Path-Space Ratio as a Molecular Shape Descriptor of Polymer Conformation

Tomas Edvinsson,*,† Gustavo A. Arteca,‡ and Christer Elvingson†

Department of Physical Chemistry, Uppsala University Box 532, S-751 21 Uppsala, Sweden, and Département de Chimie et Biochimie, Laurentian University, Sudbury, Ontario, Canada P3E 2C6

Received May 4, 2002

Polymers at interfaces exhibit properties that cannot be completely captured by descriptors of mean molecular size. Recent work in the literature shows that a combined analysis of mean size and chain entanglement provides a more discriminating approach to understanding the onset of configurational transitions in these systems. Usually, chain entanglement is characterized by properties such as the mean overcrossing number or the chain's writhe; these are powerful properties but their evaluation can be computationally demanding. In this work, we introduce a geometrical descriptor of polymer shape, termed the *path-space ratio* ζ, aimed at quantifying essential features of chain complexity, but at a lower computational cost. The descriptor includes information on chain geometry and topology. The path-space ratio ζ is built by taking into account two key ideas: (a) a dimensionless measure of length along the backbone of the polymer, and (b) the behavior of topological "knot energies". Here, we compare ζ with other approaches to quantify polymer geometry and connectivity. Particular attention is devoted to the ability of these descriptors to discriminate and quantify conformational changes in grafted polymers under compression. We show that, for these types of applications, the path-space ratio presents a fast alternative to the mean overcrossing number.

1. INTRODUCTION

In recent years, there has been considerable interest in the behavior of polymers attached to different types of surfaces or membranes. These systems are important in a number of applications including the study of lubricants and adhesives[1] as well as for biosensors and in steric stabilization.[2,3] The interest in these problems has also increased by the possibility of experiments that directly manipulate single molecules attached to a surface.[4] The occurrence of chain entanglement is a key feature behind a number of properties in these systems, including the nature of polymer shape transitions, chain response under compression, and possible hysteresis. Chain entanglement is also central for a number of other polymer properties, including flow, viscosity, and swelling of polymeric networks.[5,6] Physically, entanglements arise from the inability of a chain to *cross through* one another under conditions that preserve the main chain bonds (i.e. the chain connectivity). From a structural viewpoint, these entanglements have very different lifetimes depending on whether they are *topological* or *geometrical* in nature. Topological entanglements are "permanent" in the sense that they cannot be removed by any accessible conformational motion where the chain connectivity is maintained; this is the chemical equivalent to a homeomorphism.[7] Such situations may involve a single chain (e.g. a chemical knot) or multiple chains (e.g. links and braids, found sometimes in chemically cross-linked systems such as rubbers).

Topological effects such as knotting play a crucial role in many polymeric systems and in biopolymers. For this reason, mathematical knot theory has been shown to be a convenient tool for studying these phenomena.[7−10] Circular DNA, for example, is known to be knotted differently under the action of topoisomerase enzymes.[9,10] The complexity of these knots can be captured by a descriptor such as the mean overcrossing number (see definition below), which correlates well with the migration speed in gel electrophoresis.[11−13] In contrast, geometrical entanglements are transient, in the sense that they can be removed over time by loop diffusion (within a single chain) or chain diffusion (in a melt).[5] In these cases, we do not have mathematical knots because, topologically, the polymers are open chains. To some extent, however, polymers can act as "effectively" knotted, in the sense that they can trap local loops in knotted configurations if they become sufficiently compactified (e.g., $T \leq \Theta$-temperature). In this case, the open polymer might behave as topologically entangled over a short period of time.

In either case, characterizing the complexity of intra- and interchain entanglements is an important task in order to establish correlations between structural and physical properties. In this work, we discuss an efficient approach to quantify aspects of the self-entanglement complexity of grafted linear chains subjected to geometrical confinement. We start from the notion that purely geometrical descriptors cannot fully capture subtle differences in the trace of a polymer backbone[14,15] or discriminate among different protein folds.[16] Properties such as the mean size and the anisometry depend exclusively on the monomer position, but not on their connectivity. As a result, an approach based on these descriptors would fail to distinguish chains with similar mean sizes, yet different traces of the backbone. This lack of discrimination is a crucial drawback in general, but most particularly whenever chains with similar size differ in topology (e.g., when comparing an open and a knotted chain). On the other hand, a characterization that uses only topological information (e.g., molecular graphs, topological invariants) is also limited, since structures with very different size and energies will be considered equivalent. The most convenient approach is to characterize polymer structure and entanglement complexity by using geometrical descriptors that incorporate information on the chain connectivity. This family of descriptors include the chain writhe,[17] the mean

* Corresponding author e-mail: Tomas.Edvinsson@fki.uu.se.
† Uppsala University.
‡ Laurentian University.

10.1021/ci020269x CCC: $25.00
Published on Web 12/11/2002

overcrossing number (or average crossing number)[18,19] and knot energies.[20,21] In the case of knotted macromolecules, these properties correlate well with electrophoretic gel velocities and sedimentation rates.[10−13] In the case of linear polymers, these descriptors provide insight into the process at play during transitions that conserve the mean size.[22,23] Among these descriptors, recent work show that the mean overcrossing number has a superior performance in regard to discriminating knot complexities.[24] However, an analysis based on these descriptors is hampered by the fact that their computations are CPU-intensive. In this work, we propose and test a simple shape descriptor that provides an efficient alternative to the analysis based on more complex descriptors of chain entanglement.

The work is organized as follows. In the next section, we discuss the background and introduce the proposed molecular shape descriptor, the path-space ratio, denoted by ζ. In the following section, we test this descriptor for small model polymers on a cubic lattice and then apply it to a number of off-lattice simulations of grafted polymer chains. We then compare these results with those obtained when using a number of alternative shape descriptions. The paper ends with a summary of conclusions.

2. PATH-SPACE SHAPE DESCRIPTOR

To understand the motivation for the path-space ratio, let us first briefly consider a very simple descriptor of polymer size, the characteristic ratio, $C_n = <R_{ee}^2>/(n-1)b^2$.[25] Here, $<R_{ee}^2>$ denotes the configurationally averaged square end-to-end distance of a linear chain with bond length b and n monomers. For chains with any stiffness or excluded volume in Θ conditions, C_n approaches a limiting value as n becomes large. For a freely jointed (random) chain, C_∞=1. Despite its simplicity, C_∞ can distinguish between the average conformations of polymers such as polyethylene (C_∞=6.7) and polypropylene (C_∞=5.4) or between 1,4-cis (C_∞=4.9) and 1,4-trans polybutadiene (C_∞=5.8).[26] A comparison between unperturbed freely jointed chains and chains with variable monomer-monomer interaction has also been used to study changes in size and entanglement complexity.[27] Instead of comparing different models, we will presently quantify the shape of the polymer with a molecular shape descriptor that compares the distance along the path of the chain (the contour length) with the distance through space. As we show below, this approach provides a fast alternative to more complex shape descriptors, such as the average crossing number, for the systems studied here.

