Scaling behavior of some molecular shape descriptors of polymer chains and protein backbones

Gustavo A. Arteca
Département de Chimie et Biochimie, Laurentian University, Ramsey Lake Road, Sudbury, Ontario, Canada P3E 2C6
(Received 5 August 1993)

Some *global folding features* of macromolecular chains are characterized using a family of molecular shape descriptors. These descriptors are derived from the following basic notion: the probability of observing N "crossings" between bonds (i.e., double points or *overcrossings*) when a rigid placement of a polymer backbone is projected onto two dimensions. The approach combines simple elements of geometry and topology of linear polymers, and it quantifies the compactness and the complexity of chain entanglements in three-space. The asymptotic behavior of the shape descriptors has been determined as a function of the chain length. It is found that the configurational averages of the most probable number of overcrossings N^*, the mean number of overcrossings $\bar{N}$, and the largest probability of overcrossings A^*, obey power laws in terms of the number of monomers. The critical exponents have been estimated numerically for random-walk polymers with excluded-volume, as well as for a large number of experimental protein backbones. The results indicate that the scaling behavior is little affected by the configurational state of the polymers, since virtually the same exponents are obtained for both "swollen" and "compact" structures. The same scaling behavior is found in polymers with various excluded-volume interactions and in a set of 197 proteins. The mean number of overcrossings in proteins is well described by a simple law: $\bar{N} \approx 0.045 n^{1.4}$, where n is the number of amino acid residues. The shape descriptors for proteins show little dispersion away from the asymptotic regime, whereas a less uniform behavior is found in the radius of gyration. The results complement the analyses based on other more familiar (geometrical) descriptors, and provide some insights into the large-scale folding structure of polymer chains and proteins.

PACS number(s): 87.15.He, 82.20.Wt, 05.90.+m

I. INTRODUCTION

In this work, we study the asymptotic behavior of large-scale folding features of macromolecular chains. To describe these features, we use a family of molecular shape descriptors that characterize the type and complexity of "entanglements" in rigid polymer configurations.

Many properties of chain macromolecules depend on their fold in three-space. For instance, the protein function is strongly dependent on the tertiary structure [1–3], which is determined by the array of main-chain atoms in space as well as by their connectivity. The simple array of atoms in space is specified by the *molecular geometry*. However, other properties of polymers depend not too strongly on the details of the geometry, but rather on the *topology of the chain*. A great deal of work has been devoted to the study of the dependence of both dynamic and thermodynamic properties of polymers on their topological state (e.g., see Refs. [4–7] and others quoted therein). In this work, we follow an approach that combines elements of both the geometry *and* topology of the chain.

A rigid three-dimensional (3D) polymer configuration can be somehow characterized in terms of a number of functions derived from the molecular geometry. We shall refer to these functions as *molecular geometry descriptors*, since they do not depend on the chain connectivity. The term *molecular shape descriptor* will be reserved for functions that also take into account the connectivity. Only these latter descriptors can characterize the degree of folding in a chain in addition to its compactness.

Some geometrical parameters can be used to study large-scale (global) properties of polymers. Among them, we can mention the end-to-end chain distance and the radius of gyration [6,8–10], the principal moments of inertia (and measures of anisotropy derived from them, e.g., asphericity) [11–14], as well as the helical content in proteins [9,15]. Other geometrical descriptors, such as persistence length [6,9,10], molecular kurtosis [16], and measures of structural deformation [17–20] serve to assess the chain's flexibility. These functions describe mostly large-scale size and compactness of a backbone. Since we are presently interested in the global properties of a chain, descriptors dealing mostly with the local properties [21–28] will be excluded.

Topological methods provide an approach to describing the shape of a polymer that is valid for entire subsets of configurations. Some limited aspects of the molecular geometry and connectivity are used to characterize the topological state of a macromolecule. The resulting description is independent of properties such as polymer size and compactness. In contrast, typical properties that are important include: probability of knot formation [5,7,29–35] and enumeration of knots [5,34–37]; writhing, linking, and twisting numbers [7,38–43] loop-threading [30,44,45]; and knot invariants, including polynomials and the minimum number of crossings [34,35,46,47].

In this work, we preserve the information about each actual polymer conformation. A deformed chain is re-

garded as a new configuration with, possibly, a different molecular shape. Consequently, purely topological techniques are not convenient to us since they do not differentiate between the conformations derived from one another by homeomorphic deformations. Similarly, purely geometrical approaches (e.g., the ones based on the radius of gyration or asphericity) exclude connectivity and therefore are not convenient to recognize folding features common to two different conformations. The method we use here describes the *shape* of a macromolecule by a hybrid technique that combines elements of the chain's geometry and topology [48,49]. The molecular geometry *and* connectivity are used to determine the so-called "overcrossings" [49] or "double points" [34] in the 2D projection of a rigid polymer placement. These points are defined by pairs of bonds that cross over each other when the backbone is viewed along a given direction in three-space. The analysis of *all possible projections* provides a characterization of the type and complexity of entanglements in the polymer chain. In other words, the geometry and connectivity are used to convey aspects of the polymer "topology" (i.e., its folding features) by means of a *global shape descriptor:* the probability distribution of overcrossings. The essential features of this descriptor are very simple: *the more entangled the backbone, the more probable the observation of larger numbers of overcrossings.* The *moments of the probability distribution* provide additional descriptors to characterize the subtleties of the polymer fold. The method was originally designed to study rigid protein motifs [48–50]; recently, similar notions have been introduced to measure entanglement complexity in self-avoiding walks [51].

Overcrossing probabilities [52] have been used mostly as tools to monitor dynamic and conformational flexibility in chain and cyclic molecules [52,53]. Overcrossing descriptors appear to discriminate well between folding features [50,52] and complement the information provided by geometric properties. However, a more detailed contrast between molecular shape descriptors and molecular geometry descriptors is lacking. Moreover, very few numerical and analytical properties of the shape descriptors have been studied until now [50–53]. In this work we begin by studying a key property: *the asymptotic behavior of overcrossing descriptors for long linear polymers.* Scaling laws are known for some geometrical and topological descriptors, such as the radius of gyration [6,10,54,55], the anisotropies [14], and the minimum number of crossings in macromolecular rings [34,35]. The main goal of the present study is to establish whether or not similar scaling behaviors exist when a different viewpoint is adopted to analyze 3D molecular shape. We have thus studied numerically the asymptotic behavior of configurationally averaged 3D shape descriptors for random (linear) polymers with excluded-volume, as well as for experimental α-carbon backbones of proteins.

The article is organized as follows. In Sec. II, we discuss briefly the notion of overcrossing probability distribution, the molecular shape descriptors derived from it, and their numerical computation. Section III deals with the model of a polymer with excluded volume and the determination of the configurational average of shape descriptors. The results for the asymptotic behavior of various descriptors of polymer shape are given in Sec. IV. The analysis of protein backbones is presented in Sec. V, and the scaling relations obtained are contrasted with those of polymers. Section VI contains a summary of conclusions.

II. CHARACTERIZATION OF THE GLOBAL MOLECULAR SHAPE OF LINEAR POLYMERS

Consider a linear polymer (or protein backbone) given as sequence of $n-1$ straight-line segments defined by n identical "nodes" or "main chain atoms" ($n \geq 3$), where each segment has the same length. Since the number of atoms is proportional to the backbone *contour length,* we shall loosely refer to n also as a "chain length."

A polymer configuration is specified by the set of atomic positions ($\mathbf{r}_i$, $i=1,2,\ldots,n$) and the connectivity between the atoms. (For convenience, the geometric center of the polymer chain is chosen as the origin.) One can characterize the configuration by means of molecular geometry descriptors, such as the radius R of the smallest sphere (centered at the geometrical center of the backbone) that encloses completely the polymer chain,

$$R=\max_{\{i\}} r_i \ , \quad r_i=\|\mathbf{r}_i\| \ ; \tag{1a}$$

the instantaneous radius of gyration R_G [10],

$$R_G^2=\frac{1}{n}\sum_{i=1}^{n} r_i^2 \ ; \tag{1b}$$

or other related quantities, such as moments of inertia or asphericity [11,12,14]. These parameters mostly convey features related to the size and anisotropy of a chain. In addition, the radius of gyration and asphericity give some measure of compactness in the spatial distribution of atoms. However, in order to recognize 3D folding features, geometry alone is not enough. In this work, we describe some aspects of the polymer folding at *a given rigid conformation* (the "topological shape," in the sense used in biochemistry) by taking into account the molecular geometry and the connectivity. The essentials of the technique for shape characterization are described elsewhere [48–50]; a brief review is given below to make the work self-contained.

