# Kinetics of Internal-Loop Formation in Polypeptide Chains: A Simulation Study

Dana Doucet,* Adrian Roitberg,† and Stephen J. Hagen*
*Physics Department, and †Chemistry Department and Quantum Theory Project, University of Florida, Gainesville, Florida

ABSTRACT The speed of simple diffusional motions, such as the formation of loops in the polypeptide chain, places one physical limit on the speed of protein folding. Many experimental studies have explored the kinetics of formation of end-to-end loops in polypeptide chains; however, protein folding more often requires the formation of contacts between interior points on the chain. One expects that, for loops of fixed contour length, interior loops will form more slowly than end-to-end loops, owing to the additional excluded volume associated with the "tails". We estimate the magnitude of this effect by generating ensembles of randomly coiled, freely jointed chains, and then using the theory of Szabo, Schulten, and Schulten to calculate the corresponding contact formation rates for these ensembles. Adding just a few residues, to convert an end-to-end loop to an internal loop, sharply decreases the contact rate. Surprisingly, the relative change in rate increases for a longer loop; sufficiently long tails, however, actually reverse the effect and accelerate loop formation slightly. Our results show that excluded volume effects in real, full-length polypeptides may cause the rates of loop formation during folding to depart significantly from the values derived from recent loop-formation experiments on short peptides.

## INTRODUCTION

The formation of contacts between the residues of a disordered polypeptide chain by diffusion constitutes one elementary step in the folding of a protein, a process that can occur in a time as short as a few microseconds (1). The speed at which these intrachain loops form, especially in short polypeptides, has therefore been interpreted as a physical upper limit for the possible speed of protein folding (2). A number of studies have used laser pump-probe spectroscopy or fluorescence correlation spectroscopy to measure this timescale; for sufficiently short, flexible polypeptides the contact time approaches $\sim 10^{-8}$–$10^{-7}$ s (3–7). Loop-formation kinetics and statistics have also been addressed in theory and in simulation, both for polypeptide chains and for more general polymers (8–26).

Such studies have primarily focused on the formation of "external" or end-to-end loops, in which the two endpoints of a chain diffuse into contact with each other. In the folding of a protein, however, a more relevant case is the appearance of "internal" loops, in which two interior points on the chain, distant from the chain termini, make contact (Fig. 1 *A*). This is also the more important case in other biomolecular phenomena, such as DNA looping, which can play a key role in transcriptional regulation and other aspects of gene expression and replication (27). (The end-to-interior loop is of course a third case.) In general one expects that internal loops will form more slowly than end-to-end loops of equal contour length, if only because the additional residues extraneous to the loop contribute excluded volume in the vicinity of the two contacting residues, thereby reducing their probability of interaction. The "tails" may have additional, purely dynamical effects as well. Such considerations suggest that studies of end-to-end loop formation may overestimate the rates of the diffusional motions relevant to structure formation during early stages of folding.

We have conducted simulations to estimate the magnitude of this effect. Our objective is to determine how the rate of formation of an intrachain loop is affected by the addition of tails, or further chain segments, external to the loop. We calculate loop formation rates as follows: first, we generate a random ensemble of hard-sphere, freely jointed chains that satisfy an excluded-volume condition (Fig. 1 *B*); second, we extract from these ensembles the probability distribution, $P(r)dr$, for the separation $r$ between two residues on the chain; third, we apply the first-passage-time theory of Szabo, Schulten, and Schulten (the SSS theory (23)), which calculates the product $D\tau$ from the distribution $P(r)$. Here, $\tau$ is the mean time to formation of the intrachain contact, and $D$ is the effective diffusion constant for reconfigurations of the chain. Therefore, we simulate the equilibrium probabilities for the chain configurations, and SSS theory provides the loop-formation kinetics from those probabilities. As expected, converting an external loop to an internal loop by adding even a few additional chain segments can significantly slow the kinetics of loop closure. Interestingly, however, the slowing effect is more pronounced for loops of greater contour length. We also find that as the tail grows sufficiently long, it can actually enhance the loop formation rate. These findings constitute experimentally testable predictions for the SSS theory. Our calculations also show that contact formation times predicted by the full SSS theory can deviate significantly from those obtained in a widely used

*Submitted June 28, 2006, and accepted for publication November 28, 2006.*

Address reprint requests to Stephen J. Hagen, Physics Department, University of Florida, Museum Road and Lemerand Drive, PO Box 118440, Gainesville, FL 32611-8440.



0006-3495/07/04/2281/09 $2.00

doi: 10.1529/biophysj.106.092379

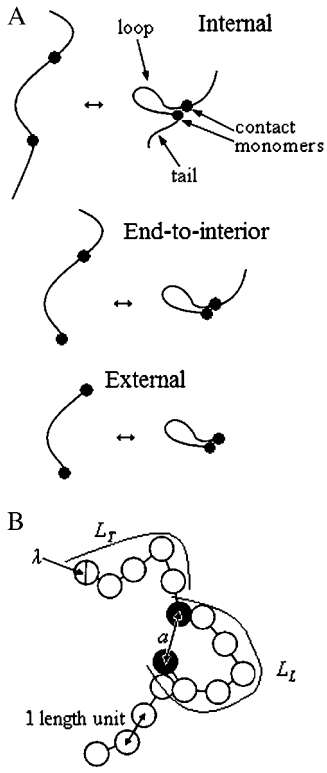


FIGURE 1 (*A*) Three different cases for loop formation in a diffusing polymer chain. The presence of tails of additional residues at one or both ends of the loop converts an external loop to an end-to-interior or internal loop, respectively. $\tau$ is the average time for formation of a contact between a particular pair of monomers. (*B*) Freely jointed chain with excluded volume. The bonds linking consecutive monomers on the chain are of unit length, and each monomer is a hard sphere of diameter $\lambda$. The tail length $L_T$ is the number of monomers in the tail. The loop length $L_L$ is the number of monomers that make up the loop, including the contact monomers. A loop is formed when the centers of the monomers of interest have a distance $a$ or less between them.

''approximate'' SSS theory, especially for the very short chains that are most commonly studied in experimental work. These results show that investigations of loop formation in polypeptides may be subject to significant excluded-volume effects: even the attachment of bulky photosensitive or fluorescent groups to the chain termini, to serve as spectroscopic reporters of contact formation, could introduce enough excluded volume to alter the dynamics that the experiment aims to study. They also indicate that the approximate form of SSS theory should be applied with caution, especially in studies on short chains.