For the present purposes, a polymer can be described by a parametrized space curve $\Gamma(t)$, $t\in[0, n]$ where $\Gamma(0)$ and $\Gamma(n)$ are the first and last points on the curve Γ. Let us consider n identical monomers, each defined by a position vector $\mathbf{r}_i$ with $i = 1, 2, ..., n$. In the most general case, these monomers are connected by curved segments of the polymer chain. Each segment γ_i can in turn be parametrized over the unit interval $s \in[0, 1]$, as shown in Figure 1

$$\gamma_i:\ [0,1] \rightarrow \mathbb{R}^3, \quad \gamma_i(0) = \mathbf{r}_i, \quad \gamma_i(1) = \mathbf{r}_{i+1} \qquad (1)$$

with continuous parametric derivatives $\dot{\gamma}_i(s) = \delta\gamma_i(s)/\delta s$, $\forall s \in [0, 1]$, so that

$$\dot{\gamma}_i(0) = \lim_{s\rightarrow 0^+} \dot{\gamma}_i(s), \quad \dot{\gamma}_i(1) = \lim_{s\rightarrow 1^-} \dot{\gamma}_i(s) \qquad (2)$$

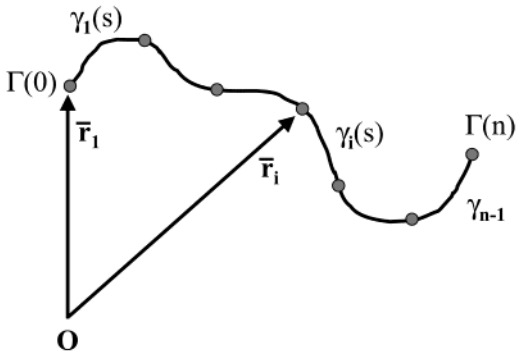

Figure 1. The parametrized representation $\Gamma(t)$ of a space curve with $t \in[0, n]$ and the stepwise parametrization with $\gamma_i(0) = \mathbf{r}_i$ and $\gamma_i(1) = \mathbf{r}_{i+1}$.

Let us now consider two geometrical properties of an open or closed curve: the mean overcrossing number and the knot energy (in our case, a "polygon" energy). These properties provide the background for the descriptor proposed in this work. An overcrossing is simply a projected bond-bond crossing within a chain. The mean overcrossing number, denoted $\bar{N}$, is the number of projected crossings N, averaged over all orthogonal projections.[14,18,28−30] For a general n-polygon, the number $\bar{N}$ can be expressed analytically as a sum of contributions by the crossings between pairs of nonconsecutive bonds[31]

$$\bar{N} = 2\sum_{i=1}^{n-3}\sum_{j=i+2}^{n-1} N_{ij} = \frac{1}{2\pi}\sum_{i=1}^{n-3}\sum_{j=i+2}^{n-1}\int_0^1\int_0^1 \frac{|(\dot{\gamma}_i(s)\times\dot{\gamma}_j(t))\cdot(\gamma_i(s)-\gamma_j(t))|}{||\gamma_i(s)-\gamma_j(t)||^3}\,dsdt \qquad (3)$$

where γ_i is the segment connecting the i and $i+1$ monomers and $\dot{\gamma}_i$ its parametric derivative along the backbone, satisfying eqs (1) and (2). Equation 3 above is a generalization of the Gaussian formula for the linking number of closed loops.[17a,32] The integrals in (3) adopt a relatively simple form in a cubic lattice,[31] in which case the curve Γ is represented by a polygon **C**. In particular, whenever the overcrossing segments are very far apart, the contribution to the crossing number due to a pair of orthogonal bonds scales as: $N_{ij} \sim r_{ij}^{-2}$, where $r_{ij} = ||\mathbf{r}_i - \mathbf{r}_j||$ is the distance between the beginning of the oriented bonds $i \rightarrow i+1$ and $j \rightarrow j+1$, denoted by monomer bead coordinates $\mathbf{r}_i$ and $\mathbf{r}_j$, respectively. In other words, the key ingredients in the computation of $\bar{N}$ are the positions of the monomer beads and their sequential connectivity. Although the function $\bar{N}$ in eq 3 is complicated, its long distance contribution can thus be represented as a double sum of inverse square distances.

A similar situation appears when considering a "knot energy" as a descriptor of chain entanglement.[33−36] A typical "knot energy" is a function defined as a double path integral over the parametrized trace of the molecular polygon **C**

$$E_f^{(k)} = \int_{C\times C}\int \frac{f(\mathbf{r},\mathbf{r}')}{||\mathbf{r}-\mathbf{r}'||^{2k}}\,d\mathbf{r}d\mathbf{r}' \qquad (4)$$

where k is a positive integer and f is a regularization function, chosen so as to eliminate the singularity associated with an actual bond-bond crossing. When considering $k = 1$ and

convenient choices for the regularization functions,[33−36] the knot energies can be written as double sums of square-inverse functions. A commonly used knot energy of type $E^{(1)}$ is the so-called Möbius energy, $E(\gamma)$,

$$E(\gamma) = \int_0^1 \int_0^1 \left(\frac{1}{||\gamma(s) - \gamma(t)||^2} - \frac{1}{D(\gamma(s),\gamma(t))^2} \right) dsdt \quad (5)$$

where $D(\gamma(s),\gamma(t))$ is the arclength of the curve from $\gamma(s)$ to $\gamma(t)$. Note that the latter descriptors of entanglement complexity are related, since for any proper rectifying line γ: $[0, 1] \rightarrow \mathbb{R}^3$, the average crossing number is bound as[33]

$$\bar{N}(\gamma) \leq \frac{1}{2\pi} E(\gamma) \quad (6)$$

Having these notions in mind, one can try to construct a simpler descriptor of entanglement for a space curve, that retains some features of both the average crossing number and knot energy. We proceed as follows. Consider a realistic polymer as described by an optimized potential energy function. By using a realistic potential energy, we ensure that bond vector (phantom) crossings are physically unachievable, thus making it unnecessary to add a regularization function or second term in eq 5. We can instead use the function *f* to incorporate information about the segment connectivity and the arclength along the backbone. In our case, we use a measure of length along the backbone of the polymer.