Let K denote a generic configuration of the backbone. No physical property of the backbone changes if we translate or rotate it rigidly, as long as no external fields, solvation effects, or other molecules are included. Our shape descriptors characterize folding features of the isolated backbone by using *all the nuclear configurations that are translationally and rotationally equivalent to K.*

A rigidly rotating polymer can appear in infinitely many *placements.* A *2D projection* can be associated with each placement, as if an imaginary "viewer" would take a snapshot of the configuration along a direction passing through the geometrical center of the backbone. One can use properties of this 2D snapshot to characterize the particular placement. By extension, by using *all possible 2D projections* we can characterize completely the polymer configuration K [49].

A *geometrical criterion* is used to quantify the *topological shape* of the polymer configuration K. Wherever two segments of the backbone "cross over" each other in front of the viewer, the 2D projection will show an actual crossing. We refer to this projected crossing as an *overcrossing*. (In knot theory, sometimes these are referred to as "double points.") Let N be the number of overcrossings at this generic 2D projection. We now introduce the *probability $A_N(n)$ of observing a rigid placement with N overcrossings in an n-atom polymer chain* [26], or simply the "overcrossing probability $A_N(n)$." The set $\{A_N(n)\}$ defines a probability distribution for the discrete variable N. We refer to the histogram representation of $A_N(n)$ vs N as the *overcrossing spectrum* of the backbone [50]. For a linear n-atom polymer (i.e., with no branching or self-intersections), this distribution satisfies

$$\sum_{N=0}^{\max N} A_N(n)=1 \ , \quad \forall n \geq 3 \tag{2}$$

where $A_N(n)=0$ for $N>\max N=(n-2)(n-3)/2$. In practice, $A_N(n)$ also vanishes for some N values below $\max N$. The function $\{A_N(n)\}$ is a *global shape descriptor*, and it conveys some essential aspects of the 3D shape of the polymer. A configuration K which is "open" (not compact) *will mostly appear in placements with low numbers of overcrossings* (i.e., it will have large $A_N(n)$ values for small N). In contrast, compact and entangled configurations *will exhibit a large number of overcrossings* [i.e., the largest $A_N(n)$ values appear for large N]. In brief, the overcrossing probabilities can distinguish between configurations with different folding features.

The overcrossing spectra are computed numerically; an algorithm for determining the overcrossings is provided in Ref. [48], and a strategy to calculate the probabilities is given in Ref. [50]. The distribution $\{A_N(n)\}$ is evaluated as follows: (a) a randomly chosen sample of m_p viewing directions is defined over the smallest sphere enclosing the backbone [Eq. (1a)]; (b) each viewing direction generates one 2D projection for which a number of overcrossings N is computed; (c) the probability $A_N(n)$ is made equal to the number of views with N overcrossings over the total number of random directions generated. The randomization algorithm we use is that of Marsaglia [56]. Accurate results for $\{A_N(n)\}$ are obtained by averaging over several computations with different m_p values.

Figure 1 illustrates these notions with one example. Here, we have considered one generic configuration of a random-walk linear polymer with excluded interaction between beads. The number of beads is $n=170$, and the distance between connected beads is constant. The top diagram displays an arbitrary 3D placement of the chain. The projection exhibits $N=59$ overcrossings. The lower diagram shows the corresponding overcrossing spectrum for this polymer configuration, once a large number of random projections are used. The spectrum superimposes the results obtained with six randomizations, including $m_p=4000$, 6000, 8000, 10 000, 15 000, and 20 000 snapshots, respectively. The oscillations in the probabilities $\{A_N(n)\}$ illustrate the typical accuracy achieved in this work. Our present results for $\{A_N(n)\}$ have an uncertainty smaller than 0.003. This accuracy is enough to carry out the analysis in next sections.

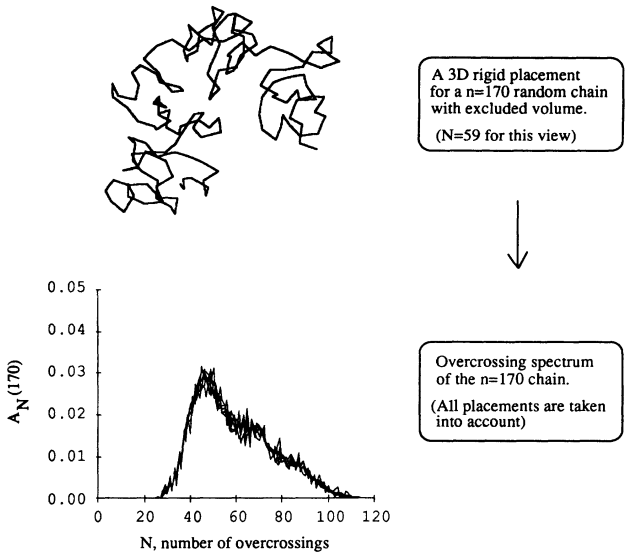

FIG. 1. Illustration of the main concepts used in this work. The top diagram shows the 2D projection of one possible 3D placement of a generic random walk with $n=170$. The projection exhibits $N=59$ overcrossings (points where two segments cross over each other). When the computation of N is repeated for all possible placements, one can construct the probability histogram of observing a given number of overcrossings (or *overcrossing spectrum*) shown in the lower diagram. The results obtained with various numbers of random placements are superimposed in the diagram. The oscillations illustrate the typical accuracy with which the probabilities $\{A_N(n)\}$ are evaluated in this work.

A number of molecular shape descriptors can be derived from the probability distribution $\{A_N(n)\}$. For instance, the most probable number of overcrossings N^* of a polymer configuration and its associated probability A^*:

$$A^*=A_{N^*}(n)=\max_{\{N\}} A_N(n) \ . \tag{3}$$

Similarly, one can analyze the moments of the distribution:

$$\overline{N^p}=\sum_{N=0}^{\max N} N^p A_N(n) \ , \quad p=1,2,\ldots \ . \tag{4}$$

In this work, we will use A^*, N^*, and $\overline{N}$ (the *mean number of overcrossings*). A magnitude similar to $\overline{N}$ has also been proposed recently to study entanglements in self-avoiding random walks in lattices [51]. Note that the descriptors N^* and $\overline{N}$ refer to the actual number of overcrossings in the *rigid* configuration. These descriptors are not related to the "minimum number of crossings" used in the literature to study knotting in polymer rings [34,35]; in this latter case, the descriptor is determined by performing homeomorphic (nonrigid) transformations on the configurations.

We use the results in Fig. 1 to illustrate the accuracy in the shape descriptors. Averaging over the six randomizations, the descriptors for this polymer configuration are $A^* \approx 0.0307 \pm 0.0007$, $N^* \approx 47 \pm 2$, and $\bar{N} \approx 59.3 \pm 0.4$ (with $n = 170$). These results are accurate; the uncertainties are larger for longer polymers.

Despite being computed from the geometrical information (i.e., atomic positions), the probability distribution $A_N(n)$ vs N does not depend *explicitly* on local features, such as the distance between the segments that overcross. These probabilities may not necessarily change much when the molecule undergoes conformational rearrangements. This property makes $\{A_N(n)\}$ a useful tool to monitor the persistence of folding features over time, e.g., along molecular dynamics trajectories [52,53]. Similarly, one can explore the relation between shape descriptors and conformations by searching randomly over the polymer's configurational space. This problem is addressed in the next section where we study the asymptotic behavior of A^* and N^* in terms of the chain length n.

III. MODEL OF A POLYMER WITH EXCLUDED-VOLUME INTERACTION AND COMPUTATIONAL STRATEGY

A simple model molecular chain with excluded-volume interaction is used in our analysis. For these chains, the configurationally averaged shape descriptors are evaluated as a function of an excluded-volume parameter.