## Previous experiments and simulations

A number of spectroscopic studies have examined contact formation in polypeptide chains. Some of the first were performed by Haas et al. (28,29), who used fluorescence resonance energy transfer to measure the probability distribution $P(r)$ for interresidue distances in synthetic oligopeptides, and to estimate the effective diffusion constant $D$ for reconfigurations of an unfolded protein (29,30). Time-resolved studies of the intrachain ligation of unfolded cytochrome *c*, a very short-range intrachain reaction, allowed more confident estimates of the speed of contact formation in disordered polypeptides (7). Bieri et al. (5) and, later, Lapidus et al. (6) initiated the use of triplet photoexcitation and quenching to study contact-formation in short, synthetic peptides. These studies showed that the contact formation rate scales approximately as the $-3/2$ power of the number of peptide bonds in the loop, with a limiting contact-formation rate exceeding $10^7\ s^{-1}$ for the shortest polypeptides. More recent studies have examined the effects of such parameters as amino acid composition, temperature, solvent composition and viscosity, and secondary-structure tendencies (4,17,25,25,31–36).

The differences between internal and external loop dynamics are much less well studied. Lee et al. (32) used fluorescence and triplet energy transfer methods to investigate contact formation between various pairs of residues within a disordered protein; some of these contacts resulted in internal loops. In fitting the decay curves to different models, they calculated a slower diffusion constant in the formation of these loops. They interpreted this as possible evidence for a drag (dynamical) effect associated with the external segments. Buscaglia et al. (25) used fluorescence resonance energy transfer to determine contact formation rates for end-to-interior loops. They added a tail of one residue to the end of a loop of 11 residues. They found that this decreased the rate of loop formation by a factor of 0.7.

The theory and simulation literature contains a number of studies of contact formation in polymer chains (8–26,37). If $N$ is the number of residues in a closed loop, the fraction of chains (at equilibrium) that contain such a loop is expected to scale as $N^{-\alpha}$, where $\alpha$ is the scaling exponent, which is found to vary in the different cases of internal and external loops. Wittkop et al. (21) and Redner (24) provide an extensive set of references on calculations of these exponents, along with their own estimates. Here we summarize only some of these earlier results. Chan and Dill (11,12) exhaustively enumerated all the configurations of a flexible chain on a cubic lattice to calculate the probabilities for loop formation in three-dimensional chains. They obtained values of $\alpha \approx 1.99$ for external loops, 2.18 for end-to-interior loops, and 2.42 for internal loops, demonstrating significant differences in the statistical behavior of the three cases. This presumably implies different loop-closure kinetics as well, although these equilibrium exponents do not directly predict the kinetic behavior.

Sheng et al. (14,15) used Monte Carlo simulations to estimate the loop-formation probabilities for internal and external loops in freely jointed chains. In their calculations, a designated pair of residues interacts via a strong and short-range attraction. Each chain conformation generated in the simulation is then classified as belonging to one of two states: a loop closed at those two residues, or else an open

coil. By fitting the $T$-dependence of the ensemble-average state (i.e., loop or coil) to a two-state melting transition, these authors obtained the entropic cost $\Delta S$ for closing a loop: they could then determine the scaling exponents $\Delta S \propto \alpha \log(N)$. They found values very similar to those of Chan and Dill (11,12). They then used the loop-coil fluctuation dynamics to estimate the scaling behavior of the contact-formation rates, based on an elementary entropic-barrier activation model of the contact-formation dynamics.

However, the Sheng et al. analysis does not directly reveal the effect on the loop-formation rate of simply converting an external loop to an internal loop by addition of a few tail segments. Further, their model for the dynamics relies on a simple two-state assumption that, though satisfactory for calculating the entropic cost of loop formation, does not take into account the influence of the distribution of chain conformations at equilibrium; however, in a first-passage-time theory, the shape of that distribution plays a critical role in the loop-closure dynamics.

Hyeon and Thirumalai (37) used a first-passage-time approach to calculate the rate of formation of interior contacts in semiflexible polymers. They derived a probability distribution for the distance between pairs of interior points in a wormlike chain, and from this distribution they generated a potential of mean force for the chain dynamics. By then developing a Kramers theory for passage over the potential barrier, they calculated the variation in the contact-formation time as a function of the chain stiffness (i.e., persistence length). The model focuses on the role of the polymer's persistence length, however, and does not account for the excluded volume of the chain itself. For this reason, the probability distribution was insensitive to the presence of tail residues outside the core loop, and the presence of tails does not suppress loop formation.

Buscaglia et al. (25) implemented the Wilemski and Fixman theory (using an expression similar to our Eq. 2 below) to calculate loop-formation rates for wormlike chains that were subject to an excluded-volume constraint. Their simulations showed that converting the 11-residue external loop to an end-to-interior loop decreased the contact formation rate. However, although their results illustrate the potential effect of excluded volume on loop formation, their data and simulations lack the resolution and scope to allow any detailed conclusion about the role of parameters such as tail length, loop contour, and contact placement (i.e., end-to-interior versus internal) in this effect. These parameters are important in the practical problem of loop formation in protein folding (for example), and therefore their role is the focus of this study.

## METHODS

### SSS theory

Although there is no universally accepted theory for the calculation of loop-formation times in polymer chains, the first-passage-time theory (SSS theory (23)) has been tested through comparisons to Langevin dynamics simulations of both Rouse chains (18) and realistic short peptides (33). Such comparisons have shown that the one-dimensional diffusion approach of SSS describes the dynamics of the peptide chain with good accuracy, and so the theory should provide useful insight into the effects of different loop configurations on the contact rate.

In this approach, one considers a fluctuating variable (here, the separation $r$ between the two residues at the loop termini) and estimates the average time required for that variable to reach a particular value. The mean first passage time is the average, over the probability distribution of the variable, of that first passage time. SSS theory is a mean-first-passage-time calculation for the rate of a diffusion-controlled reaction that occurs in a symmetric force field in one dimension ($r$) and is likely to be described by first-order (i.e., exponential) reaction kinetics. For intrachain contact formation in a polymer chain, the dynamics of $r$ are controlled by an effective force field between the two reacting residues:

$$U(r) = -k_B T \ln\left(P(r)\right). \tag{1}$$

Here, $P(r)dr$ is the equilibrium probability distribution for the distance $r$ between the residues. SSS theory uses the Smoluchowski equation to describe the diffusion of the system in this force field, and calculates the mean passage time $\tau$ of the ensemble to the formation of a contact at $r = a$:

$$\tau = \int_a^{L_L} \frac{dx}{D(x)P(x)} \left[\int_x^{L_L} P(y)dy\right]^2. \tag{2}$$

Here, $L_L$ is the contour length of the loop, $D(x)$ is the effective diffusion constant, and $a$ is the contact radius (i.e., the residue-residue separation that defines the formation of a contact). SSS theory therefore relates a kinetic quantity, $\tau$, to the equilibrium probability distribution $P(r)dr$. For the purposes of this article, we assume that the effective diffusion constant $D$ is independent of $r$ and also independent of the chain contour length. $D$ presumably depends on the viscosity of the solvent, the persistence length of the polypeptide, and the internal friction of the chain (38). From the probability distribution $P(r)$ calculated for a particular length of chain and a particular pair of residues, we can then numerically integrate Eq. 2 to obtain the product $D\tau$, describing the rate of closure of the loop.

It should be noted that SSS theory does not necessarily provide an accurate absolute rate, however. Portman (39) has shown that, given the microscopic diffusion constant $D$, the SSS (one-dimensional) and Wilemski-Fixman (closure-approximation) (40) theories lead to lower and upper bounds, respectively, on the first contact time in loop formation. That is, the SSS approach—replacing the full dynamics with diffusion on a one-dimensional potential of mean force—is a valid approach, but inserting the diffusion constant of a free monomer for the value of $D$ will lead to an underestimate of the contact time. We do not try to insert a value for $D$, as we are interested in relative changes in the contact times, rather than absolute rates. In the figures that follow we show the contact-formation time as the product $D\tau$.

### Generation of $P(r)$ distributions

We model the polypeptide as a freely jointed chain (41) of $N$ hard-sphere monomers, separated by bonds of unit length. (Fig. 1 *B*) The diameter of the hard spheres is given by $\lambda$, the excluded-volume parameter. Two monomers are said to make contact when their centers are separated by a distance $a$. For a given pair of sites on the chain, we define $R^2$ as the mean-square value of the distance $r$ between the sites. For the ideal chain (i.e., $\lambda = 0$), one has $R^2 = N$ in the limit of large $N$. In that case, the probability distribution $P(r)$ is Gaussian, regardless of whether the points define an internal or external loop.

We obtain $P(r)$ for a given freely jointed chain in the presence of excluded volume ($\lambda > 0$) by creating an ensemble of chains. Each chain in the ensemble is constructed as follows: we place one monomer at the origin, and then place a second monomer in a random location on the surface of the

sphere of unit radius that centers on the first monomer. Subsequent monomers are each placed at unit distance from the preceding monomer, and in a random location on the unit sphere surrounding that monomer. This construction ignores the excluded-volume constraint: monomers or links may cross or overlap in space. It produces an ideal chain of $N$ monomers having $N - 1$ links of unit length.

We then apply the excluded-volume condition to the chain. We calculate the distances between all nonadjacent monomers; if any interresidue distances are $<\lambda$, the chain is discarded. Otherwise, it is retained in the ensemble. We continue in this fashion until the ensemble contains at least 8 million random chains of length $N$, unbiased except for the excluded-volume constraint. Each chain in the ensemble resembles a necklace of hard spheres, each of radius $\lambda/2$ and separated from its two neighbors by a bond of unit length, and with no spheres overlapping.

We can select a pair of sites (e.g., monomers 1 and $N$) on the chain and construct a histogram of the distance $r$ between those monomers for all chains in the ensemble. Values of $r$ range from the excluded-volume parameter $\lambda$ to the loop contour length. The number of monomers in the loop is denoted by $L_L$, making the loop contour length $L_L - 1$. The histogram, when normalized to unit area, gives the equilibrium probability distribution, $P(r)$, for the distance between the selected residues (Fig. 2). By numerically integrating this $P(r)$ in the SSS expression (Eq. 2), we find the average contact time $D\tau$. We use the excluded-volume parameter $\lambda$ as the contact radius $a$: loop closure requires the two hard spheres to make their closest possible approach.

To estimate the error in our measurement of $D\tau$, we divide the 8 million chains into 10 subsets of 800,000 chains. We calculate the average contact time for each subset, and then take the error as the standard deviation of these averages for the 10 subsets.

## RESULTS AND DISCUSSION

### Probability distributions

Fig. 2 *A* shows the $P(r)$ for external loops of 10 monomer units, at the three different values of the excluded volume parameter, $\lambda = 0$, 0.5, 1.0. Not surprisingly, greater $\lambda$ shifts $P(r)$ toward larger $r$: all parts of the polymer chain move farther apart, including those that will form the loop. Fig. 2 *C* shows the effective potentials $U(r)$ (Eq. 1) that, in SSS theory, determine the diffusional dynamics of loop formation for these chains. The shape of $U(r)$ (and $P(r)$) at small $r$ has a strong influence on $\tau$, because the contact time is the time required for the system to diffuse up the steep curve at small $r$ to reach $r = \lambda$.

Fig. 2 *B* shows the effect of converting the 10-monomer external loop to an internal loop, with excluded volume fixed at $\lambda = 1$. The figure shows the probability distribution $P(r)$ for the distance between monomers 1 and 10 in a chain, as both chain termini on the ends of the loop are extended by additional segments (tails) of length 0, 1, and 10 monomers. The addition of the tail segments suppresses $P(r)$ very slightly at small $r$, leading to slightly higher probability at large $r$. This small shift indicates that an SSS calculation (Eq. 2) will find a slower rate of contact formation for larger tails.

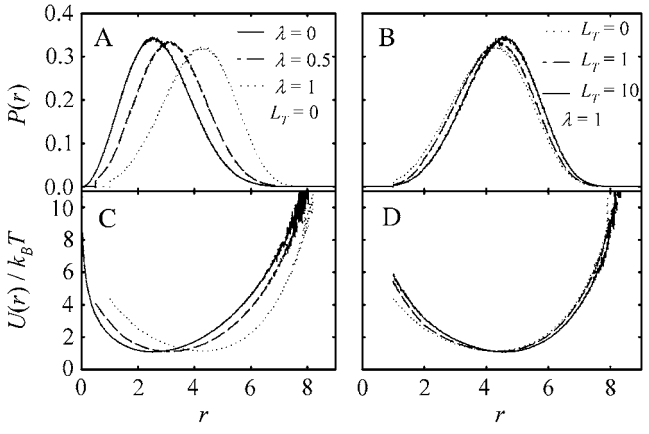


FIGURE 2 Probability distributions $P(r)$ and corresponding potentials $U(r)$ for internal and external loops. Each distribution comprises $8 \times 10^6$ simulated chains. (*A* and *C*) $P(r)$ (*A*) and $U(r)$ (*C*) for an external ($L_T = 0$) loop of 10 monomers, for different values of the excluded-volume parameter $\lambda$. Distributions shift outward in $r$ as $\lambda$ increases. (*B* and *D*) $P(r)$ (*B*) and $U(r)$ (*D*) for loops of 10 monomers with $\lambda = 1$. As tail length increases, distributions shift outward in $r$. (Note that the $\lambda = 1, L_T = 0$ case appears in all panels for comparison). Bin size is 0.01 distance units.

