymer length $\tau(N)$. The Rouse relaxation time describes the dynamics of a polymer assuming that the motion of a monomer is not correlated to the motion of monomers that are far away along the polymer chain. However, the motion of the fluid, described by hydrodynamics, couples the motion of one monomer to another when they are near each other, no matter how far apart they are along the chain. When a monomer moves, it causes a flow of fluid that in turn moves other monomers. Also, a flow of fluid that moves one monomer has a spatial extent that can move other monomers at the same time. This coupling is taken into consideration in the Zimm relaxation time. We discuss first the Rouse and then the Zimm relaxation.

The dynamics of a polymer becomes slower as the polymer length increases. The relaxation time, therefore, is controlled by the dynamics of the longest length scale—the movement of half of the polymer from one place to another. We first illustrate this using a simple elastic string model. Using this model, we derive the Rouse relaxation for a random walk that neglects the excluded volume. In the elastic string model, we

assume that the distance between adjacent monomers has the same distribution as the end-to-end distance of a polymer. This means that the Gaussian distribution applies already to the intermonomer separation. We can assume this for a very long polymer because we can always relabel our monomers to be farther apart along the chain. For example, we can label every tenth or every hundredth monomer as a new "monomer." We call the chain between the new monomers "the bond" between them. Since this only changes the number of monomers by a constant factor, it will not change the scaling of properties of the polymer. However, by relabeling the monomers, the bond between two of our new monomers itself acts like a long polymer. The relabeling idea only works when we neglect excluded volume.

The free energy of the elastic string depends on the intermonomer separation as given in Eq. (5.2.6). This is the equation for a spring where the energy is proportional to the square of the distance. The force between two adjacent monomers is then proportional to the distance between them. We can write the total force on the ith monomer as:

$$K[(r_{i+1} - r_i) + (r_{i-1} - r_i)] \quad K\frac{d^2r}{di^2} \tag{5.2.14}$$

where K is the spring constant. On the right we have taken a continuum limit with i the position along the contour of the polymer. We assume that the motion of a monomer in the fluid is overdamped, which means that the velocity of a monomer is proportional to the force upon it. Multiplying the force times the mobility of a monomer μ gives the velocity:

$$\frac{dr}{dt} \quad \mu K\frac{d^2r}{di^2} \tag{5.2.15}$$

The solution of this equation is given by exponential relaxation of spatial waves:

$$r_i(t) \quad A\cos(ki)e^{-t/\tau_k} + B\sin(ki)e^{-t/\tau_k} \tag{5.2.16}$$

where $k = 2\pi/\lambda$ is the wave vector of the oscillation of the elastic string. The boundary conditions at the ends restrict the wavelength λ to be no greater than the string length N. By inserting the solution into the differential equation we see that the relaxation time for a particular wavelength is given by

$$\tau_k = \lambda^2/\mu K(2\pi)^2 \tag{5.2.17}$$

Neglecting all the numerical coefficients gives us the longest relaxation time $(\lambda \quad N)$ as:

$$\tau(N) \quad N^2 \tag{5.2.18}$$

This is the Rouse relaxation time when excluded volume is neglected. It applies much more generally than its derivation for the elastic string model would indicate. The main reason for this generality is that the longest relaxation time involves motion of essentially the whole polymer, and therefore does not depend on the local polymer properties.

We take a different approach in order to incorporate the excluded volume and the effect of hydrodynamics in the scaling of the relaxation time. This approach relates polymer relaxation to the polymer diffusion constant. Relaxation of a polymer occurs when a significant part of the polymer (say half) is able to diffuse in a random walk across the whole volume occupied by the polymer. This means that we may write the relaxation time using a random-walk expression by assuming the distance traveled is the diameter of the polymer:

$$R(N)^2 \quad D(N)\tau(N) \tag{5.2.19}$$

This is the usual relationship of distance traveled to the diffusion constant and the time (e.g., Eq. (1.4.56)). We have used the diffusion constant of the whole polymer $D(N)$ rather than $D(N/2)$ because their scaling dependence on N is the same. $R(N)$ is a characteristic length, such as the diameter or radius of the polymer.

Since we already know the size scaling of the polymer, we must derive an expression for the diffusion constant of a polymer. The diffusion of the polymer is given by the displacement of the center of mass of the polymer. In an interval of time t, the center of mass of the polymer r_{cm} changes according to:

$$r_{cm} = \quad < r_i > = \frac{1}{N} \quad r_i \tag{5.2.20}$$

Assuming a mean field treatment, we neglect monomer-monomer correlations. Accordingly, the movement of each monomer is uncorrelated to other monomers. Each term in the sum r_i is an independent random number. The sum in Eq. (5.2.20) is like a random walk with N steps. Thus the sum over all the independent displacements of the monomers is proportional to $N^{1/2}$. The center of mass displacement is given by:

$$r_{cm} \quad N^{-1/2} \tag{5.2.21}$$

The scaling of the diffusion constant is obtained by setting this distance to be the result of a random walk of the center of mass:

$$D = \quad r_{cm}^2 / \quad t \quad N^{-1} \tag{5.2.22}$$

where t, the time for monomers to take the steps r_i, is independent of the polymer length.

Using this diffusion constant, we obtain the relaxation time from Eq. (5.2.19) as the time for diffusion of the polymer across its own radius:

$$\tau(N) = R(N)^2 / D(N) \quad N^{2\nu+1} \tag{5.2.23}$$

This is the generalization of Rouse dynamics to self-avoiding random walks. For two dimensions (three dimensions), it gives an exponent of 2.5 (2.2). Without excluded volume, $\nu = 1/2$, it reduces to the previous result.

In order to describe the relaxation of a polymer including hydrodynamics of the solvent, we start from the Navier-Stokes equation. We will not need to solve this

equation; however, we do need to know what parameters are involved in order to construct a scaling argument. The standard complete set of hydrodynamic equations

$$\frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla)\mathbf{v} = - \nabla P/\rho + \frac{\eta}{\rho}\nabla^2 \mathbf{v}$$

$$\nabla \cdot \mathbf{v} = 0$$

(5.2.24)

describe the macroscopic behavior of the motion of an incompressible fluid. In these equations **v** is the velocity of the fluid, P is the local pressure, ρ is the density which is essentially constant in an incompressible fluid, and η is the viscosity. We do not derive these equations here, but we review the origin of each term. The top equation (Navier-Stokes equation), is Newton's law $d\mathbf{v}/dt = \mathbf{F}/m$ applied at a particular location in the fluid. The left side of the equation is the acceleration of a fluid element. The second term accounts for the displacement of the accelerated fluid element. The right side of the equation is the force divided by the mass. This has two parts, the force due to the pressure gradient and the force due to the effects of shear. The second equation is the absence of a divergence of velocity (outflow of matter from a point) in an incompressible fluid. There are four equations, the three components of the top equation and the bottom equation, and four unknowns, the three components of the velocity and the pressure (divided by the density) $(\mathbf{v}, P/\rho)$.

The underlying assumption of our treatment of a polymer in a hydrodynamic fluid is that the polymer moves with the fluid in which it is located. Thus we think about the motion of the polymer as the motion of a spherical volume R^d of the fluid. Like the other mean field treatment, this approximation neglects the effects of monomer-monomer bonding on polymer motion. In order to obtain the diffusion constant of the polymer, we need to know which parameters it may depend on. A scaling argument follows from dimensional constraints. We imagine the diffusion of a spherical volume of fluid. At any instant, the velocity and pressure fields are solutions of the Navier-Stokes equations. There is one piece of information not contained in the Navier-Stokes equation—the size of the random thermal motion of the sphere of fluid. This is given by the thermodynamic expression for the average velocity of a particle at temperature T (Eq. (1.3.83)):

$$<v^2> \approx kT/m = kT/\rho R^d$$

(5.2.25)

The expression used for the mass of the fluid, $m = \rho R^d$, neglects the small mass of the polymer distributed within it. In addition to the characteristic velocity, there are only two other parameters that are relevant to the motion. One is the size, R, of the fluid volume that is moving. The other is the viscosity, η, which characterizes the fluid. The viscosity only appears in the Navier-Stokes equation in combination with the density as η/ρ.

The diffusion constant must be a function of only three parameters; $(<v^2>, R, \eta/\rho)$. The diffusion constant is related to the thermal velocity by the relationship:

$$D \approx <v^2>\tau$$

(5.2.26)

This relationship is derived later in Section 7.2.3. The time τ is not the relaxation time of the polymer. It is the characteristic time between changes in velocity of the sphere of fluid. In effect, it is the time between collisions of the sphere with the rest of

the fluid. This time depends only on the remaining two parameters $(R, \eta/\rho)$. By looking at Eq. (5.2.24) the dimensions of η/ρ can be seen to be length^2/time. The only combination of η/ρ and R that has the dimensions of time is $R^2\rho/\eta$, which must therefore be proportional to τ . Inserting this into Eq. (5.2.26) we have:

$$D \quad < \quad \mathbf{v}^2 > R^2\rho/\eta \quad \frac{kTR^2}{\rho R^d}\frac{\rho}{\eta} = \frac{kT}{\eta R^{d-2}} \tag{5.2.27}$$

This is called Stokes' law. To concretize this result we give the diffusion constant of a sphere in three dimensions, which can be obtained by solving the Navier-Stokes equation:

$$D = \frac{kT}{6\pi\eta R} \quad \frac{1}{R} \tag{5.2.28}$$

Eq. (5.2.28) is in agreement with Eq. (5.2.27) in three dimensions, and it provides the numerical prefactor for the specific case of a sphere.

There is a problem with the result of Eq. (5.2.27) for two-dimensional systems. There are two aspects to this problem. The first issue is the result itself. In two dimensions, according to Eq. (5.2.27), there is no dependence of the diffusion constant on the size, R, of the system. It turns out that a more careful treatment yields a logarithmic correction, which is nonanalytic. The second issue is the nature of the two-dimensional system that is being modeled. Any two-dimensional system that we encounter is embedded into a three-dimensional universe. A two-dimensional Navier-Stokes equation assumes one of two scenarios. The first scenario is that we have a system formed out of very long cylinders. For a polymer this would correspond to having monomers that are very extended in one dimension. The direction in which the monomers are extended is perpendicular to the direction in which the monomers are bonded to each other. It is also perpendicular to the two dimensions in which the monomers can move. Alternatively, the two-dimensional equation would correspond to having a polymer in a solvent between two solid plates whose separation is no greater than the width of a monomer. These plates allow the polymer to diffuse without sticking. Neither of these scenarios are easy to construct. It is more relevant to consider a two-dimensional problem where a polymer is trapped at an interface. The interface might be a boundary between two liquids. In this case the polymer occupies a space like that of a flat disk in the two-dimensional interface. The Navier-Stokes equation we would solve to describe the motion of the disk is a three-dimensional equation. Even though the polymer only moves in two dimensions, the diffusion constant scales the same as in three dimensions.

In order to see this we must embellish our scaling argument slightly. R would play the role of an overall scale factor of the disk shape. The radius would be given by R, and the height of the disk would be given by ζR, where ζ is a small dimensionless number. All of the scaling statements would hold as before for three dimensions leading to Eq. (5.2.27) for $d = 3$. We are not yet done, because we assumed that the height of the disk changes with the radius, which is not true for a polymer at an interface. However, for a thin disk, the interaction between the fluid and the disk only occurs at

the faces of the disk. The height is irrelevant, and we can use the same scaling for a disk whose height does not change. Thus, as in Eq. (5.2.27), the scaling of the diffusion constant is inversely proportional to the disk radius (polymer radius), $D \quad 1/R$, in the two-dimensional space. Later when we want to show a simulation of two-dimensional polymer collapse, we will choose this scaling dependence both because of its similarity to the properties of three-dimensional collapse and because it is a model of the dynamics of a polymer at an interface.

The diffusion constant of a polymer is given by either Eq. (5.2.27) or Eq. (5.2.28) with the radius given by Eq. (5.2.11). Inserting Eq. (5.2.27) into Eq. (5.2.19) gives the Zimm relaxation time:

$$\tau \quad R^d \quad N^{d\nu} \tag{5.2.29}$$

or for our modified scaling using Eq. (5.2.28) in both two dimensions and three dimensions:

$$\tau \quad R^3 \quad N^{3\nu} \tag{5.2.30}$$

The Zimm relaxation scaling in three dimensions is $3\nu = 1.8$, which is smaller than the Rouse relaxation result 2.2. In two dimensions it is either $2\nu = 1.5$ according to Eq. (5.2.29), or $3\nu = 2.25$ according to Eq. (5.2.30). For much of our discussion the differences between the various relaxation-time scaling exponents will not be significant.

This concludes our study of the structure and dynamics of polymers in good solvent. The next step is to introduce techniques for the simulation of polymers that enable us to investigate the properties of polymer collapse. We will return to scaling arguments for the same problem in Section 5.4.3.

5.3   Polymer Dynamics: Simulations

5.3.1 *Introduction to simulations of polymer dynamics*

In this section we describe several methods for simulating polymers that are both cellular automata (Section 1.5) and Monte Carlo algorithms (Section 1.7). They also illustrate the technique of space partitioning that can be used generally for parallel processing of spatially distributed systems. From a theoretical point of view, one of the most interesting features of these algorithms is that they allow Monte Carlo simulations of extended objects but are inherently parallel. With the advent of massively parallel computers, including cellular automaton machines, inherent parallelism can also be a practical advantage. The algorithms are also quite simple and they illustrate how a simple cellular automaton algorithm can be designed. Simplicity often makes it easier to work with and reason about models. A simple algorithm leads to small, fast programs, and small programs are easier to write, debug, optimize, execute, maintain and modify. One of the algorithms, the two-space algorithm, is particularly convenient and efficient and will form the basis of our simulations of polymer collapse in Section 5.4.

In general, polymer simulations, like other simulations, use either molecular dynamics or Monte Carlo dynamics (Section 1.7). Molecular dynamics simulations are

suggestive of Newtonian dynamics and are implemented by moving all atoms with small steps (no more than 10^{-2} of a characteristic interatomic distance) according to forces calculated from modeled interatomic forces. Monte Carlo simulations represent the dynamics of an ensemble of polymers by steps that take into account transition probabilities required by thermodynamics. Both of these simulation methodologies give the same results for equilibrium ensemble properties like the polymer end-to-end distance or other average structural properties. They also give the same results for dynamical properties that involve motions on a scale that is large compared to a step of an individual monomer. Large-scale motions include polymer conformational change, relaxation and diffusion. All atoms can be moved in parallel (at the same time) in molecular dynamics, which therefore appears to be ideally suited for parallel processing. However, with a processor attached to each atom, calculation of the forces requires a large number of communications between processors. Each atom must communicate to every other atom its position. Connections between processors are the limiting feature of parallel computers. Monte Carlo simulations have a fundamental advantage in that movements of monomers can be much larger and there is no need to specify forces. It is sufficient to specify the energy for a simple model polymer system. Monte Carlo simulations also can take into consideration the thermal reservoir effect of the fluid without simulating the fluid itself. Hydrodynamics, however, is not included. We will focus on the Monte Carlo method, describe why the straightforward approach to parallelization does not work, and then construct a parallel cellular automaton algorithm that does.

