Macromolecules

ARTICLE

pubs.acs.org/Macromolecules

Improved Measures for the Shape of a Disordered Polymer To Test a Mean-Field Theory of Collapse

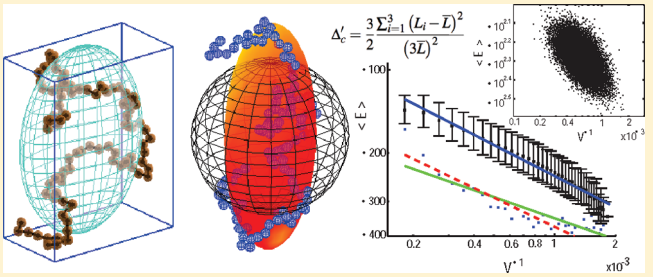
Shirin Hadizadeh,[†] Apichart Linhananta,[‡] and Steven S. Plotkin*[,†]

[†]Department of Physics and Astronomy, University of British Columbia, Vancouver, Canada

[‡]Physics Department, Lakehead University, Thunder Bay, Ontario, Canada

Supporting Information

ABSTRACT: We have generated off-lattice self-avoiding random walks (SAWs) by both Lal's pivot algorithm and by discontinuous molecular dynamics (DMD) simulations. We chose several volume measures to analyze and compare the shapes characterizing ensembles of SAWs. These included the Flory volume defined through the end-to-end distance, the volume corresponding to the radius of gyration of the polymer, the volume of a sphere drawn from the center of polymer mass that encloses or embeds all the monomers in the polymer, the corresponding ellipsoid describing the radius of gyration tensor, the volume of the smallest Cartesian box oriented in a fixed "lab" frame which encloses all polymer monomers, and the volume of the smallest box oriented along the principal axes of the radius of gyration tensor that encloses all polymer monomers. The tensor ellipsoid and principal box correlate well with each other but not with the other measures. There is a substantial amount of polymer that is excluded from the radius of gyration sphere or ellipsoid (approximately 42% and 44% respectively on average), which casts doubt on the utility of these measures in quantitatively characterizing the volume spanned by a polymer configuration. The principal box volume led to the most well-defined length distribution for the SAW in terms of the ratio of standard deviation to the mean, while the end-to-end distribution was the most broadly distributed. In the principal box analysis, polymer configurations are highly anisotropic, with stronger cubic to square prism symmetry breaking than square prism to rectangular cuboid. We introduce the *acubicity* in analogy with the asphericity and find its distribution, along with a generalization of acubicity that better discriminates polymer anisotropy. We analyze the role of the above volume measures in determining the packing fraction which enters into the "random mixing" mean-field theory for the energy of an isolated homopolymer. Here we find from both pivot and DMD simulations that no particular volume measure reproduces the mean-field scaling exponent (of unity) for the energy as a function of the reciprocal polymer volume. Instead, anomalous exponents are observed which are less than that of mean-field theory, ranging from 0.2 to 0.5 depending on the contact cutoff length. An energy function with such an anomalous scaling may be used in a simple phenomenological theory of coil−globule collapse. We find in particular that the Flory volume does not accurately describe the energy in a mean-field theory. Possible reasons for such anomalous exponents are discussed.



INTRODUCTION

Changes in volume play a significant but often under-emphasized role in governing the stability of proteins in the cell.[1−6] When proteins fold from extended unfolded state to a compact or folded native state, there is a concomitant decrease in the characteristic volume occupied by the protein as probed by osmolyte particles, and thus an increase in the volume available to osmolytes or crowding particles. In such processes that involve changes in polymer density concomitant with collapse, the effective size of the polymer is of interest. Accurate measures of the size of a self-avoiding walk, which can approximate configurations in the unfolded ensemble of a protein, have a long history.[7−9] Early on, it was recognized by Kuhn using probability arguments for ideal chains,[10] that ideal chains had inherent anisotropy. The configurations of polymer chains are better represented by generalizing quantities such as the end to end distance or radius of gyration, which are amenable to analytic formulation[11−13] to tensor quantities such as the radius of gyration tensor $\mathbb{S}$ (defined in 5 below), first studied by Šolc and Stockmayer[14,15] for on-lattice ideal chains.

Invariant polynomials constructed from the components of the radius of gyration tensor $\mathbb{S}$ allowed the anisotropy of polymer configurations to be treated analytically by field theoretical methods such as $\varepsilon = 4 - d$ expansions[16] or $1/d$ expansions.[17,18] One such anisotropy parameter is the *asphericity* Δ $(0 \le \Delta \le 1)$, defined through ratios of the trace (tr) of the radius of gyration tensor:

$$\Delta = \frac{3}{2}\frac{\mathrm{tr}\,\hat{\mathbb{S}}^2}{(\mathrm{tr}\,\mathbb{S})^2} = 1 - 3\frac{\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3}{(\lambda_1+\lambda_2+\lambda_3)^2} \qquad (1)$$

Received: February 25, 2011
Revised: May 26, 2011
Published: July 11, 2011

where $\hat{\mathbb{S}}$ is the traceless tensor $\mathbb{S}-(1/3)1\mathrm{tr}\mathbb{S}$, and the λ_i are the eigenvalues of $\mathbb{S}$. One general difficulty with theoretical approaches is that the thermodynamically averaged asphericity involves the average of a quotient of fields and so has proved too difficult to readily calculate, though Diehl and Eisenriegler have proposed a solution to this problem.[19]