To motivate our choice of *f*, let us briefly look at another geometric measure of a space curve γ, the *ropelength* $L(\gamma)$.[7] $L(\gamma)$ is the arclength divided by the thickness (where the thickness is the diameter of the largest embedded tubular neighborhood, i.e., the thickest uniform tube centered on γ). This geometric measure has been shown to bound the average crossing number of ideal knots as[37] $\bar{N}(\gamma) \leq CL(\gamma)^{4/3}$ with a universal constant $C \leq 1/4$.[38] This result tells us that the arc length along the backbone of the polymer could also be used to provide a dimensionless measure of entanglement complexity. For a polymer with *n* monomers, the backbone arclength is[39] (see Figure 1)

$$\int_0^n ||\dot{\Gamma}(t)||dt = \sum_{i=1}^{n-1} \int_0^1 ||\dot{\gamma}_i(s)||ds \quad (7)$$

with the notation used in eqs 1 and 2. Using this notation, we can define a "partial length" up to the *i*th monomer as $l_{1i}(\gamma) = \sum_{j=1}^{i} \int_0^1 ||\dot{\gamma}_i(s)||ds$. With these geometrical properties in mind, we propose a shape descriptor involving a single integral over the polymer backbone of a ratio whose numerator measures arc length and whose denominator measures the distance through space. We thus introduce a geometric descriptor of shape, *the path-space ratio*, ζ,

$$\zeta = \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{\sum_{j=1}^{i} \int_0^1 ||\dot{\gamma}_j(s)||ds}{||\gamma_1(0) - \gamma_{i+1}(0)||} \quad (8)$$

normalized over the number of bonds, $n - 1$. The denominator in (8) measures the distance between the first and the *i* th bond ending at the $(i + 1)$th bead, whereas the numerator is the chain arc length, $l_{1i}(\gamma)$, between the same beads (approximately *ib*). This descriptor thus compares the "path",

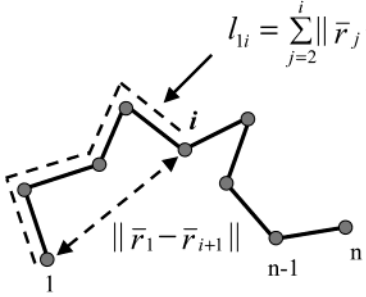

Figure 2. Schematic polymer model with *n* units and $n - 1$ bond vectors, where l_{1i} is the length along the path to the *i*th unit and $||\bar{r}_1 - \bar{r}_{1+1}||$ is the distance through space.

$l_{1i}(\gamma)$, and the corresponding distance in space, averaged over different length scales. For a general curve, however, ζ will depend on the "pivot" bead, and we now define the path-space ratio relative to a reference point α,

$$\zeta_\alpha \equiv \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{\sum_{j=\alpha}^{i} \int_0^1 ||\dot{\gamma}_j(s)||ds}{||\gamma_\alpha(0) - \gamma_{i+1}(0)||}, \quad \alpha \neq i+1 \quad (9)$$

where α can be chosen by considering the symmetry of the space curve or some convenient criterion. For $i < \alpha$, the sum in the numerator should be replaced by $\sum_{j=i}^{\alpha}$. The ratio of distances in eq 9 holds important information about the relative configuration of the curve. One could also construct a similar descriptor that is independent of the point of departure by using a sum over α. However, this would be identical to an average over all possible points of departure and thus destroy the information on relative segment connectivity, represented by length along the path of the space curve.

In the case of a wormlike chain model (e.g., the Kratky-Porod model), curved segments can be useful. In atomistic models, however, one represents both chemical bonds and bonds between monomers by bond vectors. If we consider a polymer model with *n* monomers, where the $n - 1$ segments are represented by straight lines (bond vectors), the path-space ratio for $\alpha = 1$ becomes

$$\zeta_1 = \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{l_{1i}}{||\bar{r}_1 - \bar{r}_{i+1}||} \quad (10)$$

where $l_{1i} = \sum_{j=1}^{i} ||\bar{r}_{j+1} - \bar{r}_j||$ is the length along the path to the *i*th bond (i.e. the $i + 1$ unit), as shown in Figure 2. By restricting the system to a lattice, the formulation becomes particularly simple since l_{1i} becomes *i* (in lattice units). In the present work, however, eq 9 provides the most general approach.

3. COMPUTATIONAL DETAILS

To explore the performance of the path-space ratio in a system with interesting shape transitions, we consider a single permanently grafted chain anchored to an impenetrable surface by the first monomer. This model has been subject to several analyses recently[15,27] thus providing a good example for comparisons among shape descriptors. In the present approach, each molecule has $n = 50$ subunits with a radius $a = 2$ Å. We have studied the behavior under geometrical confinement, as a model for the compression

effects found in e.g. atomic force microscopy. The subunits of a chain are connected by a harmonic potential, where the force constant is chosen to ensure only small fluctuations of the contour length. In the present simulations, the average bond length fluctuations are characterized by $<|b_i - 2a|^2> = a^2/25$, where b_i is the norm of the bond vector $\mathbf{b}_i$ between units i and $i + 1$. The stiffness of the chains is modeled by including a bending potential

$$U_b = \frac{k_b}{2}\sum_{j=1}^{n-2}(\cos\theta_j - \cos\theta_0)^2 \qquad (11)$$

where θ_0 is the angle at minimum potential energy and θ_j the angle between bond vectors $\mathbf{b}_j$ and $\mathbf{b}_{j+1}$. The force constant, k_b, was calculated self-consistently from a given persistence length $\mathscr{P}$.[40,41] In all cases the equilibrium angle θ_0 is equal to 0°. The effect of the solvent selectivity on the polymers are incorporated by a Lennard-Jones potential between nonbonded monomer units

$$U_{LJ} = 4\epsilon\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\left[\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^{6}\right] \qquad (12)$$

The energy parameter, ϵ, is varied from $0.1kT$ to $1.5kT$ where k is the Boltzmann constant, T is the absolute temperature, and $\sigma = 2a$ is equal to the subunit diameter. The persistence length has been varied between 4 and 24 Å. This covers the full range from good solvent to bad solvent conditions. For the present analysis, we have computed only the average shape features for a single polymer chain in the limit of low grafting density where the interactions from other polymers are negligible.

We have used a Monte Carlo (MC) algorithm to find the average properties of the grafted polymers. In each move, a chain unit is moved randomly along the *x*, *y*, and *z*-axes. A new total potential energy for the system is then calculated and acceptance or rejection of this move is established according to the standard Metropolis MC criterion[42] for a temperature $T = 293$ K. The constraining wall is impenetrable; moves that intersect the grafting surface are also rejected. Compression is thus enforced by geometrically confining the chains within a slab of two parallel hard-planes separated by a distance L that varies between 100 and 10 Å. For $n = 50$, a chain inside a slab with $L = 100$ Å is effectively a "free" (i.e., unconfined) mushroom; for $L = 10$ Å we find highly compressed chains (corresponding to ca. 2.5 times the average bond length). The maximum step size is chosen so that the rejection rate is slightly above 50%, giving a representative sampling of the configurational space. The system properties, such as total potential energy, average radius of gyration and the path-space ratio, relax quite fast except for the most confined systems. To have a rigorous sampling we have used 50 different initial structures for every set of parameters, with each trajectory being between 10^6 and $4\cdot10^6$ MC-steps. Coordinates were saved every 2000 steps, and equilibrium values were then calculated by using the last 200 snapshots for each MC trajectory. The resulting full set of coordinates were then used for analyzing the properties of interest. The mean overcrossing number, $\bar{N}$, is calculated with a path-integral approach[31] for the smallest structures and by a projection method[30] for the other structures.