In Monte Carlo simulations of polymers, a long chain of monomers is represented by the coordinates of each monomer. There are many different methods for describing the monomer-monomer interactions, the local structure of the polymer and the process of each move. Just as for real polymers, the local polymer structure should not affect the characteristic properties of long polymers, such as the scaling of the size, $R(N)$, or relaxation time, $\tau(N)$.

An example, illustrated in Fig. 5.3.1, is the ball-and-string model. The monomers are rigid balls of radius r_0 that are attached to nearest-neighbor monomers by strings of length d_0 that have no elasticity. The rigid balls are not allowed to overlap—the energy is infinite if they do. The strings are not allowed to break and simply prevent adjacent monomers from separating further than a distance d_0 apart. Smaller distances down to $d_0 - 2r_0$ are possible. To construct a Monte Carlo dynamics for this model we must specify the nature of a move (the matrix λ in Eq. (1.7.19)). The easiest specification is to allow one monomer at a time to move anywhere within a distance r_m from its current location. Then a single step of the Monte Carlo consists of two parts: (1) selecting a move, and (2) accepting or rejecting the move. Selecting a move consists of selecting at random one of the monomers, and selecting a direction and a distance to move the monomer with equal probability within the ball of radius r_m around its original location. The process of accepting or rejecting the move is explained as follows. There are two possibilities, either the final conformation is allowed or it is not allowed. It is allowed if two conditions are satisfied: there is no overlap of the monomer we chose with any other monomer, and the move did not take the
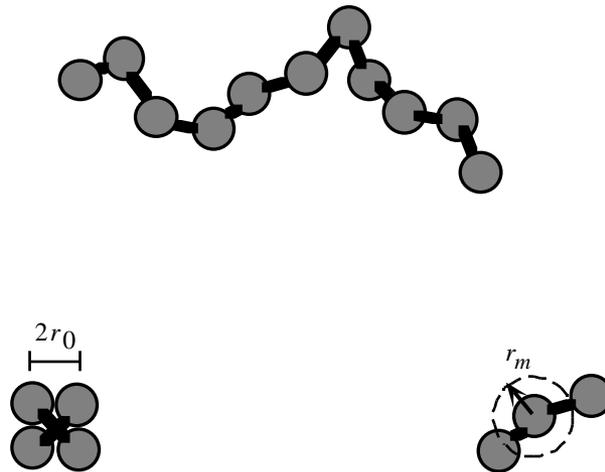
Figure 5.3.1 Illustration of an abstract polymer composed of monomers that are connected to neighbors and do not overlap "excluded" volumes. This is called the ball-and-string model. The motion of monomers is restricted so that they do not separate further than the string length, d_0, from monomers they are bonded to. Monomers have a ball radius r_0 and any two are prevented from overlapping. The strings act only as limits to the separation between monomers, and have no other physical existence. In order to ensure that the polymer cannot cross through itself, d_0 should be less than $2\sqrt{2}r_0$. As illustrated on the lower left, for a larger d_0 the polymer crosses through itself when two bonded monomers at opposite corners of the square move up out of the page and the other two bonded monomers move down into the page. In two dimensions it is enough that $d_0 < 4r_0$, preventing a monomer from passing between two other monomers.

This model polymer can be conveniently simulated using Monte Carlo displacements of individual monomers. A monomer is moved (lower right) to a randomly selected position within a radius r_m around its original position, but only if it does not then violate the structural constraints. Unlike molecular dynamics simulations, however, there are problems in moving monomers in parallel. Moving two bonded monomers might break their bond, and moving any two monomers at the same time can lead to inadvertent overlap. ∎

monomer further than d_0 away from the neighbors to which it is attached by strings. It is not allowed if either of these conditions is violated. For any monomer move that is allowed, the energy of the polymer is unchanged. For any move that is not allowed, the energy increases to infinity. Because the energy change is either zero or infinite, the temperature of the simulation does not matter. Allowed moves are accepted, moves that are not allowed are rejected.

We construct below several different ways of simulating long polymers. In all of them a simulation step consists of selecting a monomer, monomer i, from the polymer chain and performing a move subject to the following constraints:

1. The move does not "break" the polymer connectivity—monomer i does not dissociate itself from its nearest neighbors along the chain.
2. The move does not violate excluded volume—monomer i does not overlap the volume of any other monomer j.

These two constraints, connectivity and excluded volume, are sufficient to guarantee that the structural properties of a long polymer will be found.

In order to study the dynamics of a polymer, we must also guarantee that the steps taken are local steps in the space of polymer conformations. This is generally satisfied when monomer steps themselves are local. However, we must also be sure that the polymer cannot cross through itself. For the ball-and-string model, this limits the size of d_0 (see Fig. 5.3.1). For the types of models we will use, it is easy to verify that a polymer cannot pass through itself.

In naive parallel processing, a set of processors would be assigned one-to-one to perform the movement of a set of the monomers. A processor does not know the outcome of the movement of the other monomers; it can only know their position before the current step. With the two constraints (1) and (2) it would be impossible to perform parallel processing in this way, since moving different monomers at the same time is likely to lead to dissociation or overlap. Dissociation only restricts the parallel motion of nearest neighbors. However, the excluded volume constraint restricts the parallel motion of *any* two monomers, presenting a fundamental difficulty for parallel processing.

5.3.2 *Cellular automata for polymer simulations*

The idea of a cellular automaton is to think about simulating the space rather than the objects that are in it. This is useful for parallel simulation of polymers because, as long as there are no long-range interactions, the motion of monomers that are far apart must be independent of each other. Thus we can assign parallel processors to separated regions of space. When this concept is applied to a continuous space, we call the methodology space partitioning.

Space partitioning could be applied to the ball-and-string model. As shown in Fig. 5.3.2, the space would be partitioned into regions. For Monte Carlo simulations of the ball-and-string model, we could move at the same time monomers selected from regions separated by more than a distance of $2r_m + 2r_0$. At this distance, two monomers moving toward each other at the same time would not enter each other's excluded volume. This approach can work for other polymer models as well. However, it is simplest to implement and simulate for a cellular space of binary variables where the presence or absence of a monomer is indicated by a cell being ON or OFF.

To construct a polymer in a cellular space we could make a polymer model very similar to the ball-and-string model. Instead of a continuum of positions, the locations of monomers would be specified on a lattice. There is an algorithm, the bond-fluctuation algorithm, that implements such a ball-and-string model. However, in the design of a cellular automaton there is an additional feature to keep in mind. We would like to know which monomers are attached to each other solely by their relative

Figure 5.3.2 For the ball-and-string model of Fig. 5.3.1, or other polymer models without long range interactions, monomers sufficiently far apart may be moved independently and in parallel. The figure illustrates the use of space partitioning. If one monomer is selected from each shaded region, the selected monomers can be moved at the same time without chance of overlap. The location of the shaded regions should then be shifted so that all monomers can be moved. ∎

$(r_0 + r_m)$
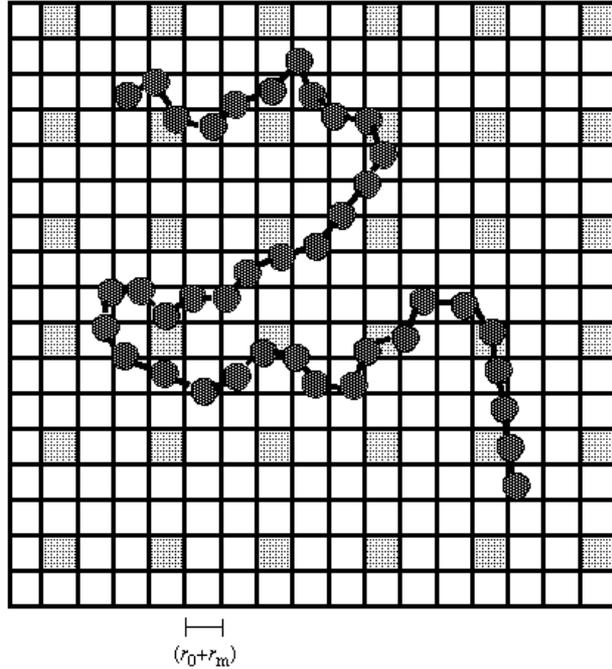
location in space. This is unlike the ball-and-string model where neighbors attached by strings can be farther apart from each other (up to a distance of d_0) than two monomers not attached by strings. All monomers, bonded or not bonded, can approach each other to a smaller distance of $2r_0$. Because of this, we have to keep track of which monomers are bonded to which monomers by labeling the monomers. For a single polymer, the label might be the monomer sequence number along the polymer, which would tell us which monomers are neighbors along the chain and which are not. This labeling would not be convenient in a cellular automaton. The idea of a cellular automaton is that the dynamics only depends on the local spatial conformation. Bonds should be specified only by the relative position of the monomers. Thus we think about a bonded neighbor as a monomer that is closer than a certain distance, and any other monomer has to be farther away. We call the space around a monomer in which its bonded neighbors are located the bonding neighborhood. We note that, since monomers that are not bonded cannot be closer to each other than bonded monomers, in any such model, the polymer chain cannot pass through itself.

A polymer model that implements this in two dimensions is shown in Fig. 5.3.3. In this model, monomers are bonded if they are adjacent either horizontally, vertically or diagonally. The bonding neighborhood is a 3×3 region around a monomer. In three dimensions, we could use a $3 \times 3 \times 3$ cube as a bonding neighborhood. Any monomer except those that are bonded must be excluded from occupying any of these sites. We can think about this as an excluded volume that is larger than a single cell, as illustrated in Fig. 5.3.4. This excluded volume applies to all monomers, except the

Figure 5.3.3 Illustration of a cellular (lattice) polymer model. In this model, monomers are considered bonded to each other if they are touching at either faces or corners. Other non-bonded monomers are not allowed to approach this close. This enables us to distinguish bonded neighbors from other monomers just by the relative location of the monomers. This property is not satisfied by the ball-and-string model. ∎

(a) (b)
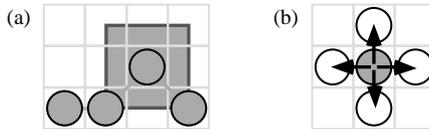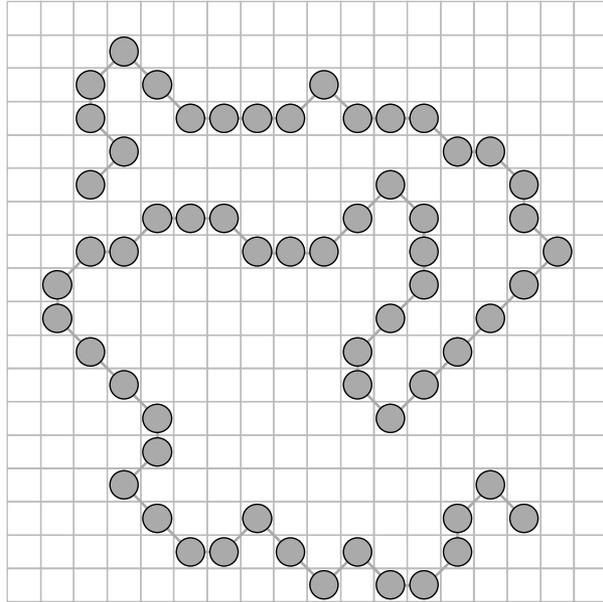
Figure 5.3.4 (a) shows the size of the effective excluded volume for the model of Fig. 5.3.3. The excluded volume is larger than a single cell. It only applies to nonbonded neighbors and prevents them from approaching the adjacent lattice sites. Note that the excluded volume illustrated does not apply to bonded neighbors which have only a one-cell excluded volume with respect to each other. (b) shows the possible moves of a monomer selected for a Monte Carlo update. There is nothing special about this choice of possible moves. We could allow diagonal moves, but the choice of possible moves must be made once and for all for a simulation. ∎

bonded neighbors. Bonded neighbors are prevented from occupying the same site, but can be adjacent. As with other variations of local polymer structure, this is not important for the structure of long polymers. If anything, this is a more realistic model for the bonding of real polymers. Bonds in real polymers are also specified by relative location of monomers—not by a labeling scheme.

The cellular space model in Fig. 5.3.3 could be simulated just like other Monte Carlo models by choosing a monomer, choosing one of the compass directions NSEW, and moving the monomer if the move does not violate either connectivity or excluded volume constraints. We can, however, turn it into a cellular automaton using

the Margolus dynamics approach of updating plaquettes (see Section 1.5.6). This is necessary because we need to conserve various quantities in this dynamics: the existence of the monomers and their bonding and excluded volume constraints. Updating plaquettes enables the implementation of conservation laws and constraints in a natural way. A Margolus dynamics for this model uses a partition of the space into plaquettes that are 3×3 regions with an additional one-cell buffer between plaquettes (Fig. 5.3.5). This enables us to move monomers around in each of the plaquettes independently of other plaquettes in the space. The easiest way to do this is to move only monomers at the central sites of the plaquettes. Choosing a direction for each monomer at a central site, we move it if the constraints allow. Fig. 5.3.6 illustrates the monomer moves that are possible and the moves that are not allowed.

After updating each of the plaquettes, we shift the plaquettes around in the space so that we can move all the monomers. We must keep in mind that even as a cellular automaton this is still a Monte Carlo algorithm. In order for the Monte Carlo algorithm to satisfy detailed balance, it is important to pick the location of the plaquettes at random. Specifically, it is necessary to allow the same location of the plaquettes to be chosen in the next time step as well. This guarantees that a particular move can be reversed. More correctly, detailed balance requires that all possible moves have the same probability of occurrence in every time step, and the random selection of a location for the plaquettes guarantees this. The complete Monte Carlo algorithm consists of selecting a location for the plaquettes, selecting a direction from NSEW for each monomer at the center of a plaquette, and moving the monomer if the move is allowed.

Figure 5.3.5 In order to convert the lattice Monte Carlo algorithm to a cellular automaton we use a Margolus dynamics that consists of 3x3 plaquettes with buffer regions as illustrated. Within each of the plaquettes, we can move the monomers around without interfering with other plaquettes. The simplest way to perform moves in the 3x3 plaquettes is given in Fig. 5.3.6. The periodicity of this partition of space is 4x4. ∎
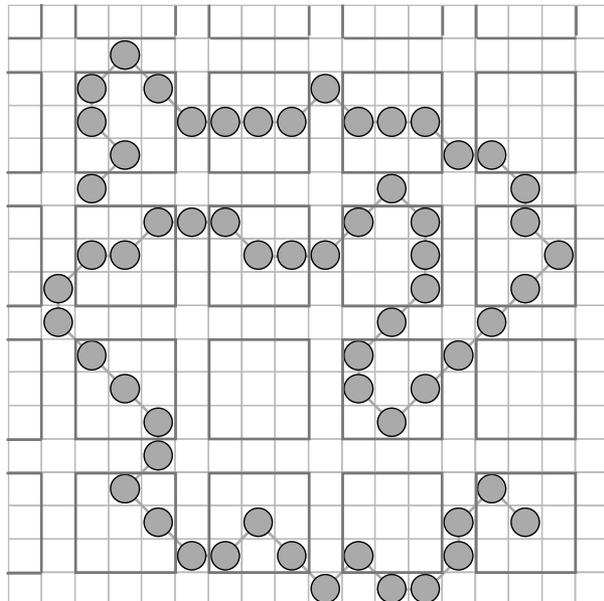
We note that the parallel version of the Monte Carlo algorithm is not, strictly speaking, a Metropolis Monte Carlo algorithm, since the parallel moves do not satisfy Eq. (1.7.19). However, it can be readily shown that a move consisting of a number of parallel independent moves, where each one of them is of the Metropolis form, satisfies detailed balance, Eq. (1.7.17). This is true because the transition matrix factors, as does the equilibrium probability distribution.