Numerical analyses have also added significant insight into the relaxation of polymer anisotropy,[20] the most distant monomers along various axes of a polymer configuration,[21,22] the thermodynamically averaged density distribution function of a polymer,[23] the distribution functions of invariants such as the asphericity above,[24] the asymmetry and collapse of polymers in porous media,[25] and the scaling and shape anisotropy of knotted and ring polymers.[26−30] The distribution of the scalar radius of gyration for an unfolded protein ensemble,[31,32] has been generalized to investigate the asphericity of such an ensemble by Pappu and co-workers,[33,34] and the analysis of asphericity, probability density, as well as intra- and intermolecular entanglements has been performed for polymers in the maximally random jammed state.[35,36]

In the presence of osmolytes or steric crowders, the process of collapse of a homopolymer, or the collapse and/or folding of a protein, can quantified through the increase in polymer density, or alternatively the packing fraction. The packing fraction is defined as the total volume of monomers constituting the polymer as probed by a zero-radius particle, divided by the "total volume of the system". Defining the total volume of the system requires some care however. For example, a box that contains all of the monomers but is either too large or improperly aligned would underestimate the packing fraction, whereas a sphere with diameter equal to the end-to-end distance of the polymer would for many conformations be too small (or may even have volume near zero), resulting in severe overestimation of the polymer density for those conformations. The same concern may be raised for the radius of gyration R_{gy}, which also does not always accurately represent the volume occupied by a polymer or protein. For example, an expanded rod-like conformation of a N residue polymer has $R_{\mathrm{gy}} = \ell(N-1)/(12)^{1/2} \approx 0.3\ell N$ as the linear size of the chain instead of the contour length $L = \ell(N-1)$. In this case, the end to end distance R_{ete} is a more accurate measure of the linear size of the extended polymer, however in this example the volume of the chain would then be poorly described by ${R_{ete}}^3 = L^3$ because of the anisotropy of the configuration. Moreover as mentioned above, for many conformations, R_{ete} underestimates the volume occupied by the polymer; for example there are numerous chain conformations with $R_{ete} \approx 0$ (though this becomes increasingly improbable in the thermodynamic limit), whereas even in a collapsed state a polymer cannot have size smaller than $R_{gy} \propto N^{1/3}$.

Accurate measures of polymer size are important because the effective volume spanned by a protein changes dramatically during folding or collapse. Such improved measures of volume facilitate an accurate statistical mechanical description of polymer−osmolyte or protein−osmolyte mixtures for example. The statistical likelihood of close approaches between parts of the polymer nonlocal in sequence depends on the effective polymer density or packing fraction, so a mean-field theory of the energy in the polymer is affected by the volume-measure used. In this paper we propose various measures to accurately represent the volume enclosing or, more accurately, characterizing a disordered polymer, and we apply these volume measures to modified mean-field theories of homopolymer energy.

In the next section, we investigate the scaling exponents for a self-avoiding off-lattice random walk, and show that a set of des Cloizeaux exponents different than those valid for the $N \to \infty$ limit more accurately describe the end-to-end distribution for chain lengths consistent with small globular proteins. After confirming the pivot-algorithm generated ensembles with discrete molecular dynamics generated ensembles, we introduce several candidate measures of volume characterizing the instantaneous conformations of a polymer. We transform the radius of gyration tensor to an effective ellipsoidal volume that reduces to the scalar radius of gyration volume in the isotropic limit. We then show that the volumes defined by the radius of gyration and radius of gyration ellipsoid generally exclude significant amounts of the polymer. We show that the various volume measures (except for the radius of gyration tensor ellipsoid) poorly correlate with a measure we analyze here which we call the principal box volume [Only in the late stages of the preparation of this manuscript had we discovered that our principal box volume, (re)introduced here as an accurate measure of the effective volume of a polymer, was in fact previously introduced long ago by Rubin and Mazur[22] in the context of "spans" of on-lattice ideal and self-avoiding walks. Similar box-volume measures (but distinct from the principal box) have also been introduced by Rawdon, Millett, Stasiak, and colleagues.[37,29]] and for protein-like sizes have substantial width compared to their mean. We then studied anisotropic symmetry breaking for the principal box distribution of polymer conformations, from both cubic to square prism and square prism to rectangular cuboid symmetry. We show that the distribution of principal box volumes itself obeys a semiempirical des Cloizeaux functional form, but with modified exponents. Corrections due to nonzero collapsed principal box volume are negligible. Lastly, we investigate the thermodynamics of a self-attractive homopolymer using various volume measures to define the packing fraction. We find that the scaling of thermal energy with volume does not obey mean-field behavior for any volume measure used. Using the nonmean-field scaling of energy with volume, along with the des Cloizeaux distribution for the principal volume as an entropy measure, we construct a free energy function that allows an analytical investigation of polymer collapse.

MEASURES AND STATISTICS OF POLYMER SIZE

Self-Avoiding Random Walk. The end-to-end distance R_{ete} or radius of gyration R_{gy} of a self-avoiding polymer chain scales with the length N of the chain as $\sim N^{\nu}$,[7,8] where ν depends of the dimensionality of the system and solvent quality. For a polymer in a good solