On the interpretation of force extension curves of single protein molecules

Dmitrii E. Makarov and Zhisong Wang
Department of Chemistry and Biochemistry and Institute for Theoretical Chemistry,
University of Texas at Austin, Austin, Texas 78712

James B. Thompson and Helen G. Hansma
Department of Physics, University of California, Santa Barbara, California 93106

Received 3 October 2001; accepted 8 February 2002

The atomic force microscope can be used to forcibly unfold and extend single polypeptide chains. The resulting force versus distance curves have been widely interpreted to arise from the loss of entropy that the unfolded polypeptide chain experiences as it is extended. Here, we have used Monte Carlo simulations of unfolded polypeptide chains to examine the average distance between the ends of a polypeptide chain as a function of the force that pulls these ends apart. We examine two types of experiments: (a) A rigid force-sensor (bead-type) experiment: The chain is subjected to a constant stretching force \( f \) and the resulting chain extension is measured. (b) A flexible force-sensor (cantilever-type) experiment: The force is measured by the deflection of a cantilever that is attached to one end of the chain. The total length of the chain plus the displacement of the cantilever is fixed. In case (b), in the limit of a large cantilever force constant, the entropic force \( f \) is related to the free energy of the chain \( F(r) \) constrained to have the end-to-end distance \( r \) by the usual thermodynamic relationship: \( f = dF/dr \). However in case (a) this relationship is invalid. The reason of its failure is that large fluctuations in the end-to-end distance \( r \) cannot be neglected at the single molecule level and so macroscopic thermodynamics relationships cannot be used. Thus the two types of experiments measure different force extension curves \( f(r) \). We compute the force extension curves for a model of a polypeptide chain in each case and find that they are significantly different. We further discuss implications of our findings with regard to the results of cantilever-type unfolding experiments. © 2002 American Institute of Physics. [DOI: 10.1063/1.1466835]

I. INTRODUCTION

Mechanical experiments performed on single biomolecules provide invaluable insights into their structure and function. Among many applications, they allow one to extract information about free energy landscapes of proteins and to directly study the mechanical characteristics of molecular motors and the proteins that perform mechanical function in living organisms. These studies have been helpful in understanding the origins of unusual strength exhibited by muscle proteins.

In a typical experiment (Fig. 1), one pulls a polymer molecule at its ends and measures the relationship \( f(r) \) between the force \( f \) exerted on the polymer and the end-to-end distance \( r \). The force-extension curves \( f(r) \) of proteins are typically strongly nonlinear. In this paper we will concentrate on the force-extension curves of unfolded proteins. In this case, it is generally believed that the origin of the force \( f \) is mostly entropic. If \( F(r) \) is the free energy of the chain constrained to have the end-to-end distance \( r \) then the force is usually calculated from the thermodynamic relationship

\[
  f(r) = \frac{dF}{dr}
\]

(in the above choice of sign, \( f \) is the force exerted on the protein).

The free energy \( F \) is related to the probability distribution \( p_0(r) \) for the end-to-end distance \( r \) in the absence of the force:

\[
  F(r) = -k_B T \log p_0(r).
\]

According to Eqs. (1) and (2), chains with Gaussian distributions \( p_0(r) \) (Ref. 20) should obey Hook’s law.

One often fits experimental dependences \( f(r) \) by the wormlike chain (WLC) model, for which an approximate analytic formula has been given:

\[
  f(r) = (k_B T/P)[0.25(1 - r/L)^2 - 0.25 + r/L].
\]

This formula has two parameters, the polymer contour length \( L \) and its persistence length \( P \), and has features that describe the polymer’s elasticity in a qualitatively correct way. In particular, it predicts that the polymer should become stiffer (i.e., \( df/dr \) should increase) as the extension approaches the contour length \( L \). Further, the observed change in contour length (as determined by fitting the WLC model), after the rupture of a single protein domain, successfully predicts the contour length of the domain that ruptured. However, quantitatively, the above formula does not always fit the experimental data well. The value of the persistence length that is required to fit the titin force-extension curves is very different from the persistence lengths of peptides found in the experiments with random coils in the absence of mechanical tension. Furthermore, different persistence lengths had to be used for different values of \( r \) in order to fit the data. For example, WLC fits for the protein spectrin needed a persistence length of 0.8 nm for forces below 50 pN and a persis-