The cellular automaton model for polymer dynamics we have constructed can be readily simulated. However, it has a problem that suggests we continue to develop better algorithms. The problem is that the polymer is locally very rigid and the possible local motions of monomers are limited. One way to think about this problem is that for very long polymers, there are two types of motion that are possible: motion of monomers perpendicular to the contour of the polymer and motion along the polymer contour. The latter includes a local stretching or compression of the polymer. For our model, the local motion is always perpendicular to the polymer contour. If we take a long enough polymer, there will be various small folds of the polymer, and motion along the large-scale polymer contour would be possible. However, this means that in order to see the dynamics of very long polymers, we need a very long chain of monomers. Since our objective is to simulate long polymer behavior using as little computation time as possible, we would be better off to have a polymer model that reproduces the long polymer behavior for short polymer chains.

One way to solve this problem is to generalize the cellular automaton model by allowing the monomers that are bonded to each other to separate by one lattice space. Bonded monomers would be located in a larger bonding neighborhood—a 5×5

Figure 5.3.6 In the simplest implementation of the cellular automaton of Fig. 5.3.5, we use the usual Monte Carlo process to update a monomer located at the central site of each plaquette. As illustrated, moves are considered only in plaquettes with monomers in the middle cell of the 3x3 plaquette. Moves that are not allowed due to connectivity or excluded volume constraints are marked by an X in the target square. ∎
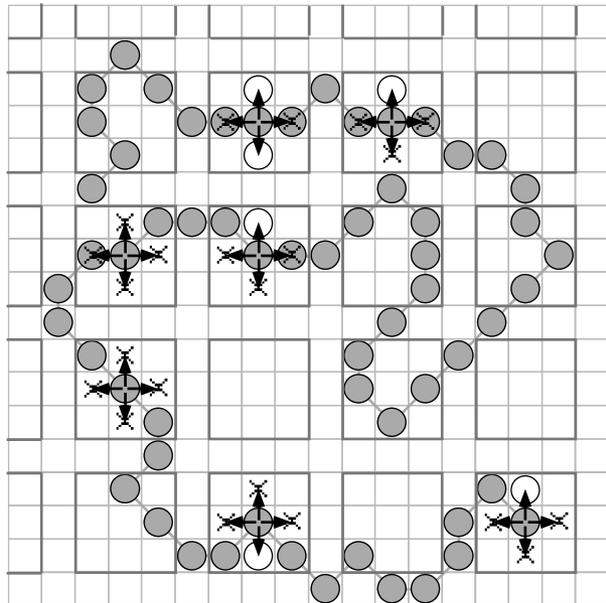
region in two dimensions, or a $5 \times 5 \times 5$ region in three dimensions. This choice of bonding neighborhood is convenient, but others could be specified as well. As before, we do not allow monomers to violate excluded volume by entering a bonding neighborhood, and we do not allow monomers to break a bond by leaving. A monomer move is accepted if monomers are not removed from nor added to the bonding neighborhood by the move. The larger bonding neighborhood allows more flexibility to the motion because adjacent monomers can move toward and away from each other, enabling local contraction and expansion of the polymer. We call this algorithm the one-space algorithm in order to contrast it with the two-space algorithm discussed next.

The problem of polymer flexibility also has a second solution—the two-space algorithm—that has some additional advantages. The simplest way to describe the two-space algorithm in two dimensions is to consider a polymer on two parallel planes (Fig. 5.3.7). The monomers alternate between the planes so that odd-numbered monomers are on one plane and even-numbered monomers are on the other. The neighbors of every monomer reside in the opposite space. We define a 3×3 region of cells around each monomer in the opposite space as its bonding neighborhood. This is the region of cells in which its neighbors reside and no other monomers are allowed to enter. To construct a polymer we place successive monomers so that each monomer has its nearest neighbors along the contour in its bonding neighborhood. The dynamics is defined, as before, by requiring that the motion of a monomer be allowed only if its movement to a new position (selected at random from NSEW directions) does not add or remove monomers from its bonding neighborhood. This preserves both connectivity, preventing loss of a neighbor, and excluded volume, preventing the addition of a neighbor (Fig. 5.3.8).
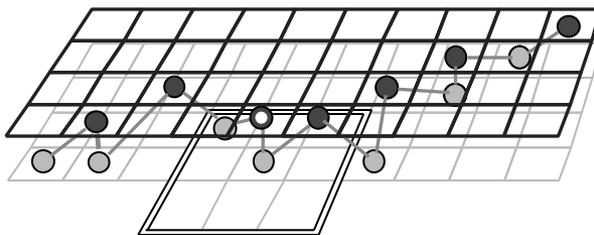
Figure 5.3.7 Schematic illustration of a two-space polymer. In two dimensions, the two spaces are parallel planes. Monomers on the upper plane are shown as circles with dark shading; monomers on the lower plane are shown as circles with light shading. Along the polymer, the monomers alternate spaces so that odd monomers are in one space (the light space) and even monomers in the other space (the dark space). Bonds are indicated by line segments between monomers. Monomers are bonded only to monomers in the other space. The "bonding neighborhood" of each monomer is a 3x3 region of cells located in the opposite plane. The bonding neighborhood of the dark monomer marked with a white dot is shown by the region with a double border. The two neighbors of this monomer, both light monomers, are located in the bonding neighborhood. ∎

Using this model an additional flexibility is achieved, because neighboring monomers can be "on top of each other," so that even the 3 × 3 bonding neighborhood allows local expansion and contraction. Even more interesting, it is possible in this dynamics to move all of the monomers in one space at the same time without concern for interference, because both connectivity and excluded volume are implemented through interactions with the other space. This allows 1/2 of the monomers to be updated in parallel. The simple parallelism of the two-space algorithm lends it to implementation on a variety of computer architectures. Because we can update 1/2 of the polymer at a time, there are two different ways to implement parallelism: space partitioning and polymer partitioning. Space partitioning is the usual cellular automaton assignment of processors to different regions of space. Polymer partitioning is the assignment of processors to different parts of the polymer. Spatial assignment is particularly convenient when a simulation is performed with a high density of monomers. For example, there is considerable interest in simulations of entangled polymers at high densities (polymer melts). Polymer assignment is convenient when the polymer occupies only a small fraction of the space. This is the case for expanded isolated polymers, or problems that might include a polymer moving in a static matrix.

To show that all the monomers in one space can be moved independently, we must show that their motion cannot result in either breaking the polymer or violating excluded volume. Since each monomer move preserves its bonded neighbors, the polymer cannot be broken. Excluded volume is different for two monomers within a space and for two monomers in opposite spaces. For two monomers in opposite spaces, the excluded volume is implemented by preventing monomers from entering
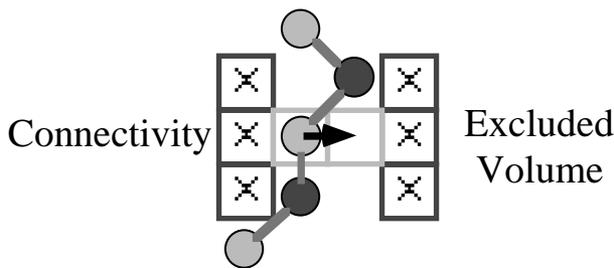
Figure 5.3.8 Illustration of the movement of a monomer in the two-space polymer algorithm. The movement of a light monomer requires checking connectivity and excluded volume in the dark space. The picture illustrates a move where the light monomer is to be moved to the right. To ensure that connectivity is not broken, we check that no monomers are left behind. This is equivalent to checking that there are no dark monomers in the three cells marked with Xs on the left. To ensure that excluded volume is not violated is equivalent to checking that there are no dark monomers in the three cells marked with Xs on the right. If there are no monomers in these cells, then no monomers are removed from or added to the bonding neighborhood of the light monomer as a result of the move. In the picture the move is allowed. ∎

each other's bonding neighborhood. The excluded volume between nonbonded monomers is the same as that shown for the first cellular automaton in Fig. 5.3.2. For bonded neighbors there is no excluded volume. For two monomers in the same space, excluded volume is just the requirement that two monomers do not move onto the same site. They can be adjacent, since they are not within each other's bonding neighborhood. In a proof by contradiction that two monomers cannot move onto the same site, assume that two monomers were to move to the same site. In this state they will have the same bonded neighbors. Since they start with different bonded neighbors and our algorithm explicitly prevents any two monomers from changing their bonded neighbors, this can not happen. There is only one exception, which we may simply avoid (or treat specially). For a polymer of length three, the two end monomers both have the same neighbor and they are not prevented by the algorithm from landing on the same site.

How do we choose the next monomer to move in the two-space dynamics? In order to preserve detailed balance, we must choose which of the two spaces to update at random. This ensures that all possible moves have equal probability in each step. Alternating spaces would not satisfy detailed balance. The order of updates of monomers within one of the spaces does not matter and may be done sequentially rather than randomly.

The two-space algorithm may be implemented in three dimensions by considering the polymer to be in a double space with a $3 \times 3 \times 3$ bonding neighborhood, and a similar generalization of possible monomer moves. If it was desired, we could also remap all of the monomers into a single space with an unusual implementation of excluded volume. As before, the local properties do not affect the long polymer scaling.

To test the algorithm, we can implement and simulate it and measure various structural properties as a function of time. The simulations we perform for these tests are in two dimensions. Rather than measuring the polymer end-to-end distance, we choose to measure the characteristic size of the polymer as given by the radius of gyration $R_g(N;t)$:

$$R_g(N;t)^2 = <(r(t) - <r(t)>)^2> = \frac{1}{N} \sum_i (r_i(t) - <r(t)>)^2$$

$$<r(t)> = \frac{1}{N} \sum_i r_i(t) \qquad (5.3.1)$$

This is just the standard deviation of monomer positions in space. As indicated, the averages are over monomers rather than over time. To initialize the simulation, we start from a straight stretched polymer that alternates from space to space. This is the easiest way to lay out a polymer initially. Simulating the polymer dynamics then results in Fig. 5.3.9. We see that after some number of steps, the polymer relaxes and fluctuates around an average polymer size that we can calculate as a time average. The value of the time average, $R_g(N)$, is indicated on Fig. 5.3.9. It is better to leave out the first part of the simulation in calculating this average in order to eliminate the effect of the improbable first configuration. For a long enough simulation, this correction is unimportant.
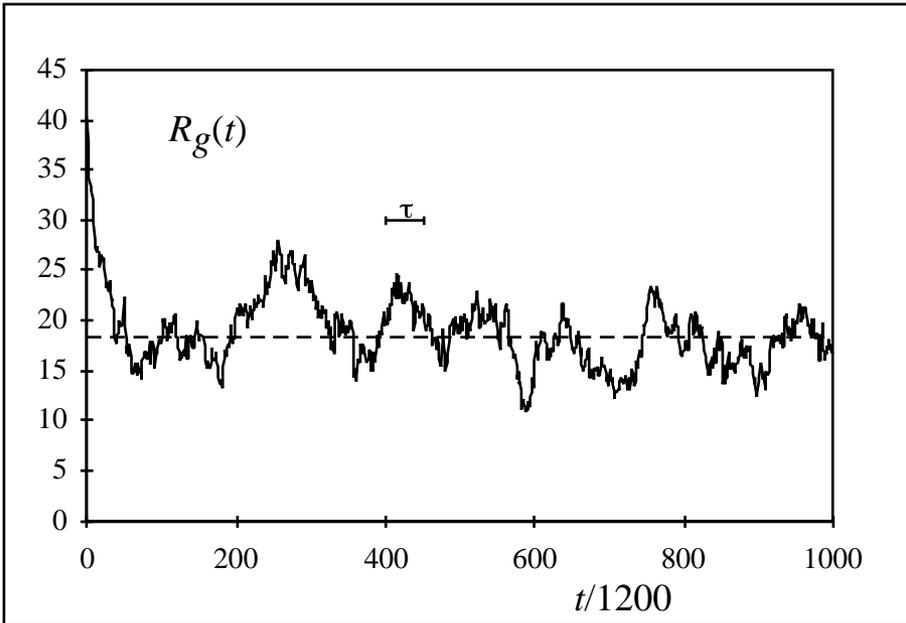
Figure 5.3.9 Plot of the characteristic polymer size, the radius of gyration, $R_g(t)$, as a function of time in a Monte Carlo simulation of the two-space algorithm. The two-dimensional polymer simulated has $N = 140$ monomers. The simulation starts from a completely straight conformation which has an unusually large size. After relaxation, the radius of gyration fluctuates around the average value, $R_g = 18.39$, indicated by the horizontal dashed line. The characteristic time over which the polymer conformation relaxes τ is the correlation time of the radius of gyration indicated by the horizontal bar. The values plotted of the radius of gyration are sampled every 1200 plane updates. There are about 50 samples in a relaxation time. ∎

In order to see the scaling properties of $R_g(N)$ we calculate the average radius of gyration for polymers of different lengths. On a log-log plot (Fig. 5.3.10) the radius of gyration is a straight line for large N. This means that it follows a power-law behavior where the slope of the line is the value of the exponent. The value of the exponent is in agreement with the expected scaling result $R_g(N) \sim N^{0.75}$ from Eq. (5.2.12) in two dimensions. We note that rather than plot $R_g(N)$ as a function of the number of monomers N, the horizontal axis of the plot is $N - 1$, which is the contour length of the polymer. For long polymers, the difference is not significant. For short polymers, this causes the results to follow more closely the long-polymer scaling behavior. The long-polymer behavior is reached for remarkably small polymer chains of only a few monomers. Since our objective is to simulate long polymers, this is a desirable result.