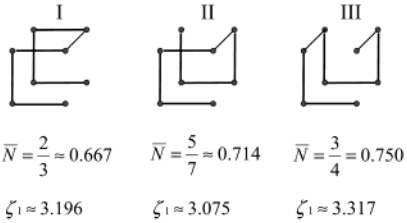

Figure 3. Simple lattice polymers and the corresponding values of $\bar{N}$ and ζ_1. The first unit is chosen to be in the bottom right corner of the cube. Conformations I, II, and III all give the same radius of gyration ($R_g = \sqrt{3}/2 \approx 0.866$) and asphericity ($\Omega = 0$) Both $\bar{N}$ and ζ_1 distinguish among these structures with same atomic coordinates but with different connectivity.

4. RESULTS AND DISCUSSION

4.1. Lattice Model Polymers. Let us consider first some small lattice model polymers in order to illustrate how the descriptor works and to what features it is sensitive. The short chains in Figure 3 have identical monomer positions but have different connectivity. The center of mass is in the middle of the unit cube; the radius of gyration is $R_g = \sqrt{3}/2 \approx 0.866$ and the anisotropy[43] $\Omega = 0$ for all chains. In contrast, $\bar{N}$ and ζ_1 discriminate between different folds of systems with identical atomic coordinates. Note that $\bar{N}$ and ζ_1 do not follow a direct correlation for these small polymers since they describe different local shape features. The average crossing number increases in the polymer series: I, II, and III, ($\bar{N} = 2/3$, 5/7, and 3/4 respectively). Both I and II have three segments in one plane, connected by a middle segment, but I has the largest number of pairs of coplanar bonds, thus yielding the lowest average crossing number. Polymer III, on the other hand, has the largest number of pairs of perpendicular noncoplanar bonds and thus produces more bond-bond projected crossings.

Calculating the path-space ratio for polymer I, for example, in Figure 3 using eq 10 we obtain $\zeta_1(I) = 1/7(1/1 + 2/\sqrt{2} + 3/1 + 4/\sqrt{2} + 5/\sqrt{3} + 6/\sqrt{2} + 7/1) = 3.196$. From the results in Figure 3, we can see that ζ_1 also manages to capture differences in fold between these small polymers with the same size. Indeed, ζ_1 gives also the highest value for polymer III, in agreement with $\bar{N}$. The larger end-to-end distance in II compared to I leads to a lower measure of chain complexity for II (cf. Figure 3). The path-space ratio, although discriminating between the structures, is not as sensitive to folding features near the reference point, a feature which comes from the weighing with the distance along the chain. In turn, this distance-weighing would appear to make the descriptor more sensitive to the global fold. Note also that ζ_1 captures the higher folding complexity in III, although it has the same end-to-end distance as I and smaller end-to-end distance than conformation II. Although these systems are strongly simplified structures, they serve to illustrate how $\bar{N}$ and ζ_1 convey shape locally.

The same picture emerges from Figure 4, where we have tested the dependence of ζ on the starting point, α. In this example, we have fixed the position of the first four segments and changed the fold for the last four segments in polymer A, B, and C, while keeping the end-to-end distance fixed. The different configurations can be regarded as a simple "unfolding process" from A to C. Table 1 gives a summary of the size and shape parameters. Note that, in general, $\bar{N}$ is not a rational number. In the case of conformation A in Figure 4, the exact result is indeed rational $\bar{N} = 5/6$, in close

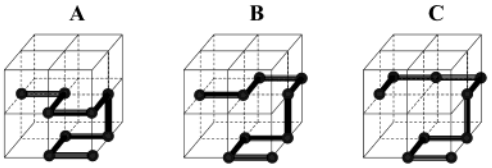

Figure 4. A representation of a local unfolding for a short model polymer. The first four segments are fixed while the change in the last four segments represents a simple unfold from configuration A to C. As in Figure 3, the first unit is at the bottom right corner of the lattice. The corresponding values for R_g, Ω, $\bar{N}$, and ζ_α are in Table 1. Here we see that ζ_1 captures the unfolding from A to C as the other descriptors, while the behavior of ζ_5 and ζ_9 illustrates that ζ_α is not sensitive to i) changes in the nearest lattice unit and ii) changes that are symmetric with respect to the reference point α.

Table 1. Shape Descriptors for Configurations A, B, and C in Figure 4[a]

polymer	R_g	Ω	$\bar{N}$	ζ_1	ζ_5	ζ_9
A	0.9686	0.02095	0.8338	3.0991	1.7986	2.4428
B	1.0657	0.03946	0.6623	2.5295	1.7986	2.4428
C	1.1863	0.05523	0.5066	2.3159	1.5913	2.4428

[a] The standard deviation for $\bar{N}$ is 0.0024, 0.0021, and 0.0011 for A, B, and C respectively.

agreement with our numerical estimation using random projections.[30] Using the method in ref 31, a straightforward yet cumbersome calculation for the other two conformers gives the following: $\bar{N}(B) = 1/3 + (3/2\pi)\arctan(\sqrt{6}/3) \simeq 0.66026...$ and $\bar{N}(C) = -1/12 + (1/2\pi)\{3\arctan(\sqrt{6}/3) + \arctan(\sqrt{6}/12) + \arctan(\sqrt{3}/3) + \arctan(4/3)\} \simeq 0.50656...$, also in good agreement with the numerical estimates. We have calculated ζ_α with $\alpha = 1, 5, 9$ which represent the first, middle, and last unit, respectively, to illustrate how the path-space ratio translates local shape features. When calculating ζ from $\alpha = 1$, we see a sequential unfolding from A to B to C, as reflected in a decrease in both $\bar{N}$ and ζ_1. Starting from the middle, on the other hand, ζ_5 is insensitive to the change from A to B, as the flip of segment 5−7 maintains the symmetry. With the same analysis, we find that ζ_9, which is the path-space ratio from the last unit, is totally unaffected by the change in the first four segments. This finding illustrates a trend observed throughout, where ζ_1 appears to be a potent shape descriptor, which in addition is very computationally efficient.