Author to whom correspondence should be addressed. Electronic mail: makarov@mail.cm.utexas.edu
tence length of 0.4 nm for forces above 50 pN. In addition, unpublished results with titin, collagen and spider silk proteins all show that power laws and exponentials fit the force-distance curves better than the WLC. Extensions and refinements of the original wormlike chain model have been suggested to improve the quality of the fits it provides; however the better accuracy of the fits may be due simply to the larger numbers of the adjustable parameters in these models.

The purpose of this paper is twofold. First, we report on a simple atomistic Monte Carlo calculation of the “entropic force” \( f(r) \) that takes excluded volume effects into account. We find the calculated force extension curves to be strongly nonlinear, in accord with what is seen in the AFM experiments performed on single proteins. Molecular dynamics simulations have been previously used to compute force-extension curves of protein domains. However calculating the equilibrium entropic force via a molecular dynamics calculation would be prohibitive in the case of an unfolded peptide (formed, e.g., by a ruptured protein domain) because of the large number of conformations that would have to be sampled and the long time it would take.

Second, we point out here that the thermodynamic relationship of Eq. (1) is generally invalid when applied to calculating force extension curves of single protein molecules such as those studied by the AFM or optical tweezers techniques. The measured force extension curve turns out to depend on the type of the experiment. Equation (1) applies to macroscopic objects with well defined \( f \) and \( r \). In single molecules, however, these quantities may undergo large fluctuations, which are affected by the way the molecule is attached to the measuring instrument. This is why Eq. (1) is generally inapplicable to single molecule experiments.

This paper is organized as follows. In Sec. II we briefly describe our simulation method and present the results of simulations. In Sec. III we introduce two types of experiments: A rigid force-sensor (bead-type) experiment and a flexible force-sensor (cantilever-type) experiment. We compute the force extension curves for each one of them and show that they are different. The general result is different from that given by Eq. (1). Thus we argue that different experimental techniques applied to the same protein molecules may generally result in different force extension curves. Finally, in Sec. IV we comment on the existing experimental data on the force extension curves of single protein molecules from the present perspective.

II. COMPUTING THE PROBABILITY DISTRIBUTION OF THE END-TO-END DISTANCE \( p_0(r) \)

We use the peptide representation shown in Fig. 2. The conformation of the molecule is specified by a set of dihedral angles, \( \{\phi_1, \phi_2, \phi_3, \ldots, \phi_N\} \) which are assumed to be the only degrees of freedom, while other geometric parameters (bond lengths and angles) remain fixed. Bending or stretching of bonds can become important when the force is large but are not considered here. The side chains \( R_1, R_2, \ldots \) are represented as single atoms. Each atom has a van der Waals radius assigned to it, whose values are taken from and the van der Waals radii are not allowed to overlap. Random chains are generated by selecting random values for the dihedral values and rejecting the chain conformations that violate steric constraints. We use the “chain growth” algorithm, in which monomers are added to the chain one at a time. If, at any step, the new residue violates steric constraints, chain growth is terminated and a new chain is started. The details of our approach will be published elsewhere. The end-to-end distance vector \( r \) is defined as the distance between the outermost \( \alpha \) carbons. By repeating this process many times, we obtain the probability distribution \( p_0(r) \) for this distance. Obviously, the probability density \( p_0(r) \) is a function of \( r = |r| \) only so we will also use the notation \( p_0(r) = p_0(|r|) \) for this quantity. The probability of finding \( r \) between \( r \) and \( r + dr \) is then equal to \( p_0(r) 4 \pi r^2 dr \).