### Contact time versus tail length: internal loops

Fig. 3 shows the contact times, calculated from the $P(r)$ and Eq. 2, as a function of tail length $L_T$, for several different values of the loop length $L_L$. Contact times, $\tau(L_T)$, are normalized to the corresponding external loop time, $\tau(0)$ (for the same loop contour length), for comparison. The figure also shows the behavior of the ideal chain—a horizontal line at constant $\tau(L_T)/\tau(0)$—for which the distribution $P(r)$ is unaffected by the addition of tails. We see that adding even a few residues to convert an external ($L_T = 0$) loop to an internal loop can significantly affect the rate of contact formation. For the $\lambda = 1$, $L_L = 10$ chain, the contact time nearly doubles with the addition of just one monomer at each terminus.

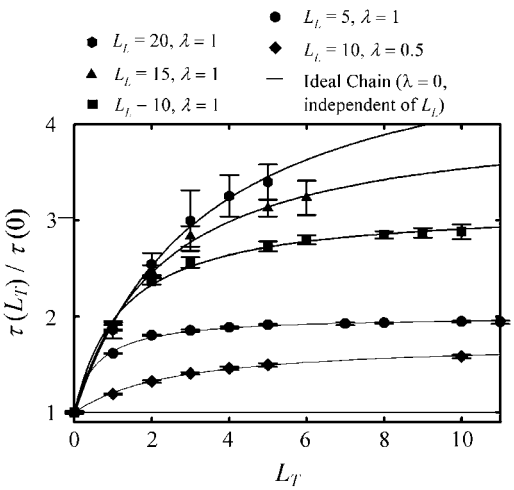


FIGURE 3 Effect of tail length on contact time $\tau$ for internal loops. Values of $\tau$ for loops with tail length $L_T$ are normalized to the value of $\tau$ for the same loop but with $L_T = 0$. The horizontal line describes the behavior of an ideal chain, for which $\lambda = 0$. Solid curves are fits to the empirical function of Eq. 3.

Not surprisingly, the influence of the tails increases as the excluded-volume parameter increases. The contact time for the $\lambda = 0.5$, 10-monomer chain increases by ~1.4 times as $L_T$ increases from 0 to 10, whereas the contact time for the $\lambda = 1.0$, 10-monomer chain increases almost threefold over the same range. It is perhaps surprising that the effect of the tails is more pronounced for longer loops than for shorter loops; the trend predicts greater suppression of internal loop formation during the folding of longer polypeptides. In the context of SSS theory, this is a consequence of the fact that the effective potential $U(r)$ (Eq. 1) at small $r$ is increased by two separate effects—lengthening the contour of the loop and adding tail residues—and $\tau$ is essentially exponential (Eq. 2) in $U(r)$. An increase of both loop contour and tail length together has a multiplicative effect in increasing $\tau$.

The variation of $\tau$ with tail length $L_T$ in Fig. 3 resembles the variation of the loop closure probability with tail length, as calculated by Chan and Dill (12) for lattice polymers: like the probability, the contact time changes rapidly with the addition of the first few tail residues and then saturates, with little additional change as the tails grow still longer. The saturation of $\tau$ is not surprising, since tails will random-walk outward from the vicinity of the loop termini; subsequent residues added to the tail have an ever-diminishing probability of residing near the contact points. It is interesting, however, that the saturation occurs more slowly for loops of greater contour length: the longer the loop, the greater the role of every residue along the tail in slowing loop formation. Fig. 3 also shows another interesting finding: the saturation effect in $\tau(L_T)$ for internal loops is quite accurately described by a simple empirical function,

$$\tau(L_T) = \left( \tau_\infty - \frac{\tau_\infty - \tau_0}{1 + \dfrac{L_T}{L_0}} \right) \tag{3}$$

(Fig. 3, *solid curves*). Here $\tau_\infty$ is the value of $\tau$ at $L_T \rightarrow \infty$ and $\tau_0$ is the value at $L_T = 0$. Table 1 gives the fit parameters, and Fig. 4 shows the fit parameters as a function of loop length. $\tau_\infty$ and $L_0$ both vary with $L_L$. The ratio of parameters $\tau_\infty/\tau_0$ scales linearly with $L_L$, showing again that the tails have a greater relative effect on longer loops. The parameter $L_0$ indicates roughly how many tail residues are required to raise $\tau$ to its limiting value. Evidently, this number equals 10–15% of the loop contour $L_L$ (Fig. 4 *B*); again, adding the initial few residues to the tail has the greatest effect on the loop rate, but subsequent residues at the tail play a role in longer loops.

**TABLE 1 Values of parameters in fit of $D\tau$ versus $L_T$ to Eq. 3**

| $L_L$ | $\lambda$ | $D\tau_0$ | $D\tau_\infty$ | $L_0$ | $\tau_\infty/\tau_0$ |
|---|---|---|---|---|---|
| 10 | 0.5 | 10.4 ± 0.1 | 18.2 ± 0.2 | 2.6 ± 0.2 | 1.75 ± 0.03 |
| 5 | 1 | 5.78 ± 0.09 | 11.61 ± 0.06 | 0.58 ± 0.04 | 2.01 ± 0.02 |
| 10 | 1 | 41 ± 2 | 132 ± 2 | 1.2 ± 0.1 | 3.18 ± 0.04 |
| 15 | 1 | 103 ± 5 | 430 ± 10 | 2.3 ± 0.3 | 4.16 ± 0.06 |
| 20 | 1 | 190 ± 10 | 990 ± 70 | 3.3 ± 0.6 | 5.2 ± 0.1 |

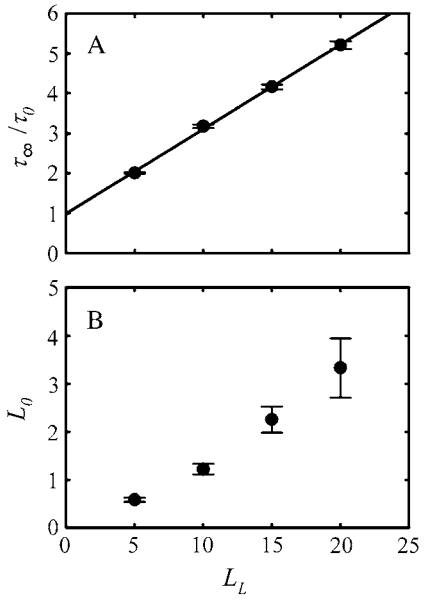


FIGURE 4 Parameters from fits of the data of Fig. 3 to the empirical function of Eq. 3. (*A*) $\tau_\infty/\tau_0$ increases linearly with loop length, indicating that the effect of the tail grows directly as the loop contour length $L_L$ increases. (*B*) $L_0$ also increases with loop length with a less than linear dependence.