The second test is to evaluate the dynamics of relaxation of the polymer. We can see from Fig. 5.3.9 that there is a characteristic time over which the polymer forgets the value of its radius of gyration. Values of $R_g(N;t)$ fluctuate with a characteristic time
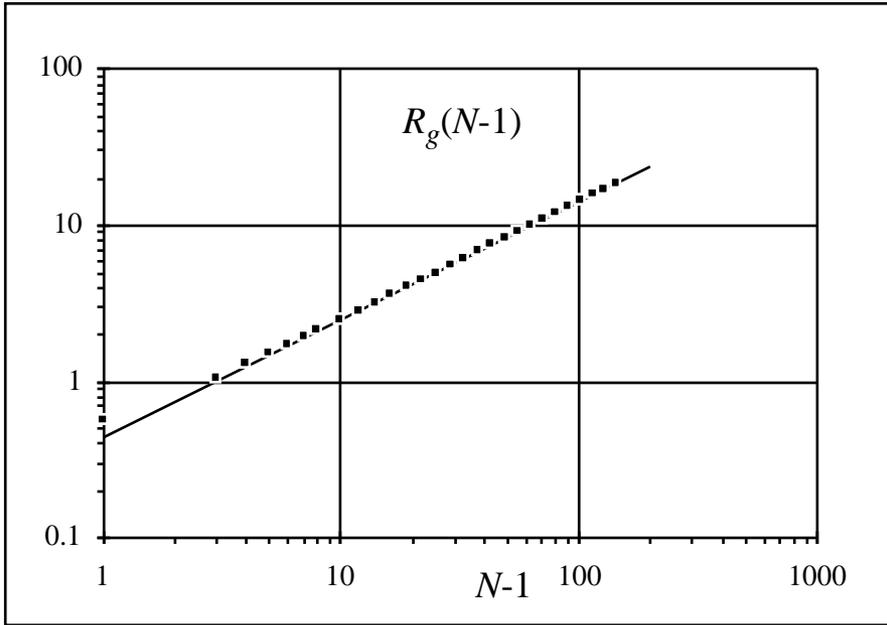
Figure 5.3.10 Plot of the average (root mean square) radius of gyration as a function of polymer length for the two-space algorithm in two dimensions. The average values are obtained by simulations like that shown in Fig. 5.3.9 using 100,000 samples and without including the first 100 samples. The horizontal axis is the number of links, $N - 1$, in the chain. The line in the figure is fitted to the data above $N = 10$ and has a slope of 0.756. This is close to the exact asymptotic scaling exponent for long polymers, $\nu = 0.75$. ∎

shown by the horizontal bar on the plot. Over shorter times than this, values of $R_g(N;t)$ are correlated. Over longer times, they are essentially independent. This characteristic time is the relaxation time, $\tau(N)$. To find a value for the relaxation time we study the correlation over time of $R_g(N;t)$ (for simplicity the dependence on N is not indicated):

$$A[R_g(t)](\Delta t) = \frac{<(R_g(t + \Delta t) - \overline{R_g})(R_g(t) - \overline{R_g})>}{<(R_g(t) - \overline{R_g})^2>}$$

$$\overline{R_g} = <R_g(t)>$$

(5.3.2)

This correlation function measures the relationship between the value of $R_g(t)$ and $R_g(t + \Delta t)$, and is a function of Δt. The averages are over time. The overall behavior of the correlation function can be readily understood. For $\Delta t = 0$ it is normalized to 1. For large Δt, where the value of $R_g(t + \Delta t)$ is independent of $R_g(t)$, the average of the product in the numerator would be the product of the averages of the two factors performed independently. Since the average of either factor is zero, the value of the cor-
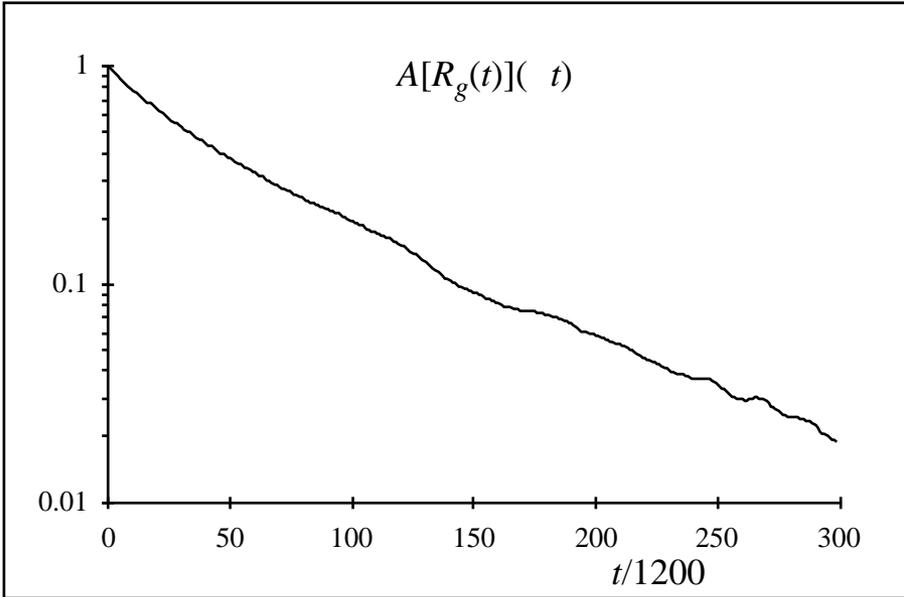
Figure 5.3.11 Plot of the autocorrelation function of the radius of gyration for a polymer of length $N = 140$. The correlation decays approximately exponentially, so the logarithm of the autocorrelation function is roughly linear in time. As in Fig. 5.3.9, the horizontal axis is marked in units of samples taken. The correlation time, τ, is the time at which the autocorrelation function drops to the value $1/e$. Using the integral method described in the text, the correlation time is $\tau = 49.09$ samples of 1200 updates each, or $\tau = 58900$ updates. Only the first 50 values shown of the autocorrelation function determine the relaxation time. ∎

relation function is zero. A plot of the correlation as a function of time is shown in Fig. 5.3.11. If the relaxation of the polymer were simple, the value of the correlation function would be an exponential in time. Since Fig. 5.3.11 is a semilog plot, it would appear as a straight line. The plot is somewhat curved, indicating that it is not a simple exponential decay. Our objective is limited to finding a characteristic relaxation time, $\tau(N)$. We can do this by finding the time at which the correlation falls to $1/e$ of its initial value. However, a better way to measure $\tau(N)$, which reduces the effect of statistical errors, is to integrate the correlation function. If we integrate out to a value $A_0 > 1/e$ then we can estimate the relaxation time using

$$\tau = \frac{\int_0^{t(A_0)} d(t)\, A[R_g](t)}{(1 - A_0)} \tag{5.3.3}$$

where $t(A_0)$ is the time interval at which $A[R_g](t) = A_0$. This formula would be exact if the correlation were exponential and without statistical error.
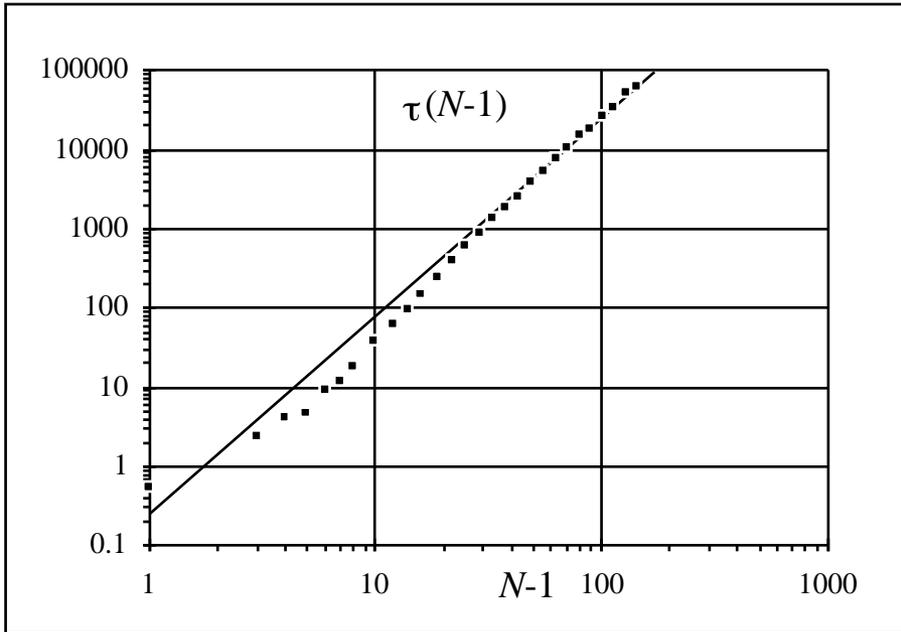
Figure 5.3.12 Plot of the relaxation time τ of polymers as a function of the number of links $N - 1$ for the two-space algorithm in two dimensions. The line in the figure is fitted to the last four points that are relaxation times for polymers longer than $N = 100$. The slope of this line is 2.51, which is close to the asymptotic scaling exponent expected for Rouse relaxation, $2\nu + 1 = 2.5$. ∎

A plot of the characteristic relaxation time $\tau(N)$ as a function of the polymer length is shown in Fig. 5.3.12. We see that $\tau(N)$ increases with length, and for long enough polymers it agrees with the Rouse power law prediction, $\tau(N) \sim N^{2.5}$. Since we have not included hydrodynamics, the Rouse scaling is to be expected, not the Zimm scaling. Short polymers do not have this behavior, but the asymptotic scaling of long polymers is satisfied for polymers longer than approximately 50 monomers. This is still quite short for a polymer, and it suggests that we can effectively simulate long polymer behavior using this algorithm. This concludes our tests of the scaling behavior of the two-space algorithm. In order to apply it to the simulation of polymer collapse, we have to modify it so that monomer-monomer aggregation can occur.

A different Monte Carlo algorithm that enables the study of the radius of gyration or other static properties, but not the dynamics of polymers, is described in Question 5.3.1.

Question 5.3.1 In Section 1.7 we discussed the possibility of considering Monte Carlo algorithms that were nonlocal. These algorithms would provide the correct equilibrium ensemble, but the dynamics would be dif-

ferent from that of local Monte Carlo algorithms. There is an interesting non-local algorithm for polymer simulations. It can be used for various polymer models, including both the one- and two-space polymers. A step of the non-local Monte Carlo moves a monomer from one end of the polymer to the other. To achieve such a change in conformation of the polymer by local monomer steps would require many steps. However, the final conformation is allowed, and so in a Monte Carlo process we can enable this transition. A complete specification of the algorithm is: Select one of the two polymer ends. Delete the end monomer. Select one of the possible neighboring locations of the monomer at the opposite end at random. If the addition does not violate excluded volume constraints, accept the move. Otherwise reject it. Convince yourself that this algorithm satisfies the requirements of a Monte Carlo process. Program and simulate it and see that the radius of gyration does satisfy the standard scaling behavior, but the relaxation does not. ❚

5.4   Polymer Collapse

5.4.1 *Introduction to polymer collapse*

The objective of this section is to develop an understanding of the kinetics of collapse. We do this first through simulations, then a scaling argument, then some more simulations. One of the goals we achieve is to obtain the scaling of the collapse time of a polymer as a function of polymer length. Our primary objective, however, is achieved when we find that the kinetic process of collapse can systematically and reproducibly restrict the possible conformations that are explored. This implies that kinetics of collapse may reproducibly lead to a desired folded conformation without exploring all of the possible conformations of the polymer.

The collapse of a polymer is controlled by the difference of the temperature $T = \quad - T$ from the -point temperature. The lower the temperature the more rapid the collapse and the more important the kinetic effects. This, it may be noted, could be used in experiments to determine the significance of kinetic effects during collapse. If kinetics play an important constructive role, then collapse that occurs too close to the -point might not result in properly folded conformations. Of course, the collapse under these circumstances is also slower.

The process of collapse involves many encounters between monomers that form weak bonds to each other, like hydrogen bonds. Some of these bonds might break and others form instead. The bonds that are formed build larger and larger aggregates. If we are concerned about the kinetic effects, then we don't have to be overly concerned about the bonds that are broken; we can focus only on the formation of bonds. We then imagine a process of irreversible sticking of monomers. One way to think about this is that the key to kinetic effects for polymers is the process of first encounter— those monomers that find each other first. A second way to think about this is that we are considering only large values of T, where the energy of a single bond becomes large compared to the temperature T and the chance of breaking it is small. This picture becomes increasingly valid for longer polymers. For long polymers, we can think

about a coarse graining process that groups monomers and bonds together. The local formation and breaking of a single bond is less relevant than the formation of clusters. The possibility of breaking up a cluster becomes less and less likely for larger and larger clusters because the bonding energy is larger and larger. It is possible to prove that for long polymers we can always think about the process of collapse as if it occurs for large values of T. This is demonstrated formally in the following paragraph.

We can discuss the thermodynamics of polymer collapse using arguments similar to those in Section 5.2, by including an additional term in the free energy that describes the interaction of the polymer with the solvent. The energy of polymer-solvent interaction is given by the energy of a monomer-solvent interaction times the number of such interactions. In a mean field picture where correlations are ignored, this would be written as:

$$F_{polymer-solvent} = \chi kTN(1 - \frac{NV}{R^d})$$ (5.4.1)

The prefactor χkT gives the energy of interaction of monomer with adjacent solvent. The rest is the number of monomers times (in parenthesis) the probability that solvent is found adjacent to a monomer. This probability is 1 minus the volume fraction of the solution occupied by monomers. Adding Eq. (5.4.1) to Eq. (5.2.9) we have

$$F(R) = F_0 + \frac{dkTR^2}{2R_0^2} + (1-\chi)kTV \frac{N^2}{R^d}$$ (5.4.2)

We see that the interaction with the solvent acts to change the effective monomer-monomer interaction. If the coefficient is negative, then the polymer self-attracts and collapses. When this happens, the free energy we have written down is not sufficient, because it has no terms that stop the radius from decreasing to a point. We need to add a term that increases with increasing monomer concentration and can stop the collapse. To do this we treat the free energy as an expansion in the concentration and add a positive term with one higher power of the concentration $c = N/R^d$:

$$F(R) = F_0 + \frac{dkTR^2}{2R_0^2} + (1-\chi)kTV \frac{N^2}{R^d} + fkT \frac{N^3}{R^{2d}}$$ (5.4.3)

where f is known as the third virial coefficient. It is convenient to rewrite this in terms of the ratio of the radius to the random-walk radius $y = R/R_0$ giving

$$F(R) = F_0 + \frac{dkT}{2} y^2 + (1-\chi)kTV \frac{N^{2-d/2}}{y^d} + fkT \frac{N^{3-d}}{y^{2d}}$$ (5.4.4)

To find the expected value of y, we take the derivative and set the result equal to zero to obtain an equation:

$$0 = y^{d+2} - (1-\chi)VN^{2-d/2} - 2f \frac{N^{3-d}}{y^d}$$ (5.4.5)

We know that $(1 - \chi)$ has to be positive for T greater than   and negative for T less than   . We can substitute a linear dependence for $(1 - \chi) = c$ T. Limiting ourselves to three dimensions, we can write Eq. (5.4.5) as:

$$0 = y^{d+2} - cV(TN^{1/2}) - 2f \frac{1}{y^d} \tag{5.4.6}$$

Looking at this equation, we see that a large value of T has the same effect as a large value of N, since the only relevant parameter that controls the collapse is $TN^{1/2}$. This shows that for long polymers, large values of N, we can always think about the collapse as occurring at low effective temperatures, or large $T_{eff} = TN^{1/2}$. This argument provides a formal justification of our treatment of collapse using monomers that encounter each other and stick (bond) irreversibly.

In Section 5.4.2 we describe simulations that indicate a possible relevance of kinetics to polymer collapse. They motivate a scaling argument, described in Section 5.4.3, which generalizes the results. Section 5.4.4 contains a discussion of the implications of the results for protein folding and other systems. Additional simulations in Section 5.4.5 explore the sensitivity of the results to the details of the polymer structure.