4.2. Geometrically Confined Polymer Mushrooms. Our previous analysis shows that the path-space ratio, ζ_α, captures some essential features of molecular shape (e.g. folding complexity), not conveyed by the mean molecular size and anisometry. In this section, we consider a more challenging example where molecular size and folding complexity are known to behave differently, namely, the case of grafted polymers under geometrical confinement.[15,27,44] (Confinement is represented here by placing the polymer between two parallel hard planes separated by a distance L, as described in section 3.)

To test the performance of ζ_α, we have computed the configurational averages of shape descriptors for 50-bead polymers with various interactions and levels of confinement. Let us start by considering the compression of a polymer with weak nonbonded interaction. A natural choice of reference unit, α, for the grafted polymers is the anchoring unit fixed on the surface. Figure 5 shows the variation in

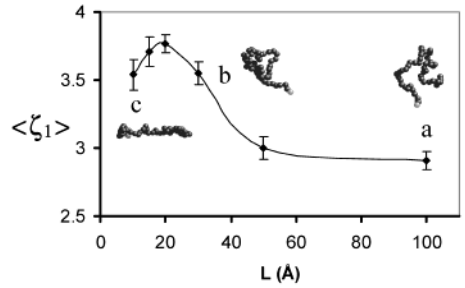

Figure 5. The configurationally averaged path-space ratio $<\zeta_1>$ versus the compression L for a grafted polymer with $\epsilon = 0.25kT$ and $P = 8$ Å. As the compression progresses the polymer conformation becomes more compact until further compression results in "flattening" (see text).

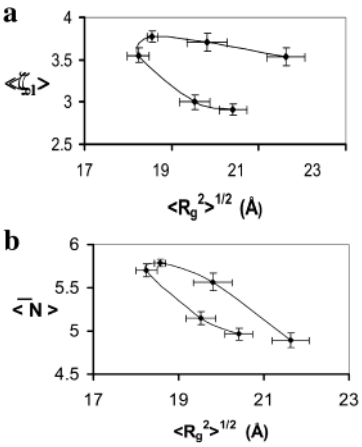

Figure 6. (a) The path-space ratio $<\zeta_1>$ versus $<R_g^2>^{1/2}$ in a "shape-transition" map for a grafted polymer with $\epsilon = 0.25kT$ and $P = 8$ Å. The lower branch ($L > 20$ Å) corresponds to the change in shape as a free mushroom is compressed. The upper branched ($L < 20$ Å) corresponds to the "strong confinements" regime.[15] (b) $<\bar{N}>$ as a function of $<R_g^2>^{1/2}$ for the same parameters as in (a).

folding features, in terms of the configurationally averaged path-space ratio, $<\zeta_1>$, for $\epsilon = 0.25kT$ and persistence length $P = 8$ Å. As the compression progresses from L = 100 Å to 10 Å, the path-space ratio increases up to a maximum around L = 20 Å and then decreases in the *regime of strong confinement* ($L < 20$ Å). At this confinement level, the chains are flattened as signaled by a decrease of $<\zeta_1>$. This can also be monitored by changes in asphericity and radius of gyration.[15,27,44] However, the extrema in chain entanglement and molecular size do not coincide. These results are illustrated in Figure 6, which shows the variation of $<\zeta_1>$ versus $<R_g^2>^{1/2}$, and $<\bar{N}>$ versus $<R_g^2>^{1/2}$ for the same system. We can recognize two distinct behaviors. First, the chains reach a minimum in $<R_g^2>^{1/2}$, but further confinement results in an increase in molecular size. This behavior can be contrasted with that for the two descriptors of chain complexity in Figure 6. Figure 6b shows that the chain entanglement, $<\bar{N}>$ does not exactly reach a maximum at the minimum of $<R_g^2>^{1/2}$. As Figure 6a shows, the path-space ratio, $<\zeta_1>$, gives similar information to that conveyed by $<\bar{N}>$. In Figure 7a we present the radius of gyration normalized with respect to the mean bond length $<||\mathbf{b}||>$ as a function of the energy parameter ϵ, for uncompressed grafted polymers with $P = 16$ Å. For the range of values $\epsilon = 0.1kT$ to $0.75kT$, the figure reveals a correlation between the path-space ratio and the radius of gyration. In this regard

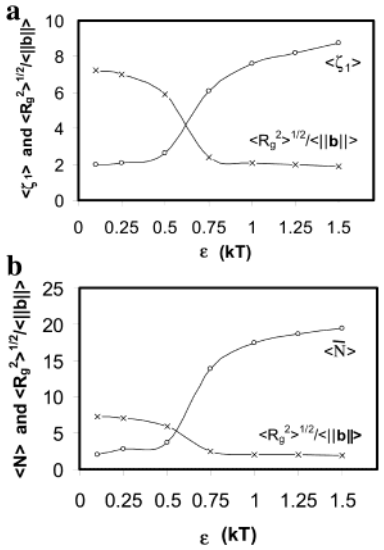

Figure 7. (a) Variation of $<\zeta_1>$ and $<R_g^2>^{1/2}$ in terms of ϵ for an uncompressed grafted chain with $P = 16$ Å. The standard deviation was 2% for small ϵ and <1% for high ϵ. (b) Variation of $<\bar{N}>$ and $<R_g^2>^{1/2}$ in terms of ϵ for an uncompressed grafted chain with $P = 16$ Å. The standard deviation was below 1% for all ϵ.

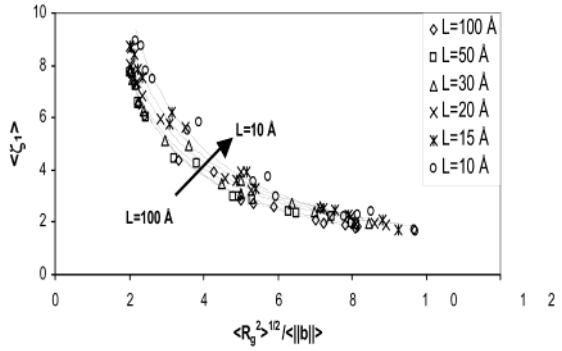

Figure 8. Shape descriptors for the collected simulations with $P = 4$, 8, 16, 24 Å and $\epsilon = 0.1$, 0.25, 0.5, 0.75, 1.0kT. The data suggest a narrow range of accessible molecular shapes.

the path-space ratio could be seen qualitatively as another descriptor of mean molecular size. However, $<\zeta_1>$ also detects other subtler effects not observed with descriptors of mean molecular size and asphericity. When the effective interaction between the beads is increased from $\epsilon = 0.75kT$ to $\epsilon = 1.5kT$, we see that $<R_g^2>^{1/2}/<||\mathbf{b}||>$ is essentially constant, but $<\zeta_1>$ indicates a further increase in folding complexity. In Figure 7b, we compare this with the behavior of $<\bar{N}>$, known to capture effects of loop entanglement.[45] Here we see that $<\bar{N}>$ shows a similar change in the interval $\epsilon = 0.75kT$ to $\epsilon = 1.5kT$, as well as a correlation between the overcrossing number and $<R_g^2>^{1/2}$ in the interval $\epsilon = 0.1kT$ to $\epsilon = 0.5kT$. Also in the high ϵ regime, $<R_g^2>^{1/2}$ still remains essentially constant, whereas $<\zeta_1>$ and $<\bar{N}>$ continue to increase. For the same curve with $P = 8$ Å (not shown), $<\zeta_1>$ increases with a factor ~1.5 in the interval $\epsilon = 0.5kT$ to $\epsilon = 0.75kT$, whereas $<\bar{N}>$ increases also in the interval $\epsilon = 0.25kT$ to $\epsilon = 0.5kT$. Despite these small differences, it is clear that $<\zeta_1>$ provides a reasonable alternative to $<\bar{N}>$ as a descriptor of molecular shape for confined polymers.