Fig. 5 shows the absolute (not normalized) contact times versus loop contour length $L_L$ for different values of the tail length, $L_T$. Not surprisingly, $D\tau$ increases sharply with loop length, although it does not follow the power law (i.e., $L_T^{3/2}$) scaling often attributed to SSS theory (see below). The figure shows that loop length and tail length have comparable effects on contact time: the effect of adding tails is by no means negligible compared to the effect of increasing the loop contour. The same can be said of excluded-volume effects; variations of $\lambda$ have large effects on $\tau$. The inset to the figure (showing a logarithmic scale) confirms the surprising finding that the relative enhancement of $\tau$ by the tails increases as the loop contour becomes longer.

## Contact time versus tail length: end-to-interior loops

Fig. 6 shows the contact time versus tail length for four end-to-interior loops, all with $\lambda = 1$. The figure shows, for example, that for a loop of 10 residues, a single tail of five residues (i.e., an end-to-interior loop) suppresses the contact rate by a factor of ~1.46. By comparison (cf. Fig. 3), the addition of two five-residue tails (i.e., forming an internal loop) suppresses the contact rate by a factor of 2.71, i.e., by a factor that is $>(1.46)^2$ or 2.13. The presence of two tails has a contact-suppressing effect greater than the product of two single-tail effects. The tails evidently interact with each other, further expanding the chain, in addition to simply blocking contacts between the loop termini.

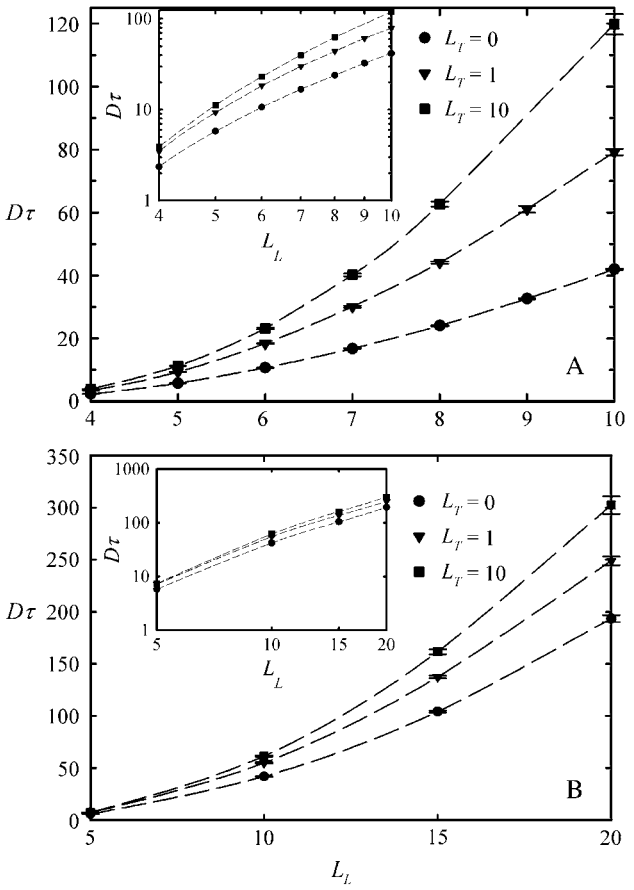


FIGURE 5 Contact time versus loop length for (*A*) internal loops and (*B*) end-to-interior loops. For all data, $\lambda = 1$. (*Inset*) Same data on a log-log scale, indicating that the contact time does not obey a power law in the loop length. Dashed lines are drawn to guide the eye.

We did not fit the data in Fig. 6 to the empirical function of Eq. 3 because the data show an interesting behavior not seen in internal loops: the contact time actually decreases for sufficiently long tails, longer than approximately seven residues. The decrease is quite weak, and one may question whether the effect is actually significant above the noise level in the figure. To verify that the contact time really is decreasing with increasing tail length, we show (Fig. 7) selected probability distributions that were plotted for end-to-interior loops, along with pairwise differences between these distributions. Fig. 7 shows the distributions $P(r)$ for three different tail lengths, attached to a 10-residue loop. Adding the first five monomers to the end-to-end loop shifts $P(r)$ to the right, as in Fig. 2. However, further increasing the tail length to 20 causes a slight broadening of the probability distribution. This can be seen more clearly in the difference between the $L_T = 5$ and $L_T = 20$ distributions; that difference, shown in the lower panel of Fig. 7, is positive at both smaller and larger $r$, but negative in the middle range ($r =$ 4–5). Therefore, relative to a loop with a short, five-monomer tail (i.e., $L_T = 5$), the loop with a long 20-monomer tail ($L_T = 20$) is described by a broader distribution $P(r)$, with a greater probability of both compact (i.e., closed loop) and more extended configurations, and a slightly lower probability of intermediate configurations. One might interpret this result by assuming that long tails tend to take an average dimension that is comparable to the average dimensions of the loop region itself. Then the tail and the loop region overlap physically; the excluded volume constraint would then tend to favor configurations of the loop residues that are either slightly more compact or slightly more extended than usual. The resulting (weak) enhancement of the loop-formation rate is an interesting and unexpected consequence of adding a tail to the loop. Although this same

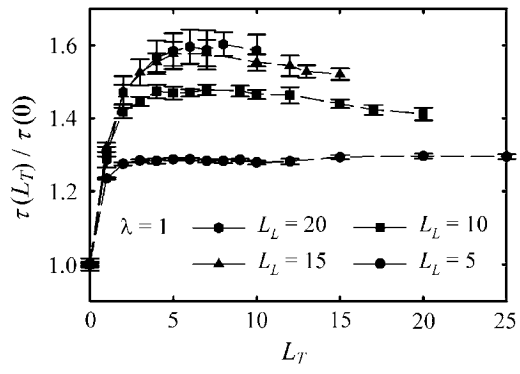


FIGURE 6 Contact time versus tail length for end-to-interior loops. Contact time initially increases with the addition of a short tail, but then decreases slightly if the tail grows sufficiently long. Dashed lines are drawn to guide the eye.