The model consists of a rod-and-bead necklace polymer chain [57–59]. These polymers are off-lattice, Pearson random walks with constant-length steps, meeting a criterion of excluded volume about each bead [60]. Attached to each "nucleus," a sphere ("bead") of radius r_{ex} is considered. The configurations permitted are those in which no "nucleus" (a node in the walk) penetrates inside the spheres about other nonconnected nodes. Similar models are commonly used to compute excluded volumes [60], simulate polymer swelling in various solvents [61], study the behavior of rigid and flexible fibers in flow fields [62], and evaluate the topological state of large polymer rings [34,35]. Self-avoiding, off-lattice models describe configurational transitions in polymers and proteins better than the lattice models [8,10,61,63–69]. Nevertheless, both approaches are expected to give the same scaling laws for configurationally averaged molecular properties [6,70].

In order to later compare these polymers with protein backbones, we consider a constant internuclear ("interbead") distance of $l = 3.8$ Å, which is the average distance between α-carbon atoms in proteins [1]. The bond angles are chosen randomly, subject to the constraint set by a given excluded-volume radius r_{ex}. The generation of polymer configurations is straightforward. After two beads have been bonded at distance l, a third bead is attached to the second at a distance l but otherwise an arbitrary location in space. (The random location in space is performed by following the procedure in Ref. [56].) Then, the distance between the third and first bead (not linked) is checked; if this distance is smaller than r_{ex}, the configuration is rejected and the walk recommences from scratch. If the conformation is accepted, a fourth bead is linked to the third with the same criterion as before, and the contacts with the nonbonded beads are checked again to decide whether the partial walk is acceptable or not. The procedure continues until a successful configuration with n beads has been generated.

Note that r_{ex} *can take a maximum value of* $2l$. At this limit value, the chain is forced to be linear. For smaller values of r_{ex}, the chain can adopt infinitely many configurations, even though the range of configurations with distinct folding is limited for large r_{ex}. This range is maximized only in the limit of $r_{ex} \to 0$. The change in the nature of the polymer configurations as a function of r_{ex} can be compared to a "temperature" or "solvent" effect [61,70,71]. In a "poor" solvent, the chain is ideal and adopts *compact* configurations (small r_{ex}), whereas in a "good" solvent, the chain appears in extended configurations characteristic of *swollen* polymers (large r_{ex}) [72]. The swollen polymer is *less flexible* since its configurational possibilities are restricted by the excluded-volume condition. The transition between these two configurational regimes can be described in terms of the single dimensionless parameter $y = r_{ex}/2l$. (The change of configurational regime is not a true phase transition in this model but a crossover phenomenon caused by a growing repulsive interaction in the finite polymer [6,70].)

Each configuration of the present polymer model has the same "energy." Therefore, configurational averages are evaluated in the microcanonical ensemble, where each configuration K_i has the same *a priori* probability. The configurationally averaged largest probability of overcrossing is thus given by

$$\langle A^* \rangle = \lim_{M \to \infty} \frac{1}{M} \sum_{i=1}^{M} A^*(K_i) , \tag{5a}$$

with similar expressions for the other shape descriptors. In addition, configurational fluctuations can be represented by standard deviations,

$$\sigma_{A^*} = \{ \langle A^{*2} \rangle - \langle A^* \rangle^2 \}^{1/2} . \tag{5b}$$

In practice, the number of configurations M is finite. In actual calculations, we have added randomly generated configurations K_i until a desired stability in the values of averages (e.g., $\langle A^* \rangle$) is achieved. The computations are quite demanding, particularly for long polymers. The evaluation of the configurational averages with an accuracy of three significant figures for a 500-bead chain with no excluded volume requires approximately 80 h of CPU on a VAX-4000 computer. The computational demand is heavier for larger r_{ex}, since many randomly generated conformations are rejected for failing the criterion of excluded volume. The present analysis is limited to polymers with $n \leq 500$. On the other hand, computations for proteins (Sec. V) include the largest ones in the Brookhaven Protein Data Bank (PDB) (ca. $n \sim 800$).

IV. RESULTS FOR THE ASYMPTOTIC BEHAVIOR OF POLYMER SHAPE DESCRIPTORS

Configurational averages of shape descriptors have been determined for polymers with the bead numbers (number of "main-chain atoms") $n=4$, 5, 10, 20, 50, 100, 150, 200, 300, 400, and 500. Several values of the radius of excluded volume were studied, ranging from $r_{ex}=0.001$ Å (practically, no excluded volume) to $r_{ex}=7.0$ Å (very large excluded volume). Since the distance between connected "atoms" is constrained to $l=3.8$ Å, the maximum value for r_{ex} is $2l=7.6$ Å. In practice, it is very time consuming to generate a large number of configurations when $r_{ex}>7.0$ Å. Moreover, the closer the parameter $y=r_{ex}/2l$ to unity, the longer the chains needed to reach the asymptotic scaling behavior.

As discussed in Sec. III, random configurations have been added until achieving an accuracy of at least ± 0.002 in $\langle A^* \rangle$, which appear to be the shape descriptor most sensitive to relative error. Since the actual calculation of each $A^*(K_i)$ value has also an error of ca. 0.002, we estimate that the results for $\langle A^* \rangle$ will not be useful for polymers much larger than $n=500$ unless another computational algorithm is adopted. To achieve the present accuracy, some 1000 configurations are needed for small values of r_{ex}; for r_{ex} close to 7.0 Å the number of configuration needed may be as small as 200. Under these conditions, the other averages are computed with three significant figures (e.g., $\langle N^* \rangle$, $\langle \bar{N} \rangle$, and $\langle R_G^2 \rangle^{1/2}$) or 2 (e.g., $\langle R \rangle$ and the standard deviations σ_R, σ_{N^*}, and σ_{A^*}).

The reliability of the configurational averages can be asserted from the molecular radii ($\langle R_G \rangle$ and $\langle R \rangle$) and the fluctuation σ_R, whose asymptotic behaviors with the polymer length are well known. The following power laws are expected as a function of n for both radii and standard deviations [6,10,54,55]:

$$\langle R_G^2 \rangle^{1/2} \approx k' n^{\nu}, \quad \langle R \rangle \approx k n^{\nu}, \tag{6}$$

with $\nu_1 \leq \nu \leq \nu_2$, where the exponents $\nu_1=0.5$ and $\nu_2=0.6$ are associated with the scaling behavior in a "poor" solvent and a "good" solvent, respectively. For large polymers with excluded volume, the asymptotic exponent predicted by field theory is $\nu \rightarrow \nu_3 \approx 0.588$ [54]. Our configurational sampling verifies these results with accuracy. The results for chains in the limit of a "poor" solvent are shown in Fig. 2, corresponding to the value $r_{ex}=0.001$ Å. A linear fitting with $50 \leq n \leq 500$ produces the results

$$\ln\langle R_G^2 \rangle^{1/2} \approx (0.500 \pm 0.001) \ln n + (0.44 \pm 0.05), \quad \mathcal{C}=0.999\,23,$$

$$\ln\langle R \rangle \approx (0.52 \pm 0.01) \ln n + (0.90 \pm 0.05), \quad \mathcal{C}=0.999\,04,$$

$$\ln\sigma_R \approx (0.51 \pm 0.02) \ln n + (-0.5 \pm 0.1), \quad \mathcal{C}=0.9949,$$

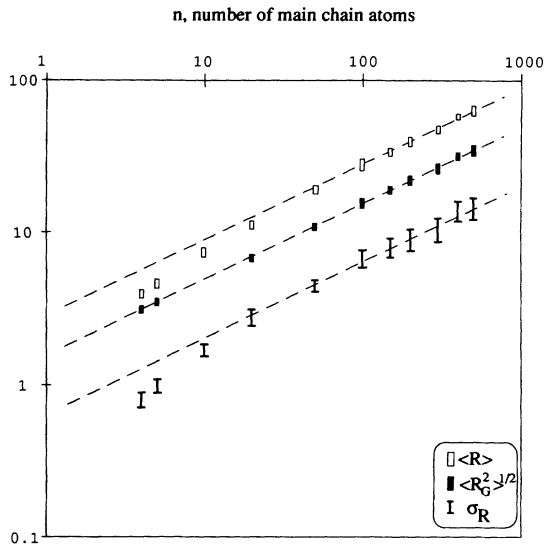

FIG. 2. Log-log plot of the configurational averages of the radius of gyration $\langle R_G^2 \rangle^{1/2}$, the radius of the smallest sphere (centered at the geometrical center) enclosing the polymer $\langle R \rangle$, and the standard deviation (fluctuation) in the latter, σ_R, as a function of the number of main-chain atoms n, for the model polymers with very small excluded volume ($r_{ex}=0.001$ Å). The dashed lines, with slope 0.5, represent the exact "ideal" behavior. The agreement ensures a reasonable configurational sampling. (The different errors bars represent the uncertainties due to variable numbers of configurations M included.)