The probability distribution \( p_0(r) \) obtained in this fashion is plotted in Fig. 3 for a peptide chain \( N = 56 \) residues long. According to Eq. (2), the logarithm of this quantity is proportional to the free energy \( F(r) \) of the chain. It is interesting to examine the asymptotic behavior of \( F(r) \) for large \( r \). For Gaussian chains, for example, the free energy is qua-

![FIG. 1. A scheme of a single molecule stretching experiment.](image1)

![FIG. 2. Representation of polypeptide chains used in the Monte Carlo simulations.](image2)
dratic in \( r, F(r) \sim C + ar^2 \), where \( C \) and \( a \) are constants. Excluded volume effects\(^{19}\) lead to a different power law, \( F(r) \sim C + ar^5 \). According to the Flory–Fisher mean field theory,\(^{19}\) the scaling exponent is related to the polymer’s dimensionality: \( \delta = (1 - 3/(d+2))^{-1} \). Thus we expect \( \delta = 5/2 \) for a three-dimensional polymer. An attempt to fit our simulation results for \( F(r) \) with \( C + ar^{5/2} \) however fails rather badly. A different power law, \( F(r) \sim C + ar^4 \), provides nearly a perfect fit of our data, see Fig. 4. From the Fisher–Flory formula, the scaling exponent \( \delta = 4 \) corresponds to the dimensionality \( d = 2 \). This and other anomalous scaling laws found in polypeptides are the subject of a separate paper.\(^{34}\)

Equation (2) then predicts a cubic dependence of the force \( f \) on the extension \( r, f \sim r^3 \). The force extension curves derived from the computed distribution \( p_0(r) \) are discussed in the next section.

III. THE FORCE EXTENSION CURVE \( f(r) \) DEPENDS ON THE TYPE OF MEASUREMENT

Here we derive the microscopic relationships between the mean values of the force \( \langle f \rangle \) and the extension \( \langle r \rangle \). We perform our derivation for two different idealized experimental arrangements shown in Fig. 5.

A. Rigid force-sensor (bead-type) experiment

In this kind of experiment [Fig. 5(a)], a constant force \( f \) is applied to the ends of the chain. This, for instance, can be achieved by attaching a magnetic bead to one end of the chain and applying a magnetic field. In this arrangement, there are no force fluctuations, \( \langle f \rangle = f \). We are interested in how the average displacement \( \langle r \rangle \) is related to \( f \).

To solve this problem we introduce the probability distribution \( p_f(r) \) of the end-to-end distance for a chain that is stretched by force \( f \). If \( U_0 \) is the energy for a given conformation of the chain in the absence of the force then

\[
U_f = U_0 - f \cdot r
\]

is the energy of the same conformation when the chain is stretched by force \( f \). Since the probability of this conformation is proportional to \( \exp(-U_f/k_BT) \), then from Eq. (3) we have

\[
p_f(r) = p_0(r) \exp(\beta f \cdot r)/Q(f),
\]

where \( \beta = (k_BT)^{-1} \). The average extension is given by

\[
\langle r \rangle = \int r p_f(r) dr = \frac{1}{Q(f)} \int r p_0(r) \exp(\beta f \cdot r) dr.
\]
FIG. 6. The force extension curve of the denatured chain that consists of 56 residues. Solid line: a fit to Eq. (1). Dots: average protein extension ⟨z⟩ for various values of the applied force. See text for details.

<table>
<thead>
<tr>
<th>f, pN</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td>0.01</td>
<td></td>
</tr>
</tbody>
</table>

FIG. 7. The probability distribution of the protein extension z in the direction of the applied force plotted for two different values of the force.

\[
\langle r \rangle = \int r p_0(r) \, d^3r \\
= Q(f)^{-1} \int r p_0(r) e^{-βr} \, d^3r \\
= Q(f)^{-1} \int e^{β(r - F(r))} \, d^3r.
\]

(5a)

Let the force \( f \) act along the z axis. Then for the resulting extension \( \mathbf{r} = x \hat{i} + y \hat{j} + z \hat{k} \), we have \( \langle x \rangle = \langle y \rangle = 0 \). Using polar coordinates and keeping in mind that the distribution \( p_0(r) \) is isotropic, one can evaluate \( \langle z \rangle = \langle r \cos \theta \rangle \) from Eq. (5a).