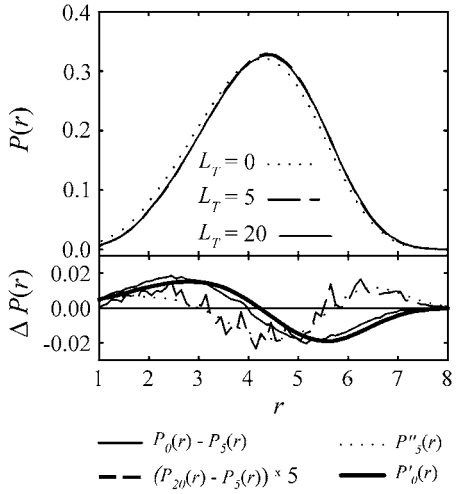


FIGURE 7 Upper panel shows probability distributions for end-to-interior loops of $L_L = 10$, for varying tail lengths $L_T = 0$, 5, and 20. The distribution $P(r)$ shifts to the right as $L_T$ increases from 0 to 5, but broadens slightly as $L_T$ further increases from 5 to 20. Lower panel shows the differences between these distributions: (*thin solid curve*) difference between the $L_T = 0$ and $L_T =$ 5 distributions; (*heavy solid curve*) first derivative $P'(r)$ for $L_T = 0$, scaled to overlap the thin solid curve; (*dashed curve*) difference between the $L_T = 5$ and $L_T = 20$ distributions; (*dotted curve*) second derivative $P''(r)$ for $L_T = 5$, scaled to overlap the dashed curve.

enhancement was not seen in the internal loops (Fig. 3), we expect that it would occur for those loops as well, once the tails become sufficiently long.

Fig. 5 *B* shows how contact time varies as a function of loop length for end-to-interior loops. The plot resembles the internal-loop case (Fig. 5 *A*), although $\tau$ shows a less drastic dependence on tail length.

The findings above make some interesting predictions for experimental studies of contact formation in polypeptides and nucleic acids. Contact formation in oligopeptides has been of particular interest recently, although only very minimal experimental data comparing internal/external loop formation in peptides has been published (25). Aside from the lack of data, the problem that arises in comparing our simulation to experiment is of course that we have treated the polypeptide as a freely jointed chain, whereas polypeptides are not freely jointed. We can, however, estimate the length of unfolded polypeptide that would correspond to one segment of the equivalent freely jointed chain. The persistence length $L_p$ of an unfolded polypeptide can be estimated from atomic force microscopy (42,43) or by other methods (25): values of $L_p \approx 0.4$–$0.44$ nm appear typical. From the 0.38-nm bond distance separating consecutive $C_\alpha$ atoms in an unfolded polypeptide, one then estimates that the equivalent freely jointed segment (i.e., the statistical segment) of an unfolded polypeptide is $2L_p \approx 0.8$–$0.9$ nm, or $\sim$2.1–2.4 residues. This argument (and the parameters in Table 1) suggests that addition of two residues on a short polypeptide chain will be almost sufficient to give the full (i.e., saturated) slowing effect in the contact-formation rate. We are aware of only one experiment that bears on this prediction: Buscaglia et al. have recently used laser spectroscopy to measure the loop formation rate in a short synthetic peptide of 11 residues (25). Based on the above estimate for the effective segment length of polypeptides, an 11-residue polypeptide chain corresponds to roughly $N = 5$ freely jointed monomers. Those authors found that adding a tail of one peptide bond reduced the contact formation rate by $\sim$30%. A tail of nine bonds appeared to give a net effect of $\sim$40%; that is, there was little additional effect on the loop formation rate. This appears generally consistent with the $\sim$30% effect seen on the contact rate of the $N = 5$ chain in Fig. 6. However, more detailed experimental studies would be needed to confirm agreement with the simulation studies described here.

### Error in SSS approximation

Szabo et al. showed that inserting the ideal chain $P(r)$ into SSS theory, Eq. 2, leads to an expression that can be approximated by a power series in $\alpha = (3/2)^{1/2}(a/R)$, where $a$ is the value of $r$ that defines the formation of a contact (23):

$$\frac{3D}{R^2}\tau = \frac{\sqrt{\pi}}{2\alpha} + \ln 2 - 1 - \frac{\sqrt{\pi}}{2}\alpha + \frac{4}{3}\alpha^2 + \ldots . \quad (4)$$

If the contact radius $a$ is much smaller than the root mean-squared distance between the loop termini, $R$, this series can be truncated at the first term, giving

$$D\tau = \frac{R^3}{3a}\sqrt{\frac{\pi}{6}}. \quad (5)$$

This simple expression predicts a scaling $\tau \propto R^3 \approx N^{3/2}$, a result often cited as the SSS prediction for the contact formation time (5,6,16–18,33,35,44). The error introduced by retaining only this first term in Eq. 4 can be substantial, however. For the simple case of a freely jointed chain with no excluded volume, Fig. 8 shows the error introduced by truncating the series (Eq. 4) at successive terms, relative to the numerically integrated SSS result, Eq. 2, for a freely jointed ideal (i.e., $\lambda = 0$) chain. The errors grow large for chains of $<\sim$20 monomers. Including only the first term results in errors of order 20% for a 25-monomer loop, and $\sim$125% for a five-monomer loop. Including the second term reduces the error only to $\sim$40% for the eight-monomer loop. Hence the first term approximation to SSS theory is relatively inaccurate for loops of $<\sim$20 monomers. If such a freely jointed loop is roughly analogous to an $\sim$40-residue peptide loop, one should not expect the first-term approximation to be very reliable when applied to most loop-formation experiments on short, synthetic peptides.