5.4.2 *Two-space simulations of collapse*

Using the cellular automaton Monte Carlo algorithms developed in Section 5.3 we can study the problem of polymer collapse. The simulation of polymer collapse starts from a set of equilibrium (expanded) polymer configurations in good solvent generated by Monte Carlo simulations. To generate these conformations, we use either the two-space algorithm with the local monomer moves or the nonlocal Monte Carlo described in Question 5.3.1. The nonlocal Monte Carlo is faster than the local Monte Carlo dynamics and yields the same equilibrium polymer conformations. However, because it is nonlocal, we cannot use it for simulating dynamics such as the collapse itself.

In order to simulate collapse of a polymer, we must allow monomers to stick to each other and form aggregates. Once aggregates form, we have to track their shape, move them as a unit, and allow continued aggregation at their boundaries as they move. This will complicate our simulations substantially. Before we try this, is there an easier way? Aggregation would be much simpler if we allowed monomers to move onto the same site of the lattice. Then the aggregate would look the same as a monomer for the simulation. We can make the simulations a little more realistic by keeping track of the aggregate mass—the number of monomers that have accumulated. Since this is easy to do, we might try to simulate collapse and see what results we find. Later we can make tests to verify or correct the results. This approach is quite similar to the way scaling relations were derived in Section 5.2—use the simplest method possible, then check if it makes sense and verify it with a more complete analysis. Whether the simple method works or not, we will have learned something valuable about what is important and what isn't. If the simple approach works we

learn that the results are general and robust. If the simple approach does not work, then by investigating what details change the results we learn what aspects of the problem or parameters play a role. An example of this approach is the treatment of the scaling of polymer size in Section 5.2. The random-walk model was not quite enough to give the correct exponent. Incorporating excluded volume was necessary and also sufficient. For simulations of polymer collapse, a simple approach that doesn't quite work is described in Question 5.4.1.

Question 5.4.1 In terms of our polymer simulation algorithms, aggregation is easiest to think about as the elimination of the excluded volume constraint. This allows one monomer to approach another and bond by entering its bonding neighborhood. Once they bond we would not allow them to separate. This means we continue to impose the connectivity constraint by not allowing a monomer to leave behind another monomer. We might simulate collapse by keeping track of each individual monomer, even if the monomers occupy the same site. However, there is a problem with this approach. Consider what happens to the diffusion constant of an aggregate. Show that the diffusion constant of an aggregate is not realistic and that this must distort the outcome of the simulations.

Solution 5.4.1 In the proposed method, monomers aggregate by moving into each other's bonding neighborhoods. As time goes on, aggregates form with many monomers bonded to each other in their mutual bonding neighborhoods. The problem with diffusion is that for an aggregate with N monomers, the diffusion constant of the aggregate becomes exponentially small with N. In order to see this, we can focus on the motion of the bonding neighborhood associated with a monomer. Assume we begin with a bonding neighborhood located at a particular place in the lattice. All of the bonded monomers are located in this bonding neighborhood. In order for the bonding neighborhood to displace by one square to the right, none of the monomers must be in the leftmost set of cells of the bonding neighborhood. Any monomer that stays on the left would veto the motion to the right. Thus, in effect, in order to move to the right all of the monomers must move to the right at the same time. Since each monomer is moving independently, the probability of this occurring decreases exponentially with the number of monomers in the aggregate. This is unrealistic. In a simulation, once aggregation starts to occur, the continued motion of an aggregate is too slow. This problem can be solved as discussed in the text, by moving aggregates as a unit. ∎

The example described in Question 5.4.1 illustrates an important feature of computer simulation and research in general. One of the most important and yet most difficult skills to learn is to distinguish a correct from an incorrect result by simple criteria. This capability is crucial when performing computer simulations because, to one degree or another, the computer acts as a black box. We are unable to verify the

performance of the simulation ourselves directly. This is a problem both for the presence of an error—bug—in the computer program, as well as an error in the methodology or approach to simulation. The latter is illustrated by Question 5.4.1. For some students, the problem of telling whether a simulation is correct may seem an impossible one. However, this is a skill that we develop. An example is the ability to determine if two numbers have been multiplied correctly. After the multiplication we can check whether the order of magnitude is correct and whether the number is even or odd. We can perform these and other tests independent of the multiplication itself. By verifying that these aspects of the multiplication are correct, we increase the likelihood that the entire multiplication is correct. When computer simulations are performed, one of our best tools to determine whether it is valid is to view the simulation as a movie. By viewing it, our qualitative concepts about the simulation can be evaluated and compared with what is observed. For this reason, it is important to build a mental model of how the simulations should appear. When the simulations and mental model do not agree, we can either correct the mental model or the simulations based upon further inquiry. This approach is not guaranteed to work, since we might have an error that falsely makes the simulation agree with our mental concept. However, performing such tests does increase the reliability of the results. In effect, it is one way we can compare two independent models of the same process. Whenever we can compare more than one model of the same process, we gain an understanding of the reliability and robustness of results.

We begin to investigate the behavior of polymer collapse using the two-space lattice Monte Carlo algorithm. Starting from an initial equilibrium conformation, polymer collapse is simulated by eliminating the excluded volume constraint. We discuss below why excluded volume may not be necessary during collapse even though it is necessary for the original polymer conformation. Once the excluded volume constraint is eliminated, the usual monomer Monte Carlo steps are taken. Monomers are no longer prevented from entering the neighborhood of another monomer; however, they continue to be required not to leave any neighbors behind. This enables monomers of the same type (odd or even) to move on top of each other. Once one monomer moves onto another monomer, they lose separate identity and become an aggregate. Aggregates are moved as a unit. This avoids the problem we found in Question 5.4.1. We keep track of the mass M of an aggregate, which is the total number of monomers that reside on the same site. We can assign a diffusion constant to the aggregate which depends on the mass of the aggregate. The most natural choice is to set the diffusion constant according to Stokes' law: $D(M) \quad 1/M^x$. As discussed at the end of Section 5.2, we use a diffusion constant in two and three dimensions that scales as $D \quad 1/R$, so in d-dimensions $x = 1/d$. By incorporating Stokes' law into the collapse, we have introduced hydrodynamics into the simulation. Hydrodynamics is not generally part of a lattice simulation. However, by explicitly setting the diffusion constant, we have incorporated its primary effect when there are aggregates present.

What role does the diffusion constant play in the simulations? The diffusion constant is proportional to the rate at which an aggregate hops from a lattice site to a lattice site (Eq. (1.4.55)). When we perform a Monte Carlo simulation, steps occur in

discrete time. The rate of hopping is represented as a probability of making a hop in a single time step. Thus we can implement the diffusion constant by controlling the probability of making a hop when an aggregate is selected to move. We choose our time scale and normalize the probabilities by setting to one the probability that a single monomer will move when chosen. All moves, of course, may be rejected if they violate other constraints.

The polymer dynamics are then simulated by selecting at random an aggregate (monomers are included as aggregates of mass 1), and moving the aggregate in one of four compass directions with a probability given by the diffusion constant, and only if connectivity constraints allow—the aggregate does not leave any neighbors behind. In order to move the aggregate with a probability given by its diffusion constant, a random number ranging between zero and one is compared with the diffusion constant. The monomer is moved only if the random number is smaller than the diffusion constant.

In order to describe the time dependence of the collapse, we must keep track of the passage of time in the simulation. Time is normally measured in a Monte Carlo simulation of polymers by choosing randomly N monomers to move in a single time interval. On average, each monomer is moved once in a time interval. During the collapse we must do a similar counting, where one time interval of the simulation consists of performing a number of aggregate moves equal to the number of remaining aggregates. Since the number of aggregates can change during the time interval, it is arbitrarily taken to be the number at the end of the time interval. As monomers are moved, a counter is incremented and compared with the number of aggregates remaining. When the number of moves exceeds the number of aggregates, a new time interval is started.

When we simulate collapse, we introduce effective interactions between monomers in the same space. This interaction is the aggregation itself. We therefore do not consider parallel processing in the simulation of collapse. While we don't take advantage of parallelism, other aspects of the two-space algorithm make it convenient for these simulations.

A sequence of frames from a simulation in two dimensions is shown in Fig. 5.4.1. Each aggregate is shown by a dot. The area of the dot is the mass of the aggregate. Most striking in these pictures is that the ends of the polymer have a special role in the collapse. The ends diffuse along the contour of the polymer, eating up monomers and smaller aggregates until the two end aggregates meet in the middle. Along the contour, away from the ends, the polymer becomes progressively smoother. The polymer becomes more and more like a dumbbell. One way to think about this process is that the polymer collapse becomes essentially a one-dimensional process along the polymer contour. We call this process end-dominated collapse.

5.4.3 *Scaling theory of collapse*

In this section we develop a scaling theory that describes the results found in the simulations. Before proceeding, we summarize a mean field model for the kinetics of polymer collapse that is relevant to conditions where the collapse occurs slowly be-

Figure 5.4.1 Frames, "snapshots," of the collapse of a single homopolymer of length $N = 500$ monomers in two dimensions. The plot is constructed by placing dots of area $M^{1/2}$ for an aggregate of mass M. This does not reflect the excluded volume of the aggregates, which is zero during this collapse simulation. Simulations that include excluded volume during collapse demonstrate similar results (Section 5.4.5). Successive snapshots are taken at intervals of approximately one-quarter of the collapse time. The initial configuration is shown at the top. The results demonstrate the end-dominated collapse process, where the ends diffuse along the contour of the polymer accreting small aggregates. ∎

cause T is very close to the θ-point. This means that monomers bind and unbind many times during the collapse. Under these conditions, the behavior of the polymer in solution is like the behavior of two liquids trying to separate because they are immiscible. This is called phase separation. The slowest process is the longest length scale separation. The final structure of the system is a sphere of polymer, so the longest length scale relaxation is the contraction of a prolate spheroid, a sausage shape, to a sphere. The longest time scale is set by the diffusion that causes the sausage to thicken as the ends contract. Despite the difference in the nature of this process from our

usual polymer relaxation, the scaling of the collapse time is essentially consistent with the Zimm relaxation of a polymer above the   -point, $\tau(N) \quad N^{3\nu}$.

In contrast to the mean field scenario, in order to investigate the kinetic effects in polymer collapse we must consider larger departures from the   -point. As discussed in Section 5.4.1, a model in which monomers stick irreversibly becomes valid progressively closer to the   -point as the polymer length increases. Below the   -point, monomers attract each other. However, unless the monomers are charged, for a long polymer the interactions are short range. Thus we will consider monomer-monomer interactions only when they come into contact. When two monomers come into contact, they stick to each other and do not separate. Therefore, the first step in understanding the kinetics of polymer collapse is identifying the order in which monomers encounter each other. In completely irreversible collapse (sticking), every encounter causes a kinetic barrier to arise, as described in Fig. 5.1.1, that does not allow the two monomers to separate. In a more realistic model there would be reversibility; however, the final conformation will evolve from a conformation established by the initial encounters. This picture is convenient for our analysis because the initial encounters between monomers occur when the polymer is expanded, and therefore we can understand it beginning from the theory we have developed for the polymers in good solvent.

When we consider qualitatively the process of polymer collapse, we realize that encounters of monomers that are distant from each other along the contour of the chain are unlikely because they are also, on average, distant from each other in space. We can therefore begin by limiting ourselves to consider aggregation as primarily a local process where a monomer forms an aggregate with neighboring monomers along the contour. This kind of aggregation is, however, inhibited by the existing bonds. Aggregation occurs when two monomers that are near each other in space move close enough to form a new bond. The easiest aggregation would occur if a monomer could move to aggregate with one of its neighbors, however, the neighbor on the other side prevents this because stepping away from the other neighbor would break an existing bond. Without curvature in the chain, the monomer is unable to move to aggregate with either neighbor, because it is bonded to the neighbor on the other side. If there is some curvature, then monomers can aggregate. The aggregation would cause the curvature to decrease, and further aggregation becomes more difficult. The same argument applies if we consider a monomer moving to bond to its second or third neighbors along the contour. These problems do not occur at the ends of the polymer. The ends, because they have only one neighbor, can move to aggregate with the monomers near them along the contour. Thus during collapse, the aggregates at the ends grow more rapidly than aggregates along the contour and eventually the polymer looks like a dumbbell. This is what was found in the polymer collapse simulations (Fig. 5.4.1).

To develop a scaling argument for irreversible collapse that distinguishes only between the ends and the average collapse along the contour, we assume that we can summarize the collapse using two variables $M(t)$ and $M_0(t)$. $M(t)$ is the average mass of an aggregate along the polymer at time t, but it does not include the end aggregates. $M_0(t)$ is the mass of the aggregate at either end of the polymer. At time t the end ag-

gregate has grown to size $M_0(t)$ and the average mass along the contour is $M(t)$. In a small time interval, each end aggregate has a probability proportional to its diffusion constant of collecting more mass by moving toward and accreting its immediate neighbor aggregate. This neighbor has an average mass $M(t)$ and is a distance a away that does not depend on time. Thus, on average $M_0(t)$ grows according to

$$\frac{dM_0(t)}{dt} \quad M(t)D_0(t)/a^2 \tag{5.4.7}$$

By Stokes' law (see the discussion at the end of Section 5.2) the diffusion constant $D_0(t)$ of the end aggregate decreases in time as its mass increases according to the relationship

$$D_0(t) \quad 1/M_0(t)^{1/d} \tag{5.4.8}$$

We assume that the two quantities $M(t)$ and $M_0(t)$ follow a power law scaling with time:

$$M_0(t) \quad t^{s_0}$$
$$M(t) \quad t^s \tag{5.4.9}$$

Inserting Eq. (5.4.9) and Eq. (5.4.8) into Eq. (5.4.7), we can ignore prefactors and set the exponents of the time on both sides equal.

$$t^{s_0-1} \quad t^s/t^{s_0/d} \tag{5.4.10}$$

Solving for s_0 in terms of s we obtain:

$$s_0 = (s+1)d/(d+1) \tag{5.4.11}$$

The two exponents are equal when $s = s_0 = d$. This would correspond to the case of uniform collapse, where there is no difference between collapse at the ends and collapse along the chain. We will find instead that s is much smaller than d, and therefore s_0 is much larger than s.

The value of s may be obtained from a second scaling argument by considering collapse of a polymer with fixed ends at their average equilibrium separation. This removes the dynamics of the end motion from the problem. The collapse of this fixed-end polymer would result in a straight rod of aggregates with an average mass $M = N/R$, where N is the number of original monomers and $R \quad N^\nu$ is the average end-to-end distance of the original polymer, which is also the length of the resulting rod. Since we have eliminated the special effects of the ends, the time over which the collapse occurs can be approximated roughly by the usual dynamics of a polymer with $\tau \quad N^z$, where $z = 3\nu$ or $z = 2\nu + 1$ for Zimm or Rouse relaxation respectively. We think about the fixed-end polymer collapse as a simple model for the collapse of the original polymer along the contour away from the ends. We are assuming that the polymer can be approximated locally by polymer segments whose ends are pinned. Over time, progressively longer segments are able to relax to rods. The average mass of the aggregates at a particular time t is then determined by the maximal segment length $N(t) \quad t^{1/z}$ that relaxes by time t. This means that

$$M(t) \sim N(t)/R(t) \sim N(t)^{1-\nu} \sim t^{(1-\nu)/z} \qquad (5.4.12)$$

or

$$s = (1-\nu)/z \qquad (5.4.13)$$

The value of s obtained from this argument (see Table 5.4.1) is small. Using Eq. (5.4.11) we also find that s_0 is much larger than s. This means that the ends will play a special role in the collapse of long polymers. The mass of the ends increases more rapidly than the average mass along the chain and eventually dominates the aggregation. We have seen this result in the simulations of collapse.