\[
\langle z \rangle = \frac{1}{2} \int 2 \pi r^3 \cos θ \sin θ p_0(r) e^{-βr} \cos θ \, dr \, dθ = \frac{4 \pi \int_0^\infty p_0(r) r/(βf)^2 [βfr \cosh(βfr) - \sinh(βfr)] \, dr}{4 \int_0^\infty dr p_0(r) r \sinh(βfr)/βf}.
\]

(5b)

The usual way to evaluate the integral in (5a) is the method of steepest descents: If the distribution \( p_0(r) \) is sharply peaked at \( \mathbf{r} = \mathbf{r}^* \) then one would find

\[
\langle r \rangle = \mathbf{r}^*.
\]

(6)

The maximum of \( p_0(r) \) is reached for \( \mathbf{r}^* \) satisfying the condition

\[
(d/dr)(λ \cdot \mathbf{r} - F) = 0,
\]

(7)

which is identical to Eq. (1). Again, if we write \( \mathbf{r}^* = x^* \hat{i} + y^* \hat{j} + z^* \hat{k} \) with \( \mathbf{k} \) along the direction of the force \( f \) then \( x^* = y^* = 0 \) and \( z^* \) can be determined from the equation \( F = dF/dr \vert_{z=z^*}. \)

If \( \mathbf{r} \) were the size of a macroscopic object then one could typically neglect the fluctuations in \( \mathbf{r} \). In other words, the distribution of \( \mathbf{r} \) would be extremely sharp and Eq. (6), and consequently Eq. (1), would be correct. For example, when using a measuring tape we do not usually need to worry about the fluctuations in its length. However such fluctuations are not negligible for a single polypeptide chain. In Fig. 6 we plot the dependences \( f(\langle z \rangle) \) and \( f(z^*) \) computed using the probability distribution \( p_0(r) \) from Sec. II. The two dependences are very different. The force extension curve \( f(z^*) \) predicted by the macroscopic thermodynamic relationship (1) is well fitted by a power law \( [F(z^*) - (z^*)^3] \) while the relationship between \( f \) and \( \langle z \rangle \) could not be fitted by any power law.

The origin of the discrepancy between the microscopic force extension curve \( f(\langle z \rangle) \) and the macroscopic law of Eq. (1) is the fact that the distribution \( p_0(z) \) is asymmetric. This is demonstrated in Fig. 7 where this distribution is plotted for two values of the force \( f \). Of course, for a macroscopically large object, the validity of Eq. (1) is guaranteed by the sharpness of the distribution \( p_0(r) \) or \( p(\langle z \rangle) \) and whether or not it is asymmetric becomes unimportant.

For Gaussian polymer chains \(^{35} \) both \( f(\langle z \rangle) \) and \( f(z^*) \) are the same. Indeed, if the probability distribution \( p_0(r) \) is Gaussian, \( p_0(r) \propto \exp(βr^2/2) \), then we have \( \langle r \rangle = \mathbf{r}^* \), regardless of the magnitude of microscopic fluctuations. Thus the deviation of the force-extension curve measured in a rigid force-sensor experiment from Eq. (1) is a consequence of the non-Gaussian nature of the end-to-end distance distribution in the case of polypeptides.

B. Flexible force-sensor (cantilever-type) experiment

The scheme of a cantilever-type experiment is shown in Fig. 5(b) and corresponds to the AFM techniques described in Refs. 6, 7, 11, 13–16 and 36. The protein is attached by its ends to a substrate and to a cantilever, which is treated here as a harmonic spring with a force constant \( k_c \). The total displacement \( \mathbf{L} \) is fixed, which is equal to the sum of the cantilever displacement \( \mathbf{l} \) and the protein extension \( \mathbf{r} \). The mean value of the cantilever displacement \( \mathbf{l} \) provides a way to measure the mean force

\[
(\mathbf{f}) = k_c (\mathbf{l}).
\]

(8)

Thus in this case the force is not a fixed but a fluctuating quantity while the total displacement \( \mathbf{L} \) has a definite value. The measured force-extension curve is the relationship between \( (\mathbf{f}) = k_c (\mathbf{l}) \) and \( \langle \mathbf{r} \rangle = \mathbf{L} - \mathbf{l} \).