Figure 8 shows all the results for $<\zeta_1>$ and molecular size during compression of model polymers with a range of

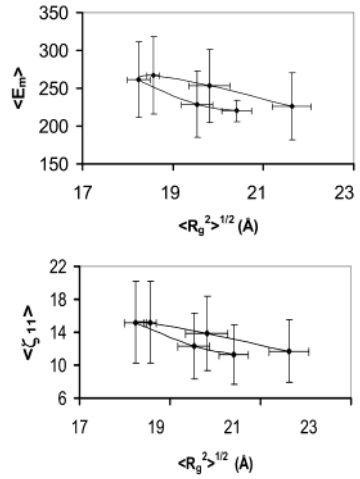

Figure 9. Möbius energy versus radius of gyration (top) and a "double-sum" path-space ratio versus the radius of gyration (bottom). The parameters are the same as in Figure 6.

stiffnesses and solvent qualities. The figure include the data for $P = 4$, 8, 16, 24 Å and $\epsilon = 0.1$, 0.25, 0.5, 0.75, 1.0kT for confinements $L = 100$, 50, 30, 20, 15, 10 Å. Here we have normalized the radius of gyration with the averaged bond length $<||\mathbf{b}||>$. (Note that the switch in behavior from low to high confinement now shows up at a larger scale.) This figure depicts the essential structural characteristics during compression for various (ϵ, P) values. The plot indicates that the polymers adopt a relatively narrow range of shape features for *all* different ϵ and P values at $L = 100$ and $L = 50$ Å where the grafted polymers are more or less unperturbed, whereas confinement below 50 Å produces a "shift" in shape features for all (ϵ, P) values. This can be compared to the corresponding behavior with $<\bar{N}>$, which shows a similar, yet narrower, range of shape descriptor values.[46] Recalling that ζ_1 is more sensitive to changes further away from the reference point (the grafting unit in this case), we can thus attribute the increase in $<\zeta_1>$ to the response of the end of the chain as the confinement increases. In the case of very strong confinement ($L < 20$ Å), the variation in $<\zeta_1>$ reflects a relative insensitivity to the position of the flattening transition (cf. Figure 6a) for high ϵ and P values.

4.3. Comparison with Other Descriptors based on Backbone Path Distances. For comparison, we have also explored two other descriptors that involve similar information included in $<\zeta_1>$. In Figure 9 (top) we can see the performance of the Möbius energy for the same parameters used in Figure 6. As the figure shows, the fluctuation in the Möbius energy is quite large, and the possibility to discriminate any structural changes during compression is lost. The same behavior is seen in Figure 9 (bottom) which shows the performance of a *double sum* path-space ratio,

$$\zeta_{11} = \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{l_{ij}}{||\bar{r}_i - \bar{r}_j||} \qquad (13)$$

which is the double sum analogue to eq 10 with l_{ij} being the distance from monomer i to j along the polymer backbone. This descriptor resembles an average of ζ_α over all possible reference points, $\alpha = 1, 2, 3, ..., n$. Also in this case, configurational fluctuations render the descriptor essentially "blind" to compressions.

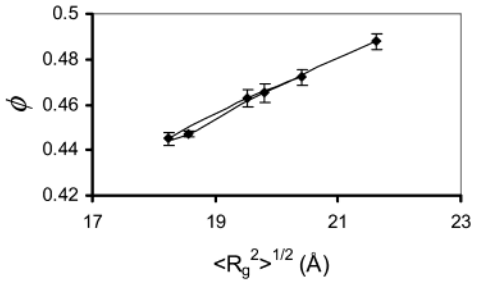

Figure 10. The leading eigenvalue for a D/D-matrix versus the radius of gyration, for the same parameters as in Figure 6.

A distance-distance matrix (D/D-matrix) proposed by Randić et al.[47−49] provides another alternative descriptor, built with similar structural information. Here, the distance in space between the two beads i and j, $||\bar{r}_{ij}||$, is divided by their closest distance along the path of the molecular graph, l_{ij}. The resulting matrix, M, with the matrix elements $M_{i,j} = ||\bar{r}_{ij}||/l_{ij}$, can then serve as a source for structure invariants. The leading eigenvalue of M has been proposed as a structural invariant and serves as a quantitative measure of folding in chains.[48,49] The authors have also suggested a sequence of matrices with successively higher powers, k, as additional sources for descriptors of folding. (The resulting matrices, $M^{(1)}$, $M^{(2)}$ ..., $M^{(k)}$, have elements $M_{i,j}{}^{(k)} = ||\bar{r}_{ij}||^k/(l_{ij})^k$.) The proposal is to use the list of the corresponding normalized leading eigenvalues, $\phi^{(1)},\phi^{(2)}$ ..., $\phi^{(k)}$, as a "folding profile". We test this idea in Figure 10 which shows the leading eigenvalue ϕ constructed from the coordinates of the grafted polymers in the simulation described in section 4.2. As we can see in Figure 10, this descriptor discriminates between the different conformations but has the same extremum as the mean molecular size. In other words, the two compression branches in Figure 6 (corresponding to free and flat mushrooms) are collapsed into one. Higher powers of k do not improve the discrimination as it can be seen in Table 2, where we have included the leading eigenvalue of the D/D-matrix, raised to the seventh power, $<\phi^{(7)}>$. This was chosen as an arbitrary but relatively large number, showing the effect of a higher power on the discrimination of the structures in our systems. For midrange to large system sizes, Randic et al.[49] suggest the use of averaged row sums. Using this approach we found the same situation as Figure 10 but with larger error bars. In summary, the use of distance matrices is not a convenient approach for the problem at hand. Once again, ζ_α appears to provide the simplest descriptor to characterize folding features using the information on distances (r_{ij}, l_{ij}).