The idea that the polymer becomes straighter with time, because regions of higher curvature aggregate more rapidly, can be made more precise. To characterize this behavior it is convenient to compare the distance, r, between two designated monomers (not necessarily the polymer ends) with the contour length, l, of the polymer connecting them. The contour length is the number of bonds between them along the chain. When aggregation occurs, the small aggregates that form, appearing like beads on the chain, decrease the effective contour length of the polymer. We can define the effective contour length by counting the minimum number of monomer-monomer bonds that one must cross in order to travel the polymer from one designated monomer to the other. Bonds formed by aggregation allow us to bypass the usual polymer contour. In this way the effective number of links in the chain decreases over time.

We consider the scaling of $r(l,t)$ as a function of l. Before collapse begins ($t = 0$), the scaling is given by $r \sim l^\nu$. This is just the usual scaling of the end-to-end distance of a self-avoiding random walk, Eq. (5.2.11), because the number of links is essentially the number of monomers. If the polymer becomes straighter, the scaling exponent will increase over time. At long enough times, the scaling will approach that of a straight line ($r \sim l$). However, the smoothing occurs first at the shortest length scales. The characteristic time over which a particular length of polymer becomes straight is the relaxation time of the polymer segment. We can approximate this as in the previous paragraph using the usual relaxation, $\tau \sim l^z$.

We can summarize the behavior of the polymer over time using a universal scaling function. This function describes how the end-to-end distance depends on the contour length as a function of time. The essential idea of the scaling function is that the behavior on different length scales can be described by the same function. Specifically we have a relationship of the form:

$$r(l,t) = l f\left(t/l^z\right) \qquad (5.4.14)$$

This relationship summarizes our previous discussion through properties of the universal scaling function $f(w)$. $f(w)$ is a constant for large values of its argument (long times) because at long times $r \sim l$. We also know that it scales as $w^{(1-\nu)/z}$ for small values of its argument in order to ensure the correct scaling of the self-avoiding random walk, $r \sim l^\nu$, at $t \sim 0$. The crossover between one behavior and the other occurs at a particular value of the argument, $w = w_0$, so that the relaxation time satisfies $t = \tau = w_0 l^z \sim l^z$.

This concludes our analytical study of irreversible collapse. This analysis has given us the tools to discuss the simulations in a way that will show us important features of the collapse. In particular we can study the behavior of the quantities $M(t)$, $M_0(t)$, r and l. Before we do so we address one of the questions we asked before about the simulations.

In our simulations, during polymer collapse we eliminated the excluded volume. Isn't the excluded volume important for collapse? We know that excluded volume is relevant to the initial polymer conformation in good solvent. Moreover, excluded volume is relevant to the final collapsed state of the polymer—without excluded volume the polymer collapses to a point. However, we see that excluded volume does not enter in the scaling argument leading to the relationship between s and s_0, Eq. (5.4.11). This argument describes the kinetics of collapse itself. Thus, we do not expect excluded volume to affect the behavior of the collapse. In particular, we do not expect it to affect the relationship between s and s_0. According to Eq. (5.4.11) this relationship depends only on the dimension d of the space. On the other hand, the value of s derived in Eq. (5.4.13) is dependent on the values of the exponents ν and z. This means that we can expect the precise values of s and s_0 to be somewhat more sensitive to the presence of an excluded volume. However, the range of possible values (Table 5.4.1) indicates that the overall behavior of the collapse should not be affected. In Section 5.4.5 we will describe simulations that have excluded volume, and find that the results are indeed similar. We can understand why the excluded volume is not important because during collapse the kinetics is primarily affected by the net attractive interaction between monomers, rather than the hard core repulsive part.

We can analyze the simulations to compare with the scaling argument by studying the average mass of the polymer except the ends, $M(t)$, and the mass of the ends, $M_0(t)$. The result of averaging many collapse simulations are shown in Fig. 5.4.2 for two and three dimensions. As indicated, two lengths of polymers were simulated in each case. Longer polymers follow the collapse of the shorter polymers but extend the curves to longer times. This is consistent with the picture of end-dominated collapse, where the only effect of a longer polymer is increasing the length of time till the end aggregates meet in the middle.

$s = (1 - \nu)/z$	$z = 3\nu$	$z = 2\nu + 1$
2 − d: $\nu = 0.75$	0.11	0.10
3 − d: $\nu = 0.6$	0.22	0.18
$\nu = 0.5$	0.33	0.25

Table 5.4.1 Values of the exponent s from the scaling relation Eq. (5.4.13) using different assumptions for ν and z. $\nu = 0.5$ would occur for a random walk without excluded volume in any dimension. The other values of ν are for self-avoiding random walks. $z = 3\nu$ is for Zimm relaxation that includes hydrodynamics. $z = 2\nu + 1$ is Rouse relaxation that does not include hydrodynamics. All of these values are small and indicate that s_0 is much larger. ∎
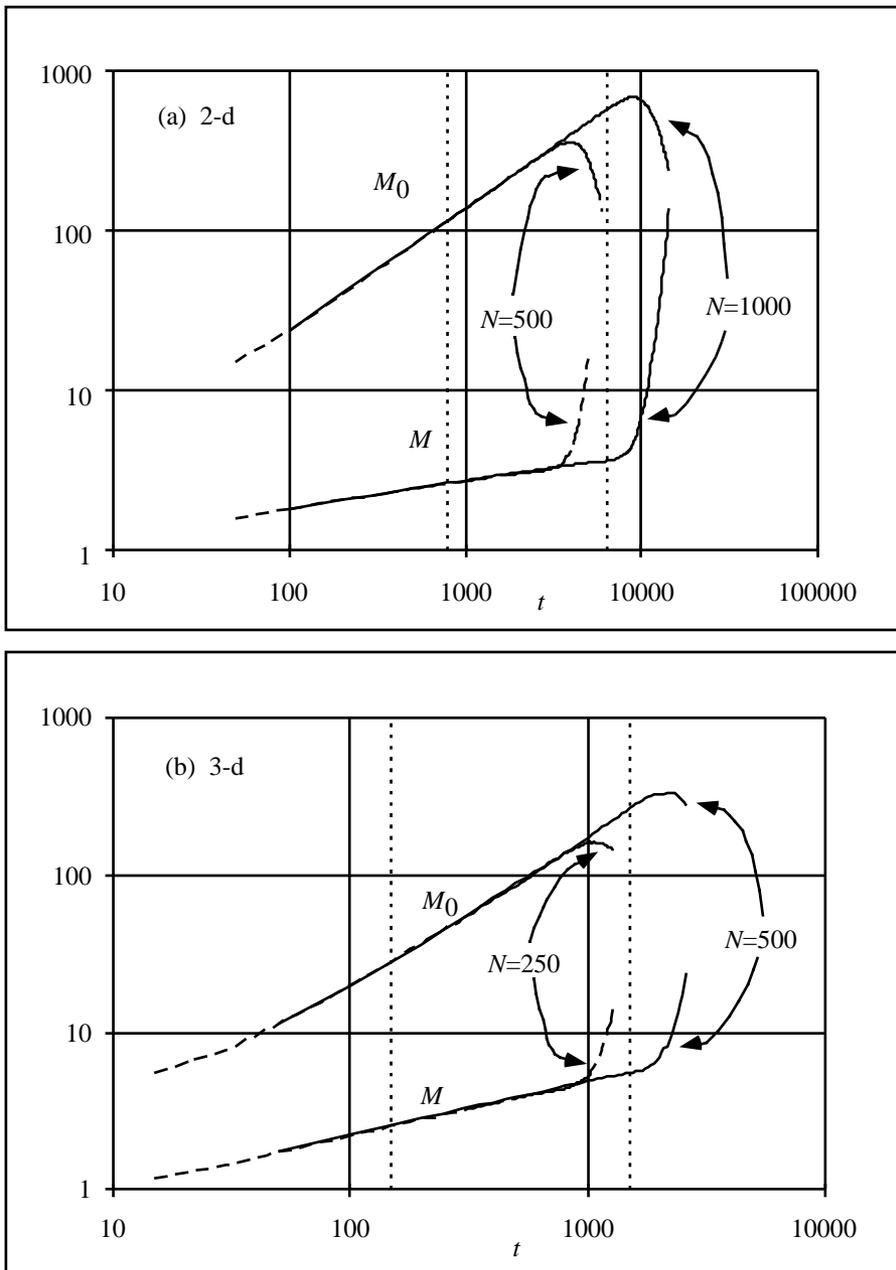
Figure 5.4.2 Plot of the time evolution during polymer collapse of the average total mass of the polymer ends $M_0(t)$, and of the average mass $M(t)$ of aggregates not including the ends. (a) shows collapse in two dimensions of polymers of length $N = 1000$ (averaged over 500 samples), and $N = 500$ (1000 samples). (b) shows collapse in three dimensions of polymers of length $N = 500$ (500 samples), and $N = 250$ (1000 samples). Scaling exponents fitted to the longer polymer collapse for times between the vertical dashed lines are given in Table 5.4.2 and discussed in the text. ∎

Both $M(t)$ and $M_0(t)$ follow a power-law scaling behavior. Exponents from the lines in Fig. 5.4.2 are given in Table 5.4.2. There are two results from our scaling analysis that we can compare with. The more reliable derivation is that of the relationship between s and s_0. A comparison is made in Table 5.4.2 by calculating the expected value of $s_0(s)$ using the scaling relation, Eq. (5.4.11), and the measured value of s. This is compared with the measured value of s_0. The agreement with the scaling relationship is striking since there are corrections which may be expected due to the neglect of the effects of small rings, or changes over time in curvature and compression of the polymer. The statistical errors are smaller than the quite small difference between the expected and measured value of s. This difference, in principle, might be real but could also be due to systematic error from the use of polymers that are not long enough to determine this level of precision.

The comparison of the value of s obtained from the simulation with the values predicted by the scaling argument (Eq. (5.4.13)) show that there is qualitative but not quantitative agreement. The value in two dimensions does not agree with that expected—it is halfway between the values expected in two and three dimensions. The value in three dimensions is very close to that expected without excluded volume ($\nu = 0.5$) and with Zimm relaxation. There are two ways to discuss this: one is to downplay the success and the other to downplay the failure. All the values of the exponent s are small, and this suggests that the general discussion in the scaling argument is essentially correct. We could also excuse the disagreement in two dimensions because we are using Stokes' law to move aggregates, and hydrodynamics is not well behaved in two dimensions (see the end of Section 5.4.2). The estimate given for three dimensions using Zimm relaxation and without excluded volume $s = 1/3$ is in coincidence with the simulations. It is hard to believe that this result can be justified except as a coincidence. While it is true that we implement hydrodynamics by performing moves of aggregates according to Stokes' law, and that we have performed the simulation without excluded volume, nevertheless, the initial conformation of the polymer should set the distances that control the collapse time. This initial polymer conformation is for a polymer with excluded volume. This would suggest that a

	s	s_0	$s_0(s)$ [Eq. (5.4.11)]	$s_0 - s_0(s)$
1-d	0	0.5	0.5	0
2-d	0.154±0.001	0.7734±0.0006	0.7695±0.0006	0.004±0.001 (0.5%)
3-d	0.337±0.002	0.982±0.005	1.003±0.002	−0.021±0.005 (2%)

Table 5.4.2 Power-law exponents for the scaling of end mass (s_0) and mass along the contour (s) during polymer collapse. The first column gives the dimension of space. The second and third columns are fitted to the simulation results between the dashed lines in Fig. 5.4.2. Fits were chosen to minimize standard errors. Errors given are only statistical—they reflect the standard deviation of the simulation data around the fitted line. The simulation results are compared to the scaling relation, Eq. (5.4.11), in columns four and five. Results in one dimension are exact, since there is no possibility of collapse along the contour, $s = 0$. ∎

smaller value of s should be expected. For our purposes, the precise value of s is not essential, but the central result of the end-dominated collapse is.

Finally, to investigate the validity of the universal scaling of polymer smoothing, Eq. (5.4.14), we plot in Fig. 5.4.3 values of r/l against the rescaled time, $w = t/l^z$, with $z = 3\nu$. This is a way of showing directly the function $f(w)$. What is important in this figure is that all the different values lie along the same curve. This is the significance of the universal scaling function. The generally good coincidence of the different curves confirms that the simulation obeys Eq. (5.4.14). Moreover, we see that the function $f(w)$ approaches the expected value, 1, at large values of w, consistent with the polymer becoming straighter with time. If we try to change the assumptions we find a poorer fit. For example, assuming $r \sim l^{0.95}$ at long times or changing slightly the value of z would lead to visibly poorer coincidence of the curves. The fit in two dimensions is somewhat less precise. This may be due to the difficulties with modeling hydrodynamics in two dimensions, or it may be due to other effects we neglected such as ring formation during collapse.
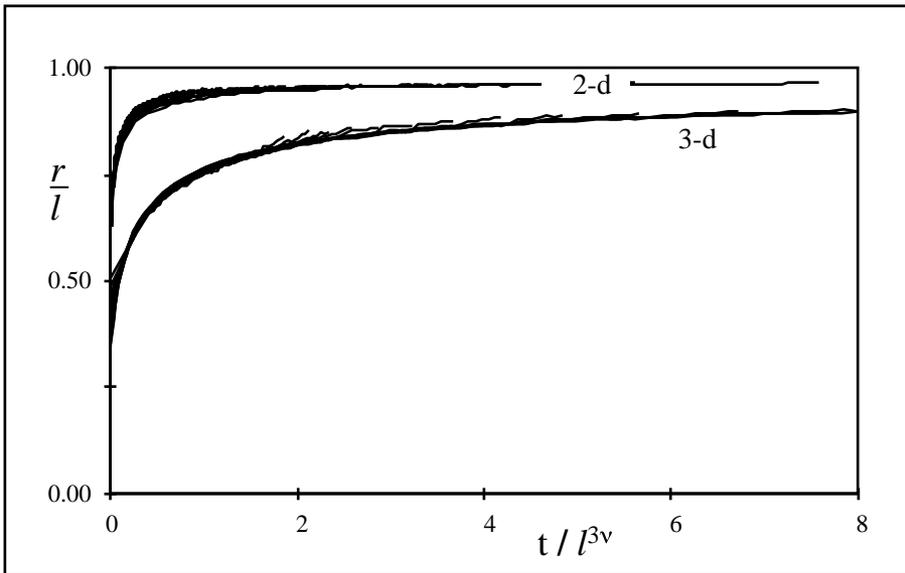
Figure 5.4.3 Plot of the rescaled end-to-end polymer segment distance, r/l, as a function of the rescaled time, $t/l^{3\nu}$. Results for both two and three dimensions are shown. In both cases the polymer contained 500 monomers and results were averaged over 200 collapses. The many curves in each case arise from different values of l. The coincidence of the curves is consistent with validity of the universal scaling relationship, Eq. (5.4.14). The plot shows the universal scaling function, $f(w)$. This function describes the structure of the polymer contour, not including the ends. For large w it approaches one, consistent with the expectation that the polymer contour becomes straight at long times. ∎

5.4.4 *Implications of end-dominated collapse*

We have found that polymer collapse in both two and three dimensions appears to reduce essentially to the behavior of a one-dimensional collapse. In one dimension we would have a starting conformation of a set of monomers along a line. Then, without excluded volume, the ends in each time step can move to occupy the same site of, and aggregate with, their neighbor. In each time step the end has equal probability to try to move away (it cannot) or to try to move on top of its neighbor. Thus, 50% of the steps, it aggregates with its neighbor. Collapse thus proceeds by the driven diffusion of the polymer ends. The driving force is the aggregation of the monomers. The only modification of simple diffusion is the change in the diffusion constant due to the accretion of mass onto the ends. In two and three dimensions, the results are similar. Ends diffuse along the contour accreting monomers and smaller aggregates.