To calculate this force extension curve, we note that the probability distribution for the cantilever displacement \( \mathbf{l} \) for any given \( \mathbf{L} \) is proportional to
\[ p_c(l) \propto p_0(L-1) \exp(-\beta k_c l^2/2) \]\n
\[ \propto \exp(-\beta F(L-1) - \beta k_c l^2/2). \]  

(9)

This immediately follows from the fact that the total energy is the sum of the elastic energy of the cantilever \( k_c l^2/2 \) and the energy of the polymer. Therefore the mean value of \( I \) is given by

\[ \langle l \rangle = \int l \exp(-\beta F(L-1)) \]

\[ - \beta k_c l^2/2) d^3V \int \exp(-\beta F(L-1) - \beta k_c l^2/2) d^3l. \]  

(10)

For small \( i \) we can expand the polymer’s free energy in powers of \( I \),

\[ F(L-1) = F(L) - \frac{dF}{dr} \bigg|_{r=L} \cdot 1 + \frac{1}{2} \frac{\partial^2 F}{\partial x^2} \bigg|_{r=L} l^2 + \cdots. \]  

(11)

Consider now the limit of a rigid cantilever \( k_c \gg F''(L) \). In other words, the cantilever force constant is larger than the effective force constant of the polymer for the given extension. In this case, the second and higher order terms in Eq. (11) can be neglected when Eq. (11) is inserted in Eq. (10). The result is

\[ \langle l \rangle = \int l \exp(-\beta F(L)) + \beta (dF/dr) \cdot 1 \]

\[ - \beta k_c l^2/2) d^3V \int \exp(-\beta F(L) - \beta k_c l^2/2) d^3l \]

\[ = k_c^{-1} dF/dr \]  

(12)

or

\[ \langle f \rangle = k_c \langle l \rangle = dF/dr, \]  

(13)

which is Eq. (1). The use of the expansion in Eq. (11) is justified if the cantilever force constant is large and therefore, according to Eq. (13), the displacement is small.

**IV. DISCUSSION**

The reason why Eq. (1) is valid in the case of a cantilever-type experiment is because the fluctuations in both the force \( f \) and extension \( r \) are small, being limited by the thermal fluctuations of the cantilever rather than the polymer itself. In the rigid force-sensor experiment the fluctuations of the extension \( r \) are determined by the force constant of the polymer itself and thus are larger. Deviations from Eq. (13) should be observed if the experiment is performed with a soft cantilever whose force constant is comparable to, or smaller than that of the polymer. Let us estimate the force constant for the polypeptide chain of 56 residues studied here. It is proportional to the curvature of the free energy plot in Fig. 4. Since our Monte Carlo data is noisy, a convenient way to evaluate this curvature is to use a power law fit for the entropy to compute the derivatives. For the extension 120 Å the effective force constant is found to be \[ |d^2F/dr^2| \approx 0.0013 \text{ N/m}. \] The typical cantilever force constant used in the AFM experiments is \( 0.06 \text{ N/m} \), more than an order of magnitude larger. Therefore the condition \( k_c \gg F'' \) is well satisfied in this case. This estimate also suggests that if one uses a cantilever with \( k_c \), an order of magnitude lower, one may see deviations from the thermodynamic relationship (1). As seen from Fig. 4, it may also be possible to increase the force constant of the polymer by stretching it more.

**ACKNOWLEDGMENTS**

This work was supported by the Robert A. Welch Foundation (Z.W. and D.E.M.), Materials Research Laboratory Program of the National Science Foundation (J.B.T.) and by NSF MCB (H.G.H.). Discussions with David Bensimon, Venkat Ganesan, Peter Rossky, and Steve Webber are gratefully acknowledged.