4.4. Alternative Descriptors Based on Simpler Projection Schemes. Another approach to quantify folding complexity is to find a simpler projection method for the average crossing number. For the sake of brevity, we will only compare approaches based on projections along Cartesian axes. A method based on projections onto the xy, yz, and zx-planes has been suggested;[50] in that approach entanglements occur only if all three projections result in a projected bond−bond crossing. The authors use this notion of chain overlap in an analysis of the *inter*-chain overlap in a dilute solution. Here, we have adopted it to study intrachain overcrossings in order to compare with ζ_α. We have briefly tested three variants of this projection scheme denoted as follows: a) $<cube>$ which is the average overcrossing number when projecting along the three axes, b) $<cube_{min}>$ which is the average overcrossing number when choosing the projection giving the least overcrossings, and finally c) $<cube_{xy}>$ when projecting along z-axis onto the xy-plane. The first projection scheme can be seen as an simplified average crossing number, the second one a coarse-grained version of the minimum crossing number, and the last one is chosen by considering the overcrossings from a "top view". (The latter approach could be useful in systems with a qualitative different direction, e.g., polymers compressed by as approaching object.)

The results of these three approaches are summarized in Table 2 which compares the mean size, the "cube" projections and the results of $<\phi^{(7)}>$ according to Randić. The following observations can be made: i) $<cube_{min}>$ correlates inversely with the D/D-matrix. Note that $<cube_{min}>$ has a maximum as a function of compression. ii) The other two projection schemes show no flattening transition, since $<cube>$ monotonically increases (within the error bars) and $<cube_{xy}>$ monotonically decreases. Thus, for these systems, simpler projection methods give no reliable shape discrimination or, as in the case with $<cube_{min}>$, a similar shape description to that of the mean molecular size.

Recently a single-projection approach has been conjectured to scale with the average crossing number.[51] A projection along one of the axis of a cubic lattice is used and the sites i on the projection of a lattice polymer which are visited $m_i > 1$ times are considered. An entity B called *opacity* is then defined by counting the pairs of visits to the same site

$$B = \frac{1}{2n}\sum_i (m_i - 1)m_i \qquad (14)$$

where n is the number of beads in a lattice with i sites. If a site has no double visit, the contribution to the *opacity* is zero. This measure disregards the connectivity in the polymer and only considers the overlap of monomer beads in an arbitrary chosen direction. Although B would appear to satisfy the correct scaling behavior for $n \gg 1$, this entity is not designed for studying folding features and it lacks the discriminating ability of the average crossing number or the path-space ratio. For instance, the three different conformations in Figure 4 give the same opacity when projecting from right to left (or vice versa). Clearly, the proper characterization of particular conformations requires a folding descriptor that incorporates all angular information, for which ζ_1 provides a simple, convenient alternative.

Table 2. Results for Projection Schemes along Cartesian Axes and the Result for the Average Leading Eigenvalue for a D/D-Matrix with the Elements Powered to 7

L (Å)	$<R_g^2>^{1/2}$ (Å)	$<cube>$	$<cube>_{min}>$	$<cube_{xy}>$	$<\phi^{(7)}>$
100	20.41 ± 0.32	4.79 ± 0.10	1.77 ± 0.10	5.05 ± 0.10	0.1067 ± 0.0007
50	19.53 ± 0.35	4.65 ± 0.10	1.78 ± 0.07	4.59 ± 0.14	0.1055 ± 0.0006
30	18.24 ± 0.25	5.36 ± 0.12	2.36 ± 0.11	4.06 ± 0.09	0.1024 ± 0.0004
20	18.56 ± 0.13	5.33 ± 0.11	2.07 ± 0.06	3.14 ± 0.04	0.1022 ± 0.0002
15	19.81 ± 0.46	5.73 ± 0.20	1.79 ± 0.09	2.34 ± 0.06	0.1051 ± 0.0006
10	21.63 ± 0.43	5.85 ± 0.20	1.11 ± 0.04	1.31 ± 0.03	0.1102 ± 0.0006

5. CONCLUSION

We have shown here that the path-space ratio, ζ_α, captures some essential features of molecular shape which go beyond those conveyed by the mean molecular size and asphericity. By using distances through space and along the backbone path, the path-space ratio is able to derive a measure of global shape from variations of local arclength and chain curvature over different scales. The same geometrical information is incorporated in descriptors of entanglement complexity and folding features, such as the mean overcrossing number. The results in this work suggest that ζ_α provides an efficient alternative to the latter descriptors.

The examples chosen in this work show that $<\zeta_1>$ is a sufficiently informative descriptor to describe the mean shape behavior of grafted polymers. It is possible that the analysis of other conformational transitions may require more detailed descriptors of geometry and connectivity. Work on extending the present descriptor will be discussed elsewhere.

ACKNOWLEDGMENT

G.A.A. is supported by NSERC (Canada) and the Canada Research Chairs' Program, and C.E. is supported by Uppsala university.