which agree with the exact exponent $\nu_1=0.5$. (Standard errors are given for slope and intercept; $\mathcal{C}$ stands for correlation coefficient.) The exact behavior (slope $=0.5$) is indicated in Fig. 2 with a dashed line. The results in the regime of a polymer swollen in a "good" solvent are given in Fig. 3, where the value $r_{ex}=7.0$ Å was used. Notice that the asymptotics requires longer chains to set in. A linear fitting with $50 \leq n \leq 500$ produces the results

$$\ln\langle R_G^2 \rangle^{1/2} \approx (0.57 \pm 0.01) \ln n + (0.44 \pm 0.05), \quad \mathcal{C}=0.999\,23,$$

$$\ln\langle R \rangle \approx (0.60 \pm 0.02) \ln n + (1.64 \pm 0.09), \quad \mathcal{C}=0.9981,$$

$$\ln\sigma_R \approx (0.58 \pm 0.03) \ln n + (0.3 \pm 0.2), \quad \mathcal{C}=0.9935,$$

which compare well with the exact result 0.588. (The latter is indicated with a dashed line in Fig. 3.) For intermediate values of r_{ex}, the scaling for the radius of gyration has an effective exponent between ν_1 and ν_2, as it corresponds to the crossover between the regimes dominated by either compact or extended chain configurations.

The results above indicate that the conformational search is thorough enough and the chains long enough to reach the scaling regime. As shown below, the asymptotic behavior for the molecular shape descriptors is determined with an accuracy similar to that of the radius.

Using the same configurational sampling, we have studied the average of the highest probability of over-

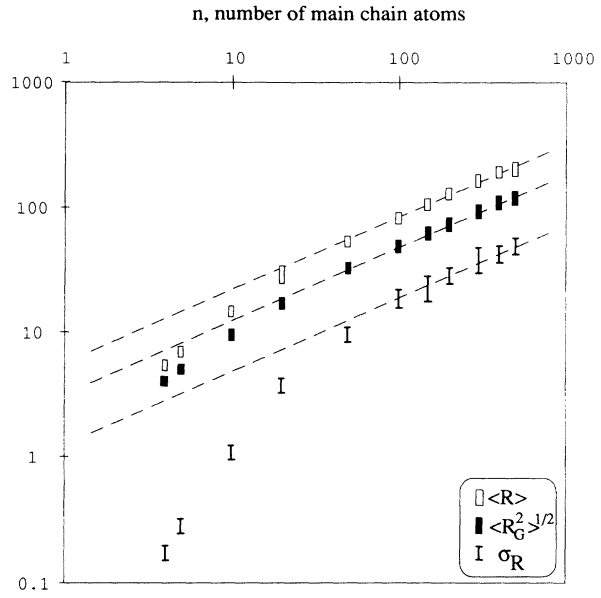

FIG. 3. Log-log plot of the configurational averages of the radius of gyration $\langle R_G^2 \rangle^{1/2}$, the radius of the smallest sphere (centered at the geometrical center) enclosing the polymer $\langle R \rangle$, and the standard deviation (fluctuation) in the latter, σ_R, as a function of the number of main-chain atoms n, for the model polymers with large excluded volume ($r_{ex}=7.0$ Å). The dashed lines, with slope 0.588, represent the expected behavior according to field theory. The agreement ensures a reasonable configurational sampling.

crossings, $\langle A^* \rangle$. Since $\langle A^* \rangle$ *decreases* with the chain length, the relative errors are larger for the long polymers. Therefore, the data have been analyzed allowing for a possible residual error, that is,

$$\langle A^* \rangle \sim a n^b + \text{const} , \tag{7}$$

and the correlation maximized over the exponent b. The results for $n > 50$ are

$$b \approx -0.99 \pm 0.02 , \quad a = 3.5 \pm 0.3 , \quad \mathcal{C} = 0.999\,919$$
$$\text{for } r_{ex} = 0.001 \text{ Å} ,$$
$$b \approx -1.01 \pm 0.02 , \quad a = 25 \pm 1 , \quad \mathcal{C} = 0.999\,959$$
$$\text{for } r_{ex} = 7.0 \text{ Å} ,$$

with a small residual, const~ 0.003. A similar behavior is observed for other excluded-volume values. As well, despite their intrinsic larger error, the results for the fluctuations in long polymers agree with these estimates ($n > 150$):

$$\ln \sigma_{A^*} \approx (-0.98 \pm 0.06) \ln n + (0.0 \pm 0.4) , \quad \mathcal{C} = 0.9958$$
$$\text{for } r_{ex} = 0.001 \text{ Å} ,$$
$$\ln \sigma_{A^*} \approx (-1.10 \pm 0.06) \ln n + (2.1 \pm 0.4) , \quad \mathcal{C} = 0.9953$$
$$\text{for } r_{ex} = 7.0 \text{ Å} .$$

Finally, bringing together the numerical results for all r_{ex} values, an estimate of the "critical" exponent in the scaling of $\langle A^* \rangle$ can be given as

$$b \approx -1.00 \pm 0.03 \quad \text{for all } r_{ex} . \tag{8}$$

Figures 4–6 illustrate this reciprocal relation between the shape descriptor and the chain length. (The error bars indicated in these and the following figures represent the uncertainty in the averaged shape descriptor due to the use of various numbers m_p of randomized projections and configurations M.) Figure 4 shows the correlation in $\langle A^* \rangle$, and its configuration fluctuation σ_{A^*}, as a function of $1/n$ for small excluded volume. Figure 5 displays the same behavior in the limit of very large excluded volume. Finally, Fig. 6 summarizes the results for various r_{ex} values. This latter figure illustrates the apparent universality of the exponent b, whereas the coefficient a shows a strong dependence on the excluded-volume interaction. The results indicate that the information on the type of dominant polymer configurations is retained in a, which appears to depend only on the parameter y, for sufficiently long chains [$a = a(y)$]. Small $a(y)$ values are characteristic of polymers in compact configurations ($y \approx 0$), whereas larger values of $a(y)$ correspond to polymers in swollen configurations ($y \approx 1$).

The occurrence of a similar power-law dependence was studied for the configurationally averaged most problem number of overcrossings: $\langle N^* \rangle \sim \alpha n^\beta$. Figure 7 gives a log-log plot for this shape descriptor as a function of n, in the case of small excluded volume. The results are an example of the behavior encountered for other r_{ex} values. The average $\langle N^* \rangle$ presents a well-defined scaling. The standard deviation σ_{N^*} reaches the asymptotic regime for polymers with $n > 500$.

The numerical fitting indicates also little dependence of the "critical" exponent β over the r_{ex} value. The results for small and large excluded volume are

$$\ln \langle N^* \rangle \approx (1.36 \pm 0.02) \ln n + (-2.39 \pm 0.08) ,$$
$$\mathcal{C} = 0.999\,781 \quad \text{for } r_{ex} = 0.001 \text{ Å} ,$$

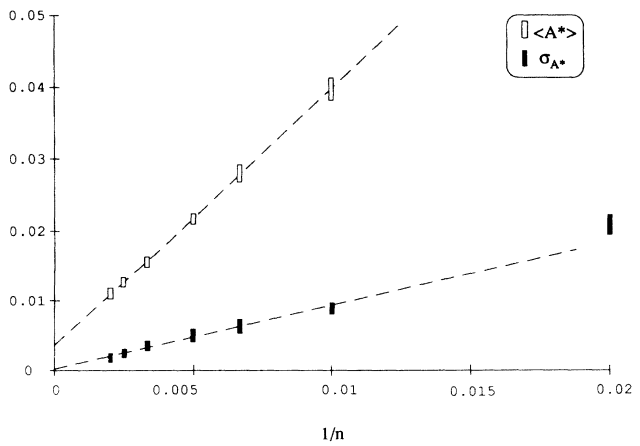

FIG. 4. Configurational averages of the largest probability of overcrossings $\langle A^* \rangle$ and its standard deviation σ_{A^*} as a function of the reciprocal number of monomers $1/n$ for $r_{ex} = 0.001$ Å.