## CONCLUSIONS

The formation of intrachain contacts in an unfolded polypeptide is an elementary step in protein folding. Recent interest in the dynamics of contact formation has been largely motivated by the goal of determining a physics-based limit to

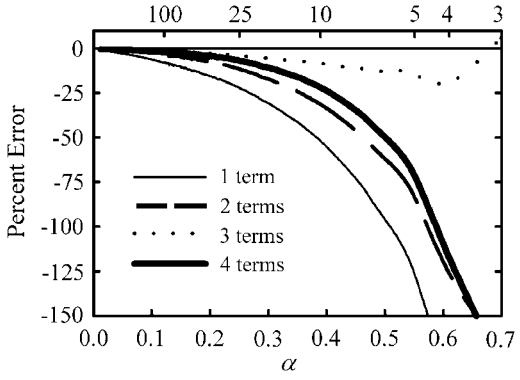


FIGURE 8 Accuracy of successive approximations to the SSS theory. The horizontal line represents the exact result of Eq. 2 from numerical integration of $P(r)$ for an ideal ($\lambda = 0$) freely jointed chain. The other curves represent the error introduced by truncating the series expansion of SSS theory (Eq. 4) at successive terms. The number of freely jointed residues in the loop appears along the top of the figure. Using only the first term in Eq. 4 (i.e., the first approximation to SSS theory) introduces an error of $\sim$100% in the contact time $\tau$ for a loop of $\alpha = (3/2)^{1/2}(a/R) = 0.5$. In fact, for a loop of approximately six residues in a freely jointed chain, one must retain at least four consecutive terms in Eq. 4 to reduce the error in $\tau$ to $<\sim$50%.

the possible speed of folding of small, designed polypeptides. We have used a simple freely-jointed chain model, with hard-sphere excluded-volume interactions, to investigate how the position of the nascent loop within the chain influences the rate of loop formation. Applying the SSS theory to the resulting distributions shows that one can expect significant slowing in the dynamics as external loops are converted (by addition of extra residues at the termini) into internal loops. For a relatively short loop of 10 links, the presence of just a few additional monomers outside the loop slows the rate of contact formation threefold. Tails appear to have an even larger effect on closure rates of longer loops, such as those that appear in folded protein structures. In this sense, experimental studies on end-to-end contact formation in short, synthetic polypeptides may substantially overestimate the limiting kinetic rates relevant to early events in protein folding. We find that a simple empirical law seems to describe the tail-length dependence of the loop-formation rate. We also observe interesting crowding effects, including a synergistic effect of two tails (relative to single tails) and a weak acceleration of the contact formation rate in the case of sufficiently long tails. These predictions are amenable to experimental test: such experiments would be of particular interest as a test of the SSS theory used here, in which the chain statistics alone essentially determine the rate of contact formation. SSS theory does not take account of any possible purely dynamical or drag effects resulting from tails.

Of course, although folding is certainly a heavily damped diffusional process, one may question whether contact formation actually does represent the most stringent physical limit on folding rates. One must expect other factors to play a role as well. The internal friction of the chains, a phenomenon that has received considerable attention in the homopolymer dynamics literature (45), appears to influence polypeptide dynamics on the loop formation timescale (microseconds or nanoseconds). It may therefore have a limiting effect on folding speed (2). Also, as mentioned above, the simple SSS theory used here does not take account of dynamical effects of introducing tails onto the termini of a loop (i.e., effects of the additional viscous drag hindering loop closure). As they are ultimately critical in determining the physical forces that set the limits on protein folding dynamics, these effects will certainly be subject to more detailed study through theory, simulation, and experiment.

S.J.H. acknowledges helpful discussions with Dr. Tom Schwartz.

The authors gratefully acknowledge funding support from the National Science Foundation, MCB 0347124 (to S.J.H.). Computer resources were provided by the Large Allocations Resource Committee through grant TG-MCA05S010 (to A.E.R. and S.J.H.).

## REFERENCES

1. Eaton, W. A., V. Munoz, P. A. Thompson, C. K. Chan, and J. Hofrichter. 1997. Submillisecond kinetics of protein folding. *Curr. Opin. Struct. Biol.* 7:10–14.
2. Kubelka, J., J. Hofrichter, and W. A. Eaton. 2004. The protein folding ''speed limit''. *Curr. Opin. Struct. Biol.* 14:76–88.
3. Eaton, W. A., V. Munoz, S. J. Hagen, G. S. Jas, L. J. Lapidus, E. R. Henry, and J. Hofrichter. 2000. Fast kinetics and mechanisms in protein folding. *Annu. Rev. Biophys. Biomol. Struct.* 29:327–359.
4. Hudgins, R. R., F. Huang, G. Gramlich, and W. M. Nau. 2002. A fluorescence-based method for direct measurement of submicrosecond intramolecular contact formation in biopolymers: an exploratory study with polypeptides. *J. Am. Chem. Soc.* 124:556–564.
5. Bieri, O., J. Wirz, B. Hellrung, M. Schutkowski, M. Drewello, and T. Kiefhaber. 1999. The speed limit for protein folding measured by triplet-triplet energy transfer. *Proc. Natl. Acad. Sci. USA.* 96:9597–9601.
6. Lapidus, L. J., W. A. Eaton, and J. Hofrichter. 2000. Measuring the rate of intramolecular contact formation in polypeptides. *Proc. Natl. Acad. Sci. USA.* 97:7220–7225.
7. Hagen, S. J., J. Hofrichter, A. Szabo, and W. A. Eaton. 1996. Diffusion-limited contact formation in unfolded cytochrome *c*: estimating the maximum rate of protein folding. *Proc. Natl. Acad. Sci. USA.* 93:11615–11617.
8. Friedman, B., and B. Oshaughnessy. 1993. Theory of intramolecular reactions in polymeric liquids. *Macromolecules.* 26:4888–4898.
9. Camacho, C. J., and D. Thirumalai. 1995. Theoretical predictions of folding pathways by using the proximity rule, with applications to bovine pancreatic trypsin-inhibitor. *Proc. Natl. Acad. Sci. USA.* 92: 1277–1281.
10. Debnath, P., and B. J. Cherayil. 2004. Dynamics of chain closure: approximate treatment of nonlocal interactions. *J. Chem. Phys.* 120: 2482–2489.
11. Chan, H. S., and K. A. Dill. 1989. Intrachain loops in polymers: effects of excluded volume. *J. Chem. Phys.* 90:492–509.
12. Chan, H. S., and K. A. Dill. 1990. The effects of internal constraints on the configurations of chain molecules. *J. Chem. Phys.* 92:3118–3135.
13. Chen, J. Z. Y., H. K. Tsao, and Y. J. Sheng. 2004. First-passage time of cyclization dynamics of a wormlike polymer. *Europhys. Lett.* 65: 407–413.
14. Sheng, Y. J., J. Z. Y. Chen, and H. K. Tsao. 2002. Open-to-closed transition of a hard-sphere chain with attractive ends. *Macromolecules.* 35:9624–9627.
15. Sheng, Y. J., P. H. Hsu, J. Z. Y. Chen, and H. K. Tsao. 2004. Loop formation of a flexible polymer with two random reactive sites. *Macromolecules.* 37:9257–9263.
16. Zhou, H. X. 2003. Theory for the rate of contact formation in a polymer chain with local conformational transitions. *J. Chem. Phys.* 118:2010–2015.
17. Zhou, H. X. 2004. Polymer models of protein stability, folding, and interactions. *Biochemistry.* 43:2141–2154.
18. Pastor, R. W., R. Zwanzig, and A. Szabo. 1996. Diffusion limited first contact of the ends of a polymer: comparison of theory with simulation. *J. Chem. Phys.* 105:3878–3882.
19. Sokolov, I. M. 2003. Cyclization of a polymer: first-passage problem for a non-Markovian process. *Phys. Rev. Lett.* 90:080601.
20. Wang, Z. S., and D. E. Makarov. 2002. Rate of intramolecular contact formation in peptides: the loop length dependence. *J. Chem. Phys.* 117:4591–4593.
21. Wittkop, M., S. Kreitmeier, and D. Goritz. 1996. The distribution function of internal distances of a single polymer chain with excluded volume in two and three dimensions: a Monte Carlo study. *J. Chem. Phys.* 104:351–358.
22. Yeh, I. C., and G. Hummer. 2002. Peptide loop-closure kinetics from microsecond molecular dynamics simulations in explicit solvent. *J. Am. Chem. Soc.* 124:6563–6568.
23. Szabo, A., K. Schulten, and Z. Schulten. 1980. 1st passage time approach to diffusion controlled reactions. *J. Chem. Phys.* 72:4350–4357.