We can analyze our results to give the scaling of the collapse time $\tau(N)$. In end-dominated collapse this is the time that passes until the end aggregates meet in the center. By this time the end aggregate has reached one-half of the mass of the whole polymer:

$$M_0(\tau(N)) \quad N \tag{5.4.15}$$

where we need to write only that the end mass is proportional to the number of monomers. Substituting Eq. (5.4.9) we obtain

$$\tau(N) \sim N^{1/s_0} \tag{5.4.16}$$

From the simulations we find that $1/s_0 = 1.293 \pm 0.001$ in two dimensions and 1.018 ± 0.005 in three dimensions (errors are statistical). Thus the collapse time is predicted to scale linearly with polymer length in three dimensions. This indicates that kinetic effects through end-aggregation accelerate the collapse from the usual relaxation time scaling of $\tau(N) \quad N^z$.

Even without a value of s_0 from the simulations, we can see that end-aggregation must accelerate the collapse from the equilibrium relaxation. The scaling relation Eq. (5.4.11) gives a minimum possible value for the exponent s_0. s_0 is a monotonic increasing function of s. The minimum value of s_0 results from setting $s = 0$ in the scaling relation. This corresponds to collapse without any aggregation along the contour. The only process that occurs is the increasing mass of the ends. The minimum value $s_0 = 2/3(3/4)$ in two (three) dimensions gives the slowest possible collapse, or the largest collapse-time scaling. Thus the maximum possible collapse-time scaling exponent $(1/s_0)$ is $3/2$ in two dimensions and $4/3$ in three dimensions. These exponents are still significantly smaller than the usual polymer relaxation-time scaling exponent, $z \quad 2$. They are also not dramatically different from the values we found using s_0 from the simulations.

When we simulate a system like a polymer, we are always concerned that our simulation results are characteristic only of the size simulated and not of the regime we are interested in. Our objective is to understand the properties of long polymers. Thus we must ask the question: Are the polymers simulated long enough to show the correct collapse mechanism for very long polymers? We can address this question

because we know how the collapse-time scales with the size of the system for end-dominated collapse. If another process were to become the dominant process for longer polymers, it would have to scale as an even lower power of N than the end-dominated collapse process. Only then could it become important for longer polymers. However, the scaling of the collapse time we found is significantly lower than other known collapse mechanisms. The mean field collapse scaling near the -point is similar to the usual polymer relaxation. Thus it is unlikely that for longer polymers a faster process will dominate.

How can we relate our discussion of polymer collapse to the problem of protein folding? Our results on the end-dominated collapse of polymers indicated that collapse can be faster than would be expected from the usual polymer dynamics. Instead of a scaling of $O(N^2)$ we have a scaling of $O(N)$. This appears to be a significant reduction in the time; however, when we consider this in light of our discussions in the previous chapter, we see that this is not as significant as we might expect. Both scaling exponents are reasonable for a polymer collapse time since there is enough time for either scaling to reach its conclusion. We allowed, in principle, for exponents up to $O(N^4)$. Even if $O(N^4)$ is generous, $O(N^2)$ appears quite possible.

Where does the difficulty then arise? The difficulty is that we have considered all of the collapsed polymer conformations as equally acceptable. All we have shown is that we can make a transition from expanded to collapsed polymers in a reasonable amount of time. We have not shown that we can select the right compact form of the polymer. This is where the simulations can provide their most important clue to the benefits of kinetics. End-dominated collapse does not give an arbitrary compact polymer. The process of sequential accretion of monomers and small aggregates by the ends should reproducibly yield a particular compact structure.

Unlike a general uniform collapse of the polymer, the end-dominated collapse proceeds by an orderly process of sequential monomer encounters. These encounters build up the aggregate compact structure (globule) in a manner that is not random. A consequence of this orderly kinetic process is that the resulting globules may be expected to be selected from a limited subset of all possible globules. This is precisely what we have been looking for—a kinetic process that might enhance the process of arriving at a specific folded protein structure. One of the most striking implications of the end-dominated collapse process is that the order of monomer encounters is essentially independent of the initial conformation of the polymer. It is not easy at this point to see what the precise nature of the globules that are formed are. However, we can note that they are likely to be formed out of two parts corresponding to the aggregate formed from one end and the aggregate formed from the other. A more subtle feature of this process is that the globule is likely to contain fewer knots than would be generally found in a globule. This is because the diffusive end motion tends to unknot the polymer, since the ends are passed through any knots rather than closing or tightening them. To discuss this formally would require defining knots in a polymer with free ends, which is a feasible but tricky task.

End-dominated collapse is also consistent with a model, called the molten globule model, that has been proposed for the kinetics of protein folding. It suggests that

there is a fast initial process of forming a compact globule followed by a rearrangement of the globule to form the final folded protein. Our simulations and scaling results describe the fast process by which a polymer makes a transition from an expanded form to a compact globule. The rearrangement process should take more time and is likely to be the limiting step in the formation of the protein. Unlike the collapse, this process requires segments of polymer to move around each other, which is a much more difficult dynamic process. The significance of the end-dominated collapse is that by preselecting the initial compact globule, the rearrangement process is shortened and does not necessarily explore all possible compact conformations of the protein before settling in the desired state.

There is another interesting feature of the end-dominated collapse relevant to proteins. Proteins, when they are formed, are not formed all at once. Instead the chain emerges sequentially from a ribosome. This process is called extrusion. It is thus quite likely that the protein in the cell (*in vivo*) performs much of the folding sequentially as it is formed. In many ways this is similar to the sequential process of end-dominated collapse. Thus the polymer during extrusion has a natural sequential process for forming a definite final folded structure. The appearance of end-dominated collapse in polymer simulations may simply reveal why folding also works when the protein is unfolded and refolded *in vitro*. One possible difference between the two environments is that the collapse process *in vitro* occurs from both ends while during extrusion it occurs from one end only.

We note that it has yet to be demonstrated experimentally that kinetics plays a significant role in protein folding, or in other processes like DNA aggregation. There is an interesting consequence of end-dominated collapse that has relevance to experimental tests. End-dominated collapse represents a significant departure from the usual rule of thumb that linear and ring polymer dynamics are similar. The primary other exception to this rule is reptation in polymer melts. When polymers are placed together at high density, the resulting fluid is called a melt. The motion of polymers in the melt is inhibited by entanglements. Essentially the only way a polymer can change position is to move along its own contour by local stretching and contraction. This is a process, called reptation, that is possible for polymer chains. For rings it is not. Other processes must become relevant for rings, and motion should be much slower. The simulations and scaling argument we have described in this chapter indicates that ring collapse should be significantly slower than linear polymer collapse. This is one of the possible ways that the predictions of these simulations could be tested by experiment.

5.4.5 *Variations in the polymer microstructure*

In this section we discuss additional simulations of homopolymer and heteropolymer collapse. The objective is to investigate how robust are the results we have found in the simple simulations discussed in Section 5.4.2 and the scaling argument in Section 5.4.3. If the results are sensitive to the choice of model, then we should doubt their applicability to real polymers. On the other hand, if the results are robust then we can feel confident that they will also be relevant to real polymers. For a variety of

systems the results suggest that collapse of long polymers is dominated by diffusion of the polymer ends, which accrete monomers and small aggregates. Collapse does not proceed uniformly along the polymer. However, in a model where *only* pairwise bonding is allowed, the collapse is uniform, since more flexible end motion does not result in continued end accretion.

All of the simulations in this section are in three dimensions and are based on the one-space cellular automaton algorithm described in Section 5.3. The one-space algorithm uses a $5 \times 5 \times 5$ bonding neighborhood for monomers. It allows adjacent monomers to separate by one lattice site providing flexibility in the polymer dynamics. Motion of monomers is performed by Monte Carlo steps that satisfy the polymer constraints.

We summarize briefly the general process of simulation of collapse, then discuss each of several models that test various aspects of the dependence of the results on the local properties of the polymer. As with the two-space simulations, the simulation of collapse starts from a set of equilibrium polymer configurations. Each of the models consists of a particular scenario for monomer-monomer sticking whereby aggregates are formed from individual monomers. In all models, once formed, aggregates are moved as a unit. The collapse simulations include only aggregation, and not disaggregation. As discussed in Section 5.4.3, the primary effect of hydrodynamics is included by scaling the diffusion constant of aggregates by Stokes' law. For the three-dimensional simulations described here, $D \quad 1/M^{1/3}$. Polymer dynamics are simulated by selecting an aggregate (monomers are induded as aggregates of mass 1) and moving the aggregate in one of the four compass directions with a probability given by the diffusion constant, and only if connectivity constraints allow—the aggregate does not leave any neighbors behind. One time interval consists of performing a number of aggregate moves equal to the number of remaining aggregates, taken to be the number at the end of the time interval.

Six different models of polymer microstructure are described in the following numbered paragraphs. These simulations are also compared with the two-space collapse simulations that did not include any form of excluded volume during collapse. All of the six models were simulated using polymers of length $N = 250$. The first two models explore variations in the collapse of homopolymers. The third and fourth models explore heteropolymer collapse. The results of these four models are shown in Fig. 5.4.4. The last two models only allow pairwise bonding. Results of these two models are shown in Fig. 5.4.5.

1. The first version of one-space collapse explores the significance of excluded volume. Since excluded volume is essential for the final structure of the polymer as well as the initial structure, it might be expected to be relevant to collapse. This was completely neglected in the two-space simulations. In these one-space simulations, during collapse, monomers are allowed to enter the bonding neighborhood. However, excluded volume is maintained during collapse by preventing monomers from occupying the same lattice site. Monomers aggregate by moving adjacent (NSEW) to other monomers. As before, aggregates are moved as a unit
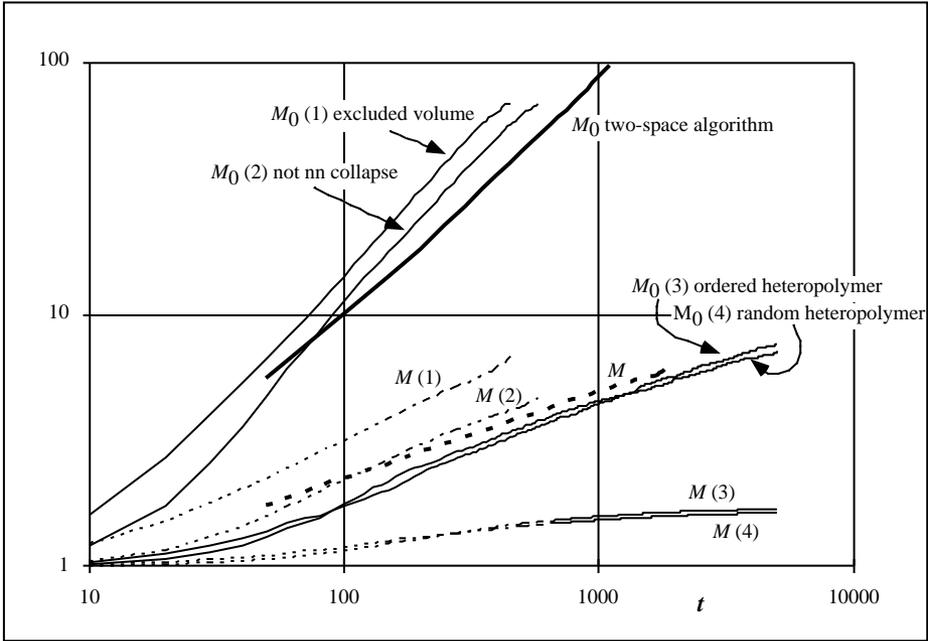
Fig. 5.4.4 Plot of the time evolution during polymer collapse in three dimensions of the average total mass of the polymer ends, $M_0(t)$, and of the average mass, $M(t)$, of aggregates not including the ends. The different lines correspond to the different models described in the text. The end mass evolution is shown by a solid line and the mass along the contour is shown as a dashed line. The bold curves are for the two-space algorithm (Fig. 5.4.2). All of the others are for variations on the one-space algorithm for polymers of length $N = 250$ and are averaged over 300 simulations: (1) for the one-space polymer including excluded volume; (2) same as (1) but preventing aggregation of nearest neighbors along the contour; (3) a model for heteropolymer collapse where only the odd monomers aggregate; (4) is similar to (3) but monomers that aggregate are selected at random. ∎

using a step probability governed by Stokes' law. This means that we must keep track of each aggregate's structure in space, and move it as a unit. After a step we must check all sites at the boundary of the aggregate in order to perform additional aggregation. From the figure we see that the inclusion of excluded volume appears to increase the rate of collapse. Conceptually, we might think about this as resulting from a decrease in the distance monomers need to travel in order to aggregate. Both exponents s and s_0 increase slightly. However, we see that excluded volume does not affect overall behavior and does not even dramatically change the values of the exponents (Fig. 5.4.4 and Table 5.4.3).

2. One of the strange features of model (1) is that monomers aggregate by directly attaching to their neighbors along the chain. Thus, much of the aggregation

	s	s_0	$s_0(s)$	$s_0 - s_0(s)$
1-d	0	0.5	0.5	0
2-d	0.154±0.001	0.773±0.001	0.769±0.001	0.004±0.001 (0.5%)
3-d	0.337±0.002	0.982±0.005	1.003±0.002	-0.021±0.005 (2%)
3-d (1)	0.484±0.002	1.102±0.004	1.113±0.002	-0.011±0.004 (1%)
3-d (2)	0.453±0.002	1.079±0.004	1.090±0.002	-0.011±0.004 (1%)
3-d (3)	0.061±0.001	0.363±0.001	0.796±0.001	
3-d (4)	0.050±0.001	0.293±0.001	0.787±0.001	

Table 5.4.3 Power-law behavior exponents fitted to the simulation results of Fig. 5.4.5. The results from the two-space algorithm (Table 5.4.2) are included for comparison in the first three lines. Fits were chosen to minimize standard errors. Errors given are only statistical. Results are compared to the scaling relation given by Eq. (5.4.11). As discussed in the text, the heteropolymer collapse results are not in agreement with this scaling relation. ∎

occurs between monomers that previously were already bonded as neighbors. This would be a particularly convenient process for end aggregation and may exaggerate its importance. The aggregated structure of model (1) is formed out of rodlike structural components. These rods arise because the attachment to nearest neighbors follows the contour of the polymer. A more realistic model would exclude such nearest-neighbor aggregation. The second version of one-space collapse is similar to model (1) except that nearest neighbors along the contour are prevented from aggregating to each other. Monomers or aggregates are forced to move around their nearest neighbor to bond to a monomer further along the chain. This prevents the simplest end monomer accretion of nearest neighbors. Simulations show, however, that not only does the end-domination persist but (Fig. 5.4.4 and Table 5.4.3) that this change does not change significantly the exponent values. Presumably this is because the need to move around neighbors to aggregate affects collapse along the contour similarly to its affect at the ends, making both more difficult. The only apparent effect is an overall slowing of the collapse as seen by the shift of the $M(t)$ and $M_0(t)$ curves to longer times. The following simulations (3)–(6) are based on model (2).