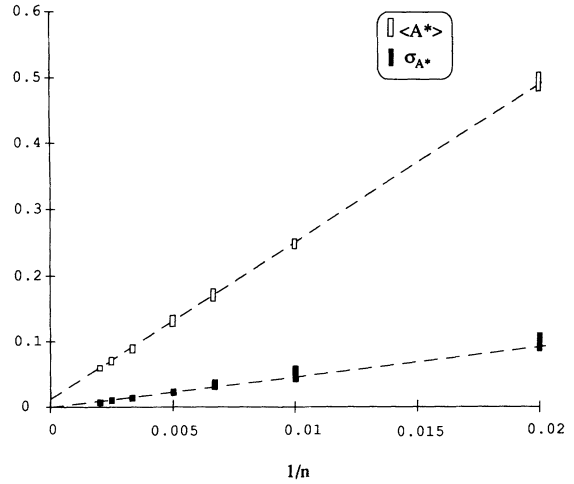

FIG. 5. Configurational averages of the largest probability of overcrossings $\langle A^* \rangle$ and its standard deviation σ_{A^*} as a function of the reciprocal number of monomers $1/n$ for $r_{ex}=7.0$ Å.

$$\ln\langle N^* \rangle \approx (1.4\pm0.1)\ln n + (-5.5\pm0.6) ,$$

$$\mathcal{C}=0.9970 \quad \text{for } r_{ex}=7.0 \text{ Å} ,$$

for $n > 100$. We have also analyzed the data allowing a residual, as done in Eq. (7):

$$\langle N^* \rangle \sim \alpha n^{\beta} + \text{const} , \tag{9}$$

and maximizing the correlation over the exponent β. The results obtained are

$$\beta \approx 1.34\pm0.02 , \quad \alpha=0.10\pm0.01 , \quad \mathcal{C}=0.999\,89$$

$$\text{for } r_{ex}=0.001 \text{ Å} ,$$

$$\beta \approx 1.31\pm0.02 , \quad \alpha=0.007\pm0.001 , \quad \mathcal{C}=0.999\,25$$

$$\text{for } r_{ex}=7.0 \text{ Å} .$$

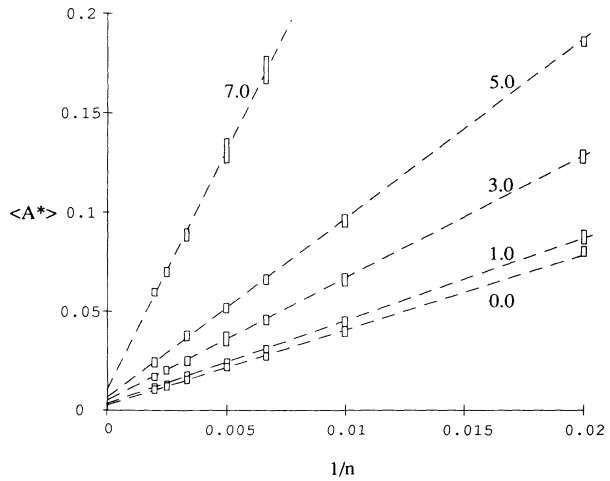

FIG. 6. Configurational averages of the largest probability of overcrossings $\langle A^* \rangle$ as a function of the reciprocal number of monomers $1/n$ or polymers with varying excluded volumes. The r_{ex} values are indicated next to the straight lines. The slopes depend on the excluded-volume interaction.

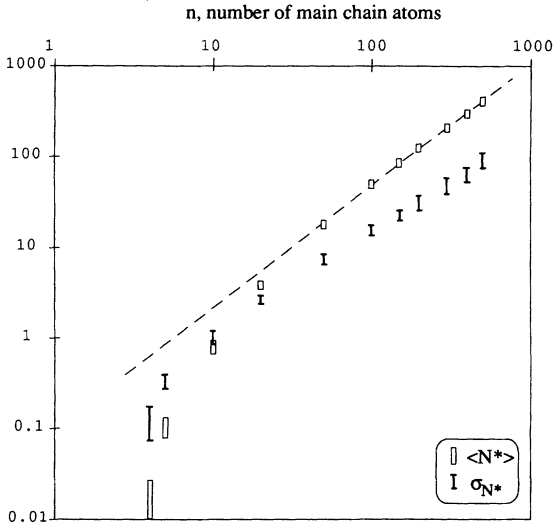

FIG. 7. Configurational averages of the most probable number of overcrossings $\langle N^* \rangle$ and its standard deviation σ_{N^*} as a function of the number of monomers n for $r_{ex}=0.001$ Å (log-log plot). Similar behaviors are found for all other values of r_{ex}.

An "effective" exponent $\beta_{eff} \approx 1.33\pm0.03$ describes (numerically) well the behavior of the shape descriptor $\langle N^* \rangle$ for medium size polymers, and it can be used conveniently to produce approximate linear plots. However, averaging over the various fittings, a more conservative estimate to the actual value of β would appear to be

$$\beta \approx 1.4\pm0.1 \quad \text{for all } r_{ex} . \tag{10}$$

The same scaling behavior is found for the average of the mean number of overcrossings, $\langle \bar{N} \rangle$. (This result improves on a rough estimate of β given in Ref. [53]. In Ref. [51], a study of a function similar to $\langle \bar{N} \rangle$ shows rigorously that β cannot be smaller than 1.)

Within the accuracy of the present calculations, any dependence of this exponent upon the excluded-volume radius r_{ex} cannot be discerned. This is illustrated in Fig. 8, which shows the results for various r_{ex} values in a plot with the "effective" exponent $\beta_{eff}=1.33$. Whereas the same scaling is found in all cases, the coefficient α depends strongly on the excluded volume. As found in the analysis of $\langle A^* \rangle$, the coefficient $\alpha=\alpha(y)$ contains all the information on the dominant polymer configurations. Large $\alpha(y)$ values correspond to compact polymers, whereas small values are characteristic of polymers in mostly swollen configurations [i.e., an opposite behavior to $a(y)$].

The present results are not enough to estimate an analytical form for the functions $a(y)$ and $\alpha(y)$. However, they must satisfy some known limit properties. For very large excluded volume ($y \to 1$), the molecular chains adopt only *perfectly linear* configurations, and therefore show no overcrossings (i.e., $\langle A^* \rangle \to 1$ and $\langle N^* \rangle \to 0$). Consequently, in the asymptotic limit, we find

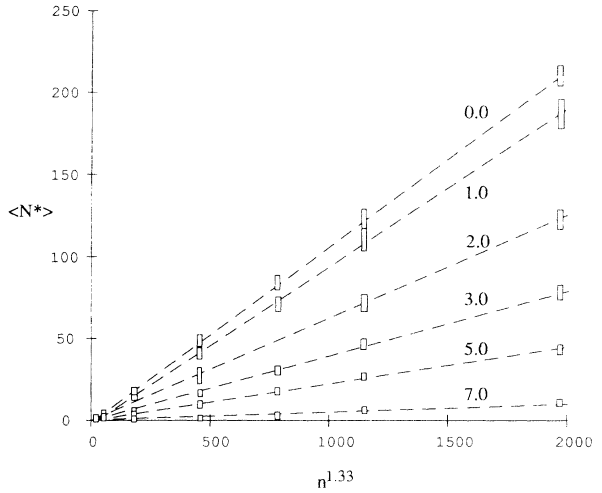

FIG. 8. Linearized plot of the configurational averages of the most probable number of overcrossings $\langle N^* \rangle$ as a function of the number of monomers n for various values of excluded volume. [The effective exponent $\beta_{\text{eff}}=1.33$ is used only for illustration purposes since it provides the optimum regression of the form $\langle N^* \rangle = \alpha n^{\beta} + \text{const}$ for not too large values of n. The actual scaling exponent is probably closer to $\beta \approx 1.4$ (see text). The r_{ex} values are indicated next to the straight lines.]

$$\lim_{n\to\infty} \lim_{y\to 1-} \frac{a(y)}{n} = 1 , \tag{11a}$$

$$\lim_{y\to 1-} \alpha(y) = 0 . \tag{11b}$$

That is, the coefficient $a(y)$ for $\langle A^* \rangle$ must be infinitely large for infinitely long and stretched polymers, whereas the coefficient for $\langle N^* \rangle$ (or $\langle \bar{N} \rangle$) must be zero. The results indicate as well that $a(y)$ is a monotonously increasing function of y, with no inflection points. [A very rough approximation to the present numerical value is $a(y) \sim 4.1/(1-y)$.] In contrast, $\alpha(y)$ appears to be a monotonically decreasing function of y, exhibiting one inflection point. For large excluded volume, $\alpha(y)$ decreases almost linearly with y. [With the data in this work, the inflection point is not far from $y \sim 0.26$ ($r_{\text{ex}} \sim 2.0$ Å).]

In summary, it is found that the molecular shape descriptors of the polymer model can be written asymptotically in terms of two multiplying factors: on the one hand, one function that depends on the excluded volume, that is, on the type of dominant configurations in the chain (swollen or compact); on the other hand, an exponential of the chain length, where the exponent seems independent of the excluded volume. We shall see in the next section that results for proteins suggest also that these exponents depend little on the "swollen" or "collapsed" state of the polymer.

V. SCALING BEHAVIOR OF SHAPE DESCRIPTORS FOR PROTEIN BACKBONES

The asymptotic behavior of the shape descriptors has also been studied for a set of 197 protein backbones. The overcrossing probabilities $\{A_N(n)\}$ are computed using the same approach employed for the polymers. For each protein, only one or a handful of experimental (x-ray, NMR) configurations are available, and therefore we drop the notation of configuration average in the descriptors. For single configurations, the most convenient property is $\bar{N}$, since it is less sensitive than A^* and N^* to computational error due to the randomization of rigid backbone projections.

The protein backbones are given by the sequence of α-carbons associated with the amino acid residues. The backbone structures have been extracted from the Brookhaven PDB [73,74] and selected in an unbiased manner searching for both structural variety and possible chain lengths. To our knowledge, our present survey contains the most varied set of proteins on which an analysis of molecular shape descriptors has been carried out to date. Previous works in the literature (e.g., Refs. [75,76]) have included up to 90 proteins, most of which are globular and contain no more than 300 amino acid residues ($n \leq 300$). In contrast, our set of proteins contains both globular and irregular proteins, together with other exhibiting familiar folding patterns [1,77]. The number of α-carbons for each protein, n, has been taken as the number of backbone atoms in the *actual* structures deposited in the PDB, and *not* as the ideal number of residues in the complete primary sequence. The largest proteins in our set correspond to two chains of glycogen phosphorylase, 1GPA ($n=828$) and 1GPB ($n=823$), which are the longest backbones included in the PDB [78]. The smallest protein in our analysis is deaminooxytocin (1XY1, $n=8$), which has $N^*=0$, as it is the typical case in isolated α helices and β sheets [48,49]. The smallest structure with $N^*>0$ is a chain of insulin (9INS, $n=30$). The shape descriptors have been evaluated upon averaging six different computations of the overcrossing spectrum of each protein, corresponding to 4000, 6000, 8000, 10 000, 15 000, and 20 000 random projections. Different random seeds were used in each case. The results for the shape descriptors are accurate up to three significant figures. The present survey is thorough enough to draw clear conclusions on their scaling behavior.

Dewey [76] found for a set of 43 (mostly globular) proteins that the radius of gyration scaled as $R_G \sim n^{\nu}$, with $\nu \approx 0.35 \pm 0.03$, with "slight deviations" from this behavior. This result was in support of the "collapsed" polymer model, which requires $\nu = \frac{1}{3}$ below the theta point [6,79]. We have reanalyzed this behavior using the present set of proteins, and the results are shown in Fig. 9. The dispersion from a linear behavior is large and the values for the exponent ν result from a poor correlation,

$$\nu \approx 0.38 \pm 0.05 , \quad \mathcal{C} = 0.819$$

$$\text{for } n > 55 \quad (166 \text{ proteins}) ,$$

$$\nu \approx 0.37 \pm 0.03 , \quad \mathcal{C} = 0.921$$

$$\text{for } n > 150 \quad (107 \text{ proteins}) ,$$

where the statistical 95% confidence intervals are given. This outcome adds a note of caution with respect to

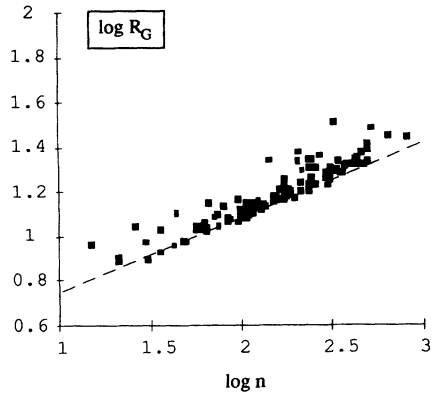

FIG. 9. Log_{10}-log_{10} plot of the radius of gyration of a set of 197 proteins as a function of the number of amino acid residues n. The set has been selected without structural bias; the proteins included represent a random sampling of all the possible folding features know to date. (The dashed line has slope $\frac{1}{3}$, and it represents the behavior expected in the "collapsed polymer" model. Clearly, the power law $n^{1/3}$ is satisfied only by the smallest proteins associated with a given n value.)

drawing universal conclusions about the state, collapsed or extended, of proteins as polymers. Our results show, though, that the "collapsed" state probably characterizes the proteins exhibiting the smallest size compatible with a given number of residues n. As Fig. 9 shows, the "lower bounds" for R_G follow a $\frac{1}{3}$-power law very closely (dashed line), that is,

$$\min(R_G)_n \sim n^{1/3} , \tag{12}$$

where $(R_G)_n$ indicate a value at constant n. For proteins with a larger radius of gyration (for a given n), no conclusion can be deduced since the dispersion is quite large.

In contrast, the shape descriptors N^* and $\bar{N}$ show smaller deviations, and better defined asymptotic behaviors can be deduced within the same set of proteins. Figure 10 displays the results for N^* and $\bar{N}$ vs n, in a log-log plot. (The short proteins with $N^*=0$ are of course excluded from the diagram on the left of Fig. 10.) Both descriptors exhibit the *same* behavior, although $\bar{N}$ has a smaller deviation. A linear fit gives the results

$$\ln\bar{N} \approx (1.37\pm0.05)\ln n + (-2.9\pm0.2) ,$$
$$\mathcal{C}=0.981 , \quad n>55 ,$$
$$\ln\bar{N} \approx (1.42\pm0.03)\ln n + (-3.2\pm0.1) ,$$
$$\mathcal{C}=0.9933 , \quad n>150 ,$$

with 95% confidence levels. The shape descriptor N^* produces a comparable critical exponent $\beta \approx 1.41\pm0.08$ ($n>55$). From the results above, an empirical law for the global shape of all proteins is derived,

$$\bar{N} \approx 0.045 n^{1.4} , \tag{13}$$

where each significant figure is deemed accurate. The same behavior holds approximately for N^*. Equation (13) is valid even for small proteins; note that it gives $\bar{N}<1$ for $n<9$, which is a reasonable approximation of the result for real peptide chains [50].

The numerical fittings above give a conservative estimate $\beta \approx 1.4\pm0.1$ for proteins, in agreement with the exponent found in the case of linear polymers with various r_{ex} values [cf. Eq. (10)]. Our results for the radius of gyration (Fig. 9) indicate that the considered protein backbones appear in many configurational states and not only as "collapsed" structures. Consequently, the fact that a clear scaling behavior is found for N^* and $\bar{N}$, whereas this is not the case for R_G, suggests that the exponent β depends little on the swollen or collapsed nature of the protein configuration. That is, all the results for shape descriptors of proteins appear to be consistent with the observations for linear polymers.

The descriptor A^* shows a larger dispersion, but its scaling behavior is still better defined than that of R_G. A linear fitting with all proteins gives

$$\ln A^* \approx (-0.98\pm0.04)\ln n + (1.6\pm0.3), \quad \mathcal{C}=0.961 ,$$

with 95% confidence levels, which corresponds to the approximate behavior $A^* \approx 5n^{-1}$. As found for the descriptor $\bar{N}$, the asymptotic form of shape descriptor A^* and the exponent b obtained for proteins cannot be distinguished from the results in Sec. IV for the polymer model [cf. Eq. (8)].