24. Redner, S. 1980. Distribution functions in the interior of polymer chains. *J. Phys. A.* 13:3525–3541.

25. Buscaglia, M., L. J. Lapidus, W. A. Eaton, and J. Hofrichter. 2006. Effects of denaturants on the dynamics of loop formation in polypeptides. *Biophys. J.* 91:276–288.

26. Descloizeaux, J. 1980. Short-range correlation between elements of a long polymer in a good solvent. *J. Phys. (France).* 41:223–238.

27. Matthews, K. S. 1992. DNA looping. *Microbiol. Rev.* 56:123–136.

28. Haas, E., M. Wilchek, E. Katchalskikatzir, and I. Z. Steinberg. 1975. Distribution of end-to-end distances of oligopeptides in solution as estimated by energy-transfer. *Proc. Natl. Acad. Sci. USA.* 72:1807–1811.

29. Grinvald, A., E. Haas, and I. Z. Steinberg. 1972. Evaluation of distribution of distances between energy donors and acceptors by fluorescence decay. *Proc. Natl. Acad. Sci. USA.* 69:2273–2277.

30. Haas, E., E. Katchalskikatzir, and I. Z. Steinberg. 1978. Brownian-motion of ends of oligopeptide chains in solution as estimated by energy-transfer between chain ends. *Biopolymers.* 17:11–31.

31. Krieger, F., B. Fierz, O. Bieri, M. Drewello, and T. Kiefhaber. 2003. Dynamics of unfolded polypeptide chains as model for the earliest steps in protein folding. *J. Mol. Biol.* 332:265–274.

32. Lee, J. C., H. B. Gray, and J. R. Winkler. 2005. Tertiary contact formation of $\alpha$-synuclein probed by electron transfer. *J. Am. Chem. Soc.* 127:16388–16389.

33. Lapidus, L. J., P. J. Steinbach, W. A. Eaton, A. Szabo, and J. Hofrichter. 2002. Effects of chain stiffness on the dynamics of loop formation in polypeptides. Appendix: Testing a 1-dimensional diffusion model for peptide dynamics. *J. Phys. Chem. B.* 106: 11628–11640.

34. Nau, W. M., and X. Y. Zhang. 1999. An exceedingly long-lived fluorescent state as a distinct structural and dynamic probe for supramolecular association: an exploratory study of host-guest complexation by cyclodextrins. *J. Am. Chem. Soc.* 121:8022–8032.

35. Hagen, S. J., C. W. Carswell, and E. M. Sjolander. 2001. Rate of intrachain contact formation in an unfolded protein: temperature and denaturant effects. *J. Mol. Biol.* 305:1161–1171.

36. Neuweiler, H., A. Schulz, M. Bohmer, J. Enderlein, and M. Sauer. 2003. Measurement of submicrosecond intramolecular contact formation in peptides at the single-molecule level. *J. Am. Chem. Soc.* 125: 5324–5330.

37. Hyeon, C., and D. Thirumalai. 2006. Kinetics of interior loop formation in semiflexible chains. *J. Chem. Phys.* 124:104905.

38. Hagen, S. J., L. L. Qiu, and S. A. Pabit. 2005. Diffusional limits to the speed of protein folding: fact or friction? *J. Phys. Condens. Matter.* 17:S1503–S1514.

39. Portman, J. J. 2003. Non-Gaussian dynamics from a simulation of a short peptide: loop closure rates and effective diffusion coefficients. *J. Chem. Phys.* 118:2381–2391.

40. Wilemski, G., and M. Fixman. 1974. Diffusion-controlled intrachain reactions of polymers. 2. Results for a pair of terminal reactive groups. *J. Chem. Phys.* 60:878–890.

41. Grosberg, A. Y., and A. R. Khoklov. 1994. Statistical Physics of Macromolecules. AIP Press, Woodbury, NY.

42. Oberhauser, A. F., P. E. Marszalek, H. P. Erickson, and J. M. Fernandez. 1998. The molecular elasticity of the extracellular matrix protein tenascin. *Nature.* 393:181–185.

43. Rief, M., M. Gautel, F. Oesterhelt, J. M. Fernandez, and H. E. Gaub. 1997. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science.* 276:1109–1112.

44. Jones, C. M., E. R. Henry, Y. Hu, C. K. Chan, S. D. Luck, A. Bhuyan, H. Roder, J. Hofrichter, and W. A. Eaton. 1993. Fast events in protein-folding initiated by nanosecond laser photolysis. *Proc. Natl. Acad. Sci. USA.* 90:11860–11864.

45. Manke, C. W., and M. C. Williams. 1985. Internal viscosity of polymers and the role of solvent resistance. *Macromolecules.* 18: 2045–2051.