3. We now consider heteropolymers. Thus far our discussions, both in simulations and in scaling arguments, have not distinguished between different monomers. However, there are significant differences in the bonding of different monomers in heteropolymers. In order to investigate the effect of such variation, we take an extreme case where there are some monomers that bind and some that do not bind at all. This is a simple model of proteins that takes into account the difference between hydrophobic and hydrophilic monomers. In our language, this is the same as monomers that want to collapse by aggregation and monomers that do not. The third one-space model of collapse includes both kinds of monomers. Using the one-space algorithm as in (2), only odd monomers are allowed to aggregate. This is an ordered 50% hydrophilic version of this model of hy-

drophilic/hydrophobic collapse. The way the collapse works is that monomers that do not want to bind at all must be on the outside of an aggregate. The aggregate has an interior filled with bonding monomers and a surface of nonbonding monomers. These nonbonding monomers prevent further aggregation and thus slow the continued formation of larger aggregates. The process of increasing shielding is not included in our scaling arguments, so Eq. (5.4.11) should not be expected to apply, and it doesn't. Despite the shielding of continued collapse by nonbonding monomers, the collapse is still dominated by end motion for the length and time scale simulated. The overall collapse is significantly slowed—the exponent s_0 is reduced to a third of its value for the homopolymer case. There is also some indication that the collapse would completely saturate in this case and would not go through to completion. The overall fraction of hydrophobic (nonbonding) monomers must be reduced to reach complete aggregation.A reduction in the proportion of hydrophobic monomers would make the simulation results more like the previous homopolymer models.

4. The fourth version of one-space collapse is similar to the third version but tests the relevance of the order in model (3). Instead of alternating the hydrophobic and hydrophilic monomers along the chain,they are placed at random along the chain. The collapse behavior is almost the same as in (3). Opportunistic collapse, which results from convenient local arrangement of bonding monomers,speeds the collapse at first. However, the scaling of the masses is slightly lower, and eventually collapse is slightly slower at later times.

5. In order to make end-dominated collapse as unfavorable as possible, we must eliminate the advantage that is gained by the high mobility of the ends. We can do this by eliminating entirely the continued aggregation. The fifth version of one-space collapse starts from the same conditions as (2);however, collapse only includes pairwise bonding. Once two monomers are bonded, other monomers that become adjacent are not aggregated. This is not like protein folding, since each amino acid can form two hydrogen bonds with other amino acids. Also, there is additional bonding due to van der Waals forces. Nevertheless, we can consider the pairwise bonding model as a version of collapse.End dominance in this case would correspond to a progressive pairing from the ends inward. A plot of the pair-density along the chain (Fig. 5.4.5) shows that,despite a tendency toward more rapid pairing at the ends, the collapse is essentially uniform. Thus in this model we do not have end-dominated collapse.

6. The sixth and final version of collapse (Fig. 5.4.5) is like the fifth version where pairwise bonding is allowed;however, only even-odd monomer combinations are allowed for pairwise bonding. The results do not differ significantly from (5).

We have learned from these simulations that in heteropolymer collapse the values of the exponents s and s_0 may change; however, only when we go to an extreme and unrealistic model of polymer aggregation do we lose the end-dominated collapse entirely.
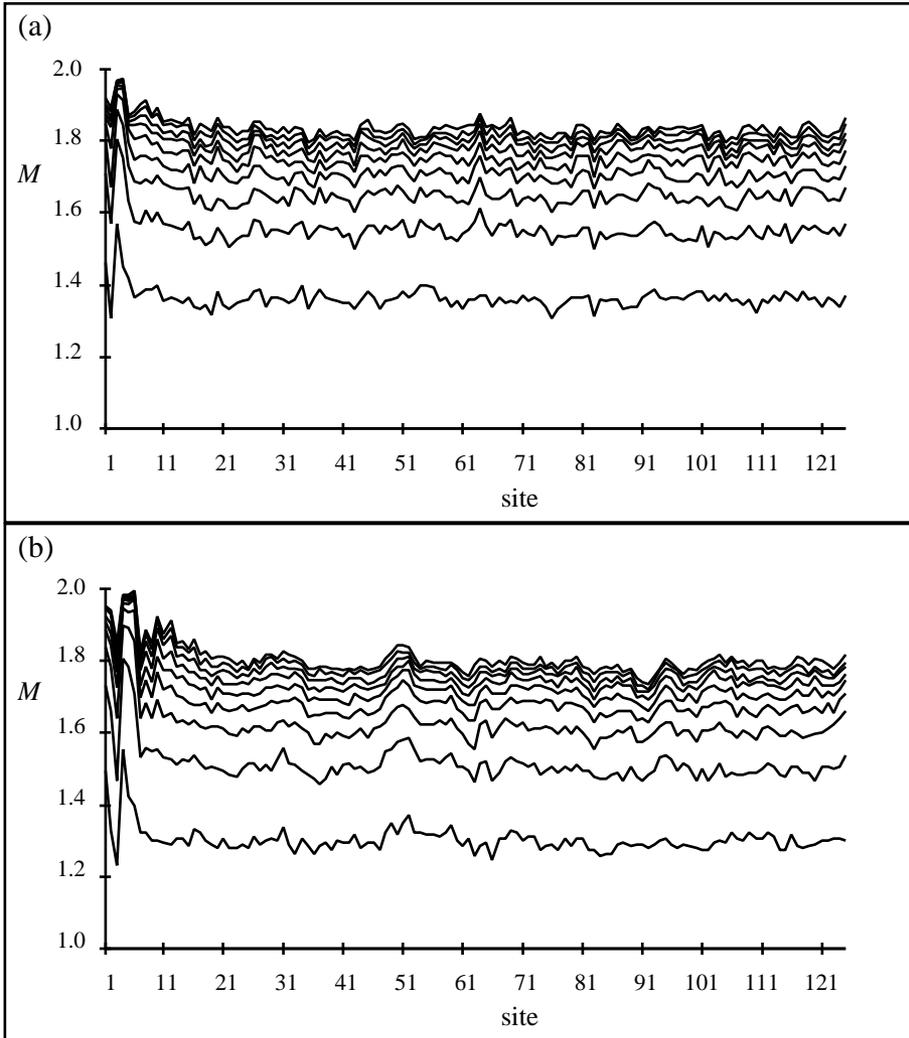
Figure 5.4.5 Plots of the average aggregate mass at every site along the contour when collapse only allows pairwise bonding. Each curve shows the averages at a particular time, with later times having higher masses. Collapse is uniform except for a tendency for the monomers near the end to pair up first. The minimum mass is 1 and the maximum mass is 2. The collapsing polymer has a length $N = 250$. By symmetry only the first 125 sites are shown. (a) shows the case (5) of arbitrary monomer bonding with the exception of no nearest-neighbor bonding; (b) shows the case (6) of only odd to even monomer bonding. Progressively later times are shown separated by (a) 50 updates and (b) 100 updates. For clarity only the first ten times are shown. ∎

The scaling time $\tau(N) \sim N^{1/s_0}$ can also be calculated for the first four one-space models of collapse. For homopolymer collapse $1/s_0 = 0.907 \pm 0.003$ in model (1) and, 0.927 ± 0.003 in model (2). Thus the collapse time still scales approximately linearly with polymer length in three dimensions. In contrast, the time for heteropolymer collapse with nonbonding monomers scales as $1/s_0 = 2.75 \pm 0.007$ in model (3) and 3.41 ± 0.008 in model (4). This is much slower than the heteropolymer collapse. It also approaches the limit of our allowed exponents for protein folding discussed at the beginning of Chapter 4.

Question 5.4.2 One way to think about the effect of various properties of the microstructure of a polymer is as a change in how the diffusion constant of an aggregate grows with its mass. For example, if there are monomers that do not aggregate at all, they become like a surface coating on an aggregate that prevents further aggregation. Rather than model this as a limitation in aggregation, we could simplify the effects by modeling them as a progressively larger diffusion constant that would also limit continued aggregation. Find the relationship between s and s_0 for different values of x in $D \sim 1/M^x$. Then simulate the collapse of a polymer for different values of x and see if the scaling relationship between s and s_0 continues as before. How should this affect the value of s?

Solution 5.4.2 The generalized form of Eq. (5.4.11) is:

$$s_0 = (s + 1)/(x + 1) \qquad (5.4.17)$$

Simulations of collapse when x is varied are shown in Fig. 5.4.6. The end-dominated collapse occurs for all values of x that are simulated. The scaling relationship continues to be satisfied.

The size of s appears to be nearly constant. A quite reasonable value for s_0 may be obtained from a single value for s in each dimension and use of the scaling relation Eq. (5.4.17). In three dimensions, s appears to decrease slowly with increasing x. ∎

5.4.6 *Conclusions*

In this chapter our objective was to understand how kinetic processes could accelerate the formation of a selected final polymer structure. We found that there is a distinctive kinetic process that occurs in the initial stages of polymer collapse that results in a characteristic order of monomer-monomer encounters. This suggests that it might be possible to design a sequence of amino acids that fold into a particular structure using this order of events. The polymer might not reach the desired structure if the polymer were to explore all possible conformations. We have not demonstrated that this process applies specifically to proteins, but the robustness of the end-dominated collapse to variations in models suggests that it should play some role. Our analysis has been specific to polymers. How can we generalize this discussion to apply more generally to complex systems? The most important feature of these simulations

(a) 2-d

(b) 3-d
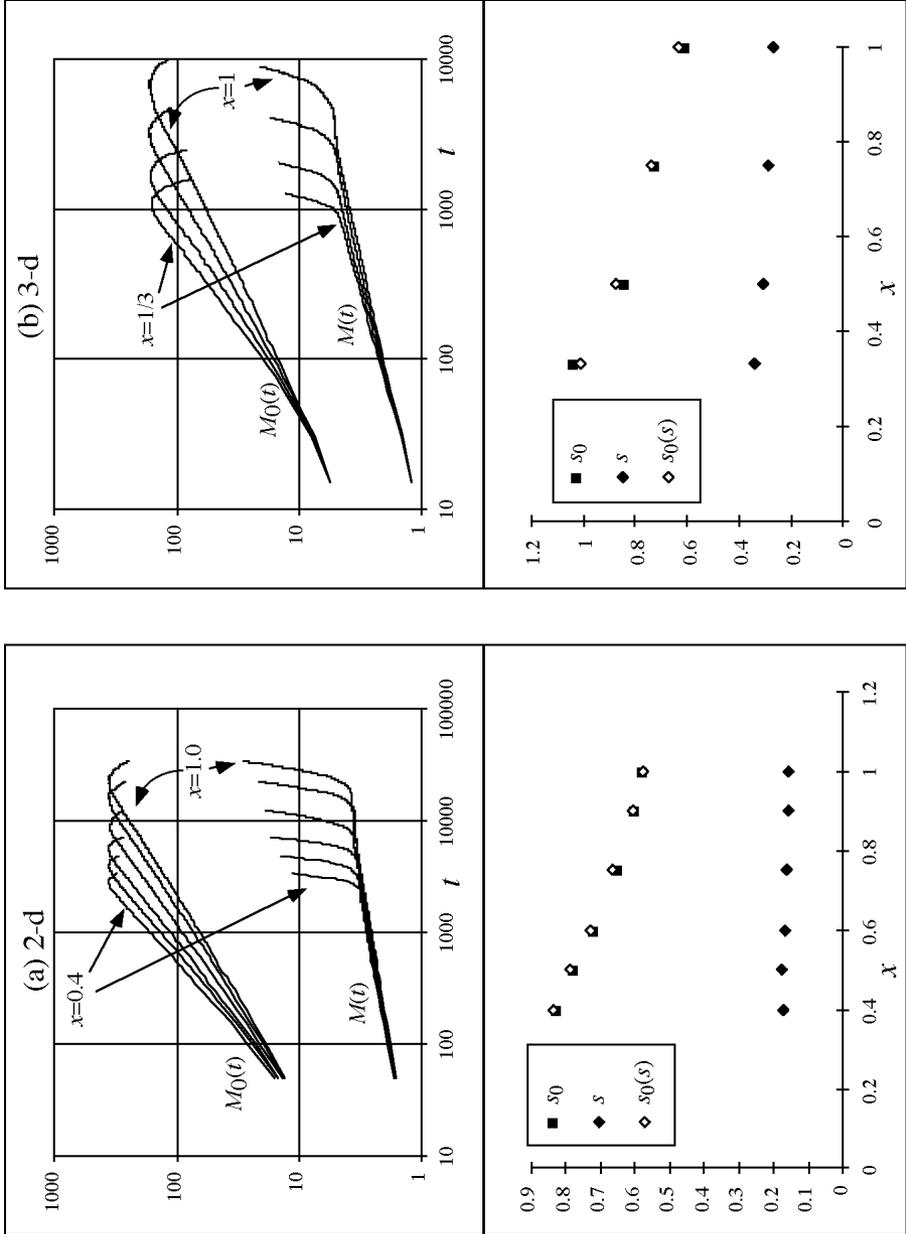
Figure 5.4.6 Exploration of the variation of polymer collapse with x, where the diffusion constant scales as $D \sim 1/M^x$. In the top panels of (a) and (b) are plots of the time evolution during polymer collapse of the average total mass of the polymer ends $M_0(t)$ and of the average mass $M(t)$ of aggregates not including the ends. (a) shows collapse in two dimensions of polymers of length $N = 250$ (500 samples) for $x = \{0.4, 0.5, 0.6, 0.75, 0.9, 1.0\}$. (b) shows collapse in three dimensions of polymers of length $N = 250$ (500 samples) for $x = \{0.5, 0.75, 1\}$. The value at $x = 1/3$ is from polymers of length $N = 500$. In the bottom panels the scaling exponents s_0 and s obtained from fits to the plots of $M_0(t)$ and $M(t)$ are plotted as a function of x. The value of $s_0(s)$ obtained from the scaling relation Eq. (5.4.16) is also plotted. It is in good agreement with the values of s_0 found in the simulations. ∎

is the recognition that there may be a natural sequentiality to events. Such sequentiality does not necessarily mean that a desired structure will be attained; however, it provides an opportunity for control over the final structure. This leads us to the topic of self-organization and organization by design, which we address in the following two chapters.