Figure 11 makes explicit the comparison between polymers and proteins by superimposing the results for N^* in proteins with the linearized diagram for the same descriptor in polymers (cf. Fig. 8 using the exponent $\beta_{eff}=1.33$). Note that the polymer model was given a distance between beads of $l=3.8$ Å and therefore it can properly be contrasted with protein backbones. The figure illustrates that protein backbones follow the same behavior as the polymers, and that their shape descriptors are neatly bound between the values corresponding to $r_{ex}=1.0$ and 3.0 Å. For proteins, the numerical results give $\alpha \approx 0.045$ and $a \approx 5$, which are close to the values corresponding to $r_{ex} \sim 2.0$ Å in the polymer model. As discussed before, this r_{ex} value corresponded to random-walk polymer configurations that were in the

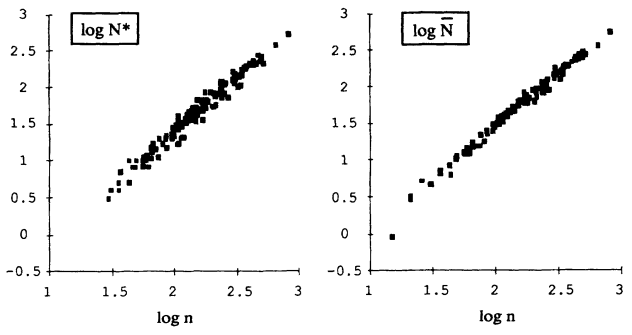

FIG. 10. Log_{10}-log_{10} plot of the shape descriptors N^* (most probable number of overcrossings) and $\bar{N}$ (mean number of overcrossings) as a function of the number of amino acid residues (n) in a survey of 197 protein backbones. The molecular backbones have been extracted from the Brookhaven PDB, and cover proteins with very different folding features. The two descriptors exhibit the same power-law behavior in terms of n.

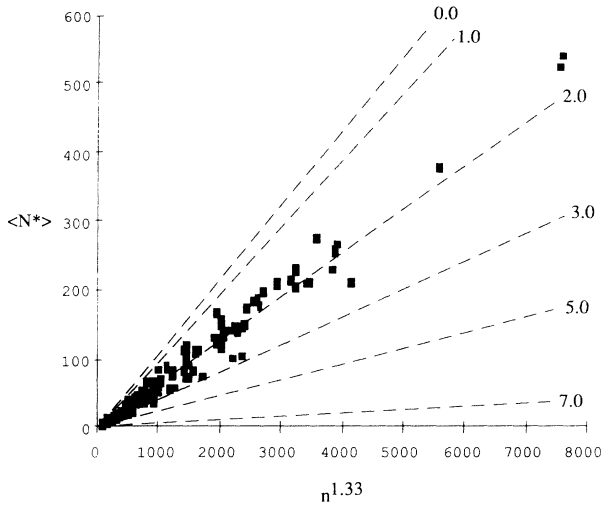

FIG. 11. Superimposed results for the shape descriptor N^* as a function of chain length for protein backbones and polymers with excluded volume (with the same distance between connected atoms). The straight lines represent the results for the configurationally averaged descriptor $\langle N^* \rangle$ of linear polymers (cf. Fig. 8). The values next to the straight lines indicate the radius r_{ex} for the excluded volume. Each square represents the N^* value for one protein backbone. The effective exponent $\beta_{eff}=1.33$ is used only for illustration purposes, since it provides an approximately linearized plot for all the polymers.

transition regime between compact and swollen [i.e., near the infection point in the function $\alpha(y)$]. This coincidence suggests that the set of proteins is found, on the average, on a similar "intermediate" regime. This is again consistent with the fact, illustrated in Fig. 8, that the radius of gyration shows a scattering of values *away* from the regime corresponding to only "collapsed" backbones.

Two additional comments can be added. First, it is observed in Fig. 11 that the shape descriptors for proteins come close to those of "swollen" polymers (e.g., those with $r_{ex} > 3.0$ Å) *only* for quite small values of n. This result is consistent with the fact that very open conformations are found only in rather small objects, such as α helices, pure β sheets, and irregular proteins. Finally, it can be noted that the radius r_{ex} that would fit the average protein behavior more closely is not far from the accepted van der Waals radius of aliphatic carbon (ca. 1.75 Å [80]). It is unclear at this moment whether or not this latter coincidence is of any significance.

VI. CONCLUSIONS

In this work, we have studied the 3D molecular shape description of a simple random-walk model of polymer chains specified by two parameters, namely, the chain length n and the dimensionless excluded-volume radius $y = r_{ex}/2l$. The results obtained suggest that these two parameters appear separated in the analytical (asymptotic) form of the configurationally averaged shape descriptors, i.e.,

$$\langle D \rangle \sim d(y) n^g , \quad n \gg 1 , \tag{14}$$

where d is a function that depends on the shape descriptor D (e.g., $D \equiv A^*$, N^*, or $\bar{N}$) and, through y, on the dominant configurational state of the polymer. In contrast, the exponent g appears to depend little (if at all) on y. With respect to this latter exponent, Eqs. (8), (10), and (13) represent the main numerical results, where $g \equiv b \approx -1.00$ for $D \equiv A^*$, and $g \equiv \beta \approx 1.4$ for $D \equiv N^*$ and $\bar{N}$ are obtained. The uncertainty in these exponents gives an upper bound to their possible dependence on the excluded volume: it should be below 5%, on average. This can be contrasted with the scaling behavior of the radius of gyration (or $\langle R \rangle$), which exhibits a ca. 20% variation in the critical exponents as a function of y. Regarding the function $d(y)$, its qualitative dependence with the shape of the dominant configurations [Eqs. (11)] is probably universal. That is, for a given polymer length, a should decrease with the complexity of the backbone's entanglements, whereas α should increase. The actual form of the functions will, though, depend on the details of the potential-energy function describing the molecular chain.

The asymptotic behavior of the same shape descriptors has also been surveyed in a large set of protein backbones, exhibiting very diverse folding features. The results obtained indicate the following interrelated facts: (1) A single scaling behavior is found for $\bar{N}$, N^*, and A^* throughout the whole set of proteins, whereas R_G does not follow a clear single law; (2) The exponents characterizing the scaling behavior of shape descriptors appear not to depend on the swollen or collapsed nature of the protein configuration; (3) The shape descriptors scale with the number of amino acid residues (n) in the same manner as found for the polymers.

We believe that the finding of well-defined correlations between global descriptors of backbone entanglements (such as A^*, N^*, and $\bar{N}$) and the protein's size is a valuable piece of information. From a conceptual point of view, the analysis provides an insight into describing the 3D structure of protein folds. Such viewpoints are needed as a larger number of structures are deposited in data banks [81]. The shape descriptors convey information on aspects of the "topology" of the fold not readily available when using geometrical descriptors such as R_G, and therefore can serve as a complementary tool. From the practical side, relations such as Eq. (13) (and a similar one for the standard deviation in the descriptor) can be used to assess the likelihood of some conformations in three-space for a given sequence of amino acid residues. This is a central issue in molecular modeling and protein engineering, where one tries to rationally determine the 3D structure of a protein from its primary sequence (e.g., Refs. [2,67,82,83]). With the present criterion, a "trial" folding pattern with an $\bar{N}$ value very far from $0.045n^{1.4}$ could, in the first instance, be discarded. In order to explore these possibilities, a more detailed analysis of *all* proteins available in current data banks (close to 500 in full release) is presently under way.