REFERENCES AND NOTES

(1) Fleer, G. J.; Cohen-Stuart, M. A.; Scheutjens, J. M. H. M.; Cosgrove, T. *Polymers at Interfaces*; Chapman and Hall: London, 1993.
(2) van Zanten, J. H. Terminally Anchored Chain Interphases: Their Chromatographic Properties. *Macromolecules* **1994**, *27*, 6797.
(3) Ji, H.; de Gennes, P. G. Adhesion via connector molecules: the many-stitch problem. *Macromolecules* **1993**, *26*, 520.
(4) Strick, T.; Allemand, J.-F.; Croquette, V.; Bensimon, D. The manipulation of single biomolecules. *Phys. Today* **2001**, *54*, 46.
(5) de Gennes, P.-G. *Scaling Concepts in Polymer Physics*; Cornell University Press: Ithaca, NY, 1979.
(6) Wiegel, F. W. In *Phase transitions and critical phenomena*; Domb, C.; Lebowitz, J. L., Eds.; Academic Press: London, 1983.
(7) Whittington, S. G., Sumners, D. W., Lodge, T., Ed.; *Topology and geometry in polymer science*; Springer: New York, 1998.
(8) Kholodenko, A. L.; Vilgis, T. A. Some geometrical and topological problems in polymer physics. *Physics Rep.* **1998**, *298*, 251.
(9) Mesirov, J. P., Schulten, K., Sumners, D. W., Ed.; *Mathematical approaches to biomolecular structure and dynamics*; Springer: New York, 1996.
(10) Stasiak, A., Katritch, V., Kauffman, L. H., Ed.; *Ideal knots*; World Scientific: Singapore, 1998.
(11) Stasiak, A.; Katritch, V.; Bednar, J.; Michoud, D.; Dubochet, J. Electrophoretic mobility of DNA knots. *Nature* **1996**, *384*, 122.
(12) Katritch, A.; Bednar, J.; Michoud, D.; Scharein, R. G.; Dubochet, J.; Stasiak, A. Geometry and physics of knots. *Nature* **1996**, *384*, 142.
(13) Vologodskii, A. V.; Crisona, N. J.; Laurie, B.; Pieranski, P.; Katritch, V.; Dubochet, J.; Stasiak, A. Sedimentation and electrophoretic migration of DNA knots and catenanes. *J. Mol. Biol.* **1998**, *278*, 1.
(14) Arteca, G. A. Fluctuations in molecular size, entanglement complexity, and anisometry along molecular dynamics trajectories of short α-helices. *Biopolymers* **1995**, *35*, 393.
(15) Edvinsson, T.; Elvingson, C.; Arteca, G. A. Variations in molecular compactness and chain entanglement during the compression of grafted polymers. *Macromol. Theory Simul.* **2000**, *9*, 398.
(16) Arteca, G. A.; Tapia, O. Characterization of fold diversity among proteins with the same number of amino acid residues. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 642.
(17) (a) Fuller, F. B. The writhing number of a space curve. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 815. (b) Orlandini, E.; Tesi, M. C.; Whittington, S. G.; Sumners, D. W.; Janse van Rensburg, E. J. The writhe of a self-avoiding walk. *J. Phys. A* **1994**, *27*, L333.
(18) Janse van Rensburg, E. J.; Sumners, D. W.; Wasserman, E.; Whittington, S. G. Entanglement complexity of self-avoiding walks. *J. Phys. A* **1992**, *32*, 1607.
(19) Arteca, G. A. Overcrossing spectra of protein backbones: Characterization of 3D molecular shape and global structural homologies. *Biopolymers* **1993**, *33*, 1829.
(20) O'Hara, J. Energy of a knot. *Topology* **1991**, *30*, 241.
(21) Moffatt, K. The energy spectrum of knots and links. *Nature* **1990**, *347*, 367.
(22) Grosberg, A. Yu.; Feigel, A.; Rabin, Y. Flory-type theory of a knotted ring polymer. *Phys. Rev. E* **1996**, *54*, 6618.
(23) Arteca, G. A.; Velázquez, I.; Reimann, C. T.; Tapia, O. Unfolded in vacuo lysozyme folds into native, quasi-native and compact structures. *Phys. Rev. E* **1999**, *59*, 5981.
(24) Huang, J.-Y.; Lai, P. Y. Crossings and writhe of flexible and ideal knots. *Phys. Rev. E* **2001**, *63*, 021506.
(25) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Wiley: New York, 1969.
(26) Mattice, W. L.; Suter, U. W. *Conformational Theory of Large Molecules*; Wiley: New York, 1994.
(27) Arteca, G. A.; Edvinsson, T.; Elvingson, C. Compaction of grafted wormlike chains under variable confinement. *Phys. Chem. Chem. Phys.* **2001**, *3*, 3737.
(28) Arteca, G. A.; Mezey, P. G. The shapes of macromolecular backbones: Three-dimensional characterization by spherical shape maps. *Biopolymers* **1992**, *32*, 1609.
(29) Arteca, G. A.; Nilsson, O.; Tapia, O. Global characterization of protein secondary structures. Analysis of computer-modeled protein unfolding. *J. Mol. Graph.* **1993**, *11*, 193.
(30) Arteca, G. A. Scaling behavior of some molecular shape descriptors of polymer chains and protein backbones. *Phys. Rev. E* **1994**, *49*, 2417.
(31) Arteca, G. A. Path integral calculation of the mean number of overcrossings in an entangled polymer network. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 550.
(32) Fuller, F. B. Decomposition of the linking number of a closed ribbon: A problem from molecular biology. *Proc. Natl. Acad. Sci. U.S.A.* **1978**, *75*, 3557.
(33) Freedman, M. H.; He, Z.-X.; Wang, Z. Möbius energy of knots and unknots. *Annals Math.* **1994**, *139*, 1.
(34) Kusner, R. B.; Sullivan, J. M. In *Proceedings of the Georgia International Topology Conference, 1993*; Kazez, W. H., Ed.; Am. Math. Soc. Press: 1997.
(35) Simon, J. In *Mathematical approaches to biomolecular structure and dynamics*; Mesirov, J. P., Schulten, K., Sumners, D. W., Eds.; Springer: New York, 1996. (ref 9).
(36) Buck, G.; Orloff, J. A simple energy function for knots. *Topology Applications* **1995**, *61*, 205.
(37) Cantarella, J.; Kusner, R. B.; Sullivan, J. M. Tight knot values deviated from linear relation. *Nature* **1998**, *392*, 237.
(38) Buck, G. Four-thirds power law for knots and links. *Nature* **1998**, *392*, 238.
(39) Thorpe, J. A. *Elementary Topics in Differential Geometry*; Springer-Verlag: New York, 1979.
(40) Allison, S. A.; Sorlie, S. S.; Pecora, R. Brownian dynamics simulations of wormlike chains: dynamic light scattering from a 2311 base pair DNA fragment. *Macromolecules* **1990**, *23*, 1110.
(41) Håkansson, C.; Elvingson, C. Semiflexible chain molecules with nonuniform curvature. 1. Structural properties. *Macromolecules* **1994**, *27*, 3843.
(42) Frenkel, D.; Smit, B. *Understanding Molecular Simulation*; Academic Press: San Diego, 1990.
(43) Rudnick, J.; Gaspari, G. The asphericity of random walks. *J. Phys A* **1986**, *19*, L191; Shape of random walks *Science* **1987**, *237*, 384.
(44) Arteca, G. A. Molecular shape transitions in grafted polymers under geometrical confinement. *Int. J. Quantum Chem.* **1996**, *60*, 1515. Shape transitions in polymer mushrooms compressed by a finite-size obstacle **1997**, *65*, 519.
(45) Arteca, G. A. A detailed shape characterization of regular polypeptide conformations. *Can. J. Chem.* **1995**, *73*, 241.
(46) Edvinsson, T.; Elvingson, C.; Arteca, G. A. Effect of compression on the molecular shape of polymer mushrooms with variable stiffness. *J. Chem. Phys.* **2002**, *116*, 9510.
(47) Hall, L. H. In *Computational Chemical Graph Theory*; Rouvray, D. H., Ed.; Nova Scientific: New York, 1990.
(48) Randic, M. On characterization of chemical structure. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 672.
(49) Randic, M.; Krilov, G. On a characterization of the folding of proteins. *Int. J. Quantum Chem.* **1999**, *75*, 1017.
(50) Brostow, W.; Drewniak, M.; Medvedev, N. K. Brownian dynamics simulation of dilute polymer solutions: chain overlap and entanglements. *Macromol. Theory Simul.* **1995**, *4*, 745.
(51) Grassberger, P. Opacity and entanglement of polymer chains. *J. Phys. A* **2001**, *34*, 9959.

CI020269X