ACKNOWLEDGMENTS

I would like to thank Mr. Mark Payette for his collaboration in the numerical evaluation of shape descriptors for proteins. This work has been supported by the Fonds de Démarrage (Laurentian University) and an operating grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

[1] C. Brändén and J. Tooze, *Introduction to Protein Structure* (Garland, New York, 1991).
[2] *Proteins: Form and Function,* edited by R. A. Bradshaw and M. Purton (Elsevier, Cambridge, 1990).
[3] R. Jaenicke, Prog. Biophys. Mol. Biol. **49**, 117 (1987).
[4] S. F. Edwards, Proc. R. Soc. London, Ser. A **385**, 267 (1982).
[5] F. W. Wiegel in *Phase Transitions and Critical Phenomena,* edited by C. Domb and J. L. Lebowitz (Academic, London, 1983), Vol. 7, pp. 101–149.
[6] P. G. de Gennes, *Scaling Concepts in Polymer Physics* (Cornell University Press, Ithaca, 1985).
[7] D. W. Sumners and S. G. Whittington, J. Phys. A **21**, 1689 (1988).
[8] M. V. Volkenstein, *Configurational Statistics of Polymeric Chains* (Interscience, New York, 1963).
[9] T. M. Birshtein and O. B. Ptitsyn, *Conformations of Macromolecules* (Interscience, New York, 1966).
[10] P. J. Flory, *Statistical Mechanics of Chain Molecules* (Interscience, New York, 1969).
[11] M. Bishop and C. J. Saltiel, J. Chem. Phys. **88**, 3976 (1988).
[12] H. W. Diehl and E. Eisenriegler, J. Phys. A **22**, L87 (1989).
[13] O. Nilsson and O. Tapia, J. Mol. Struct. **256**, 295 (1992).
[14] A. Baumgärtner, J. Chem. Phys. **98**, 7496 (1993).
[15] V. Daggett, P. A. Kollman, and I. D. Kuntz, Biopolymers **31**, 1115 (1991).
[16] K. B. Lipkowitz and M. A. Peterson, J. Comput. Chem. **14**, 121 (1993).
[17] J. D. Honeycutt and D. Thirumalai, Biopolymers **32**, 695 (1992).
[18] D. Rojewska and R. Elber, Proteins Struct. Func. Genetics **7**, 265 (1990).
[19] K. Kuczera, J. Kuriyan, and M. Karplus, J. Mol. Biol. **213**, 351 (1990).
[20] D. S. Hartsough and K. M. Merz, J. Am. Chem. Soc. **114**, 10 113 (1992).
[21] N. S. Goel and R. L. Thompson, *Computer Simulations of Self-Organization in Biological Systems* (Macmillan, New York, 1988).
[22] Y. Isogai, G. Némethy, S. Rackovsky, S. L. Leach, and H. A. Scheraga, Biopolymers **19**, 1183 (1980).
[23] V. N. Maiorov and G. M. Crippen, J. Mol. Biol. **227**, 876 (1992).
[24] J. Åqvist and O. Tapia, J. Mol. Graph. **5**, 30 (1987).
[25] D. Zachmann, W. Heiden, M. Schlenkrich, and J. Brickmann, J. Comput. Chem. **13**, 76 (1992).
[26] B. K. Shoichet and I. D. Kuntz, J. Mol. Biol. **221**, 327 (1991).
[27] F. Jiang and S.-H. Kim, J. Mol. Biol. **219**, 79 (1991).
[28] M. L. Connolly, Biopolymers **32**, 1215 (1992).
[29] M. Delbrück, Proc. Symp. Appl. Math. **14**, 55 (1962).
[30] G. M. Crippen, J. Theor. Biol. **51**, 495 (1975).
[31] E. J. Janse van Rensburg and S. G. Whittington, J. Phys. A **23**, 3573 (1990).
[32] S. G. Whittington, Proc. Symp. Appl. Math. **45**, 73 (1992).
[33] C. E. Soteros, D. W. Sumners, and S. G. Whittington, Math. Proc. Cambridge Philos. Soc. **111**, 75 (1992).
[34] K. Koniaris and M. Muthukumar, Phys. Rev. Lett. **66**, 2211 (1991).
[35] K. Koniaris and M. Muthukumar, J. Chem. Phys. **95**, 2873 (1991).
[36] J. des Cloizeaux and M. L. Mehta, J. Phys. A **40**, 665 (1979).
[37] J. P. J. Michels and F. W. Wiegel, Phys. Lett. A **90**, 391 (1982).
[38] F. B. Fuller, Proc. Symp. Appl. Math. **14**, 64 (1962); Proc. Natl. Acad. Sci. USA **68**, 815 (1971).
[39] M.-H. Hao and W. K. Olson, Biopolymers **28**, 873 (1989).
[40] A. H. Louie and R. L. Somorjai, J. Theor. Biol. **98**, 189 (1983).
[41] M. Le Bret, Biopolymers **18**, 1709 (1979).
[42] P. De Santis, S. Morosetti, and A. Palleschi, Biopolymers **22**, 37 (1983).
[43] J. H. White, N. R. Cozzarelli, and W. R. Bauer, Science **241**, 323 (1988).
[44] M. L. Connolly, I. D. Kuntz, and G. M. Crippen, Biopolymers **19**, 1167 (1980).
[45] T. Kikuchi, G. Némethy, and H. A. Scheraga, J. Comput. Chem. **7**, 67 (1986).
[46] D. W. Sumners, J. Math. Chem. **1**, 1 (1987).
[47] G. A. Arteca and P. G. Mezey, J. Mol. Graph. **8**, 66 (1990).
[48] G. A. Arteca and P. G. Mezey, Biopolymers **32**, 1609 (1992).
[49] G. A. Arteca, J. Math. Chem. **12**, 37 (1993).
[50] G. A. Arteca, Biopolymers **33**, 1829 (1993).
[51] E. J. Janse van Rensburg, D. W. Sumners, E. Wasserman, and S. G. Whittington, J. Phys. A **25**, 6557 (1992).
[52] G. A. Arteca, J. Comput. Chem. **14**, 718 (1993).
[53] G. A. Arteca, in *Advances in Computational Biology,* edited by H. O. Villar (JAI, Greenwich, CT, 1994), Vol. 1, p. 1.
[54] J. C. Le Guillou and J. Zinn-Justin, Phys. Rev. B **21**, 3976 (1980).
[55] M. Muthukumar and B. G. Nickel, J. Chem. Phys. **86**, 460 (1987).
[56] G. Marsaglia, Ann. Math. Stat. **43**, 645 (1972).
[57] D. C. Rapaport, J. Phys. A **11**, L213 (1978).
[58] D. C. Rapaport, J. Chem. Phys. **71**, 3299 (1979).
[59] A. Bellemans, J. Orban, and D. van Belle, Mol. Phys. **39**, 781 (1980).
[60] A. J. Barrett and B. C. Benesch, J. Chem. Phys. **97**, 9454 (1992).
[61] L. A. Johnson, A. Monge, and R. A. Friesner, J. Chem. Phys. **97**, 9355 (1992).
[62] S. Yamamoto and T. Matsuoka, J. Chem. Phys. **98**, 644 (1993).
[63] H. S. Chan and K. A. Dill, J. Chem. Phys. **90**, 492 (1989).
[64] H. S. Chan and K. A. Dill, J. Chem. Phys. **92**, 3118 (1990).

[65] A. Kolinski, M. Milik, and J. Skolnick, J. Chem. Phys. **94**, 3978 (1991).
[66] E. M. O'Toole and A. Z. Panagiotopoulos, J. Chem. Phys. **97**, 8644 (1992).
[67] J. Skolnick and A. Kolinski, J. Mol. Biol. **221**, 499 (1991).
[68] H. S. Chan and K. A. Dill, Macromolecules **22**, 4559 (1989).
[69] L. M. Gregoret and F. E. Cohen, J. Mol. Biol. **219**, 109 (1991).
[70] K. F. Freed, *Renormalization Group Theory of Macromolecules* (Wiley, New York, 1987).
[71] H. Yamakawa, *Modern Theory of Polymer Solutions* (Harper and Row, New York, 1971).
[72] K. A. Dill and D. Shortle, Annu. Rev. Biochem. **60**, 795 (1991).
[73] F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, D. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, J. Mol. Biol. **112**, 535 (1977).
[74] S. Pascarella and P. Argos, Protein Eng. **5**, 121 (1992).
[75] C. Abad-Zapatero and C. T. Lin, Biopolymers **29**, 1745 (1990).
[76] T. G. Dewey, J. Chem. Phys. **98**, 2250 (1993).
[77] J. S. Richardson, Methods Enzymol. **115**, 359 (1985).
[78] L. L. Walsh (private communication).
[79] I. M. Lifshitz, A. Y. Grosberg, and A. R. Khokhlov, Rev. Mod. Phys. **50**, 683 (1978).
[80] A. Gavezzotti, J. Am. Chem. Soc. **105**, 5220 (1983).
[81] J. M. Thornton, Curr. Opinion Struct. Biol. **2**, 888 (1992).
[82] H. S. Chan and K. A. Dill, Phys. Today **46** (2), 24 (1993).
[83] B. Rost, R. Schneider, and C. Sander, Trends Biochem. Sci. **18**, 120 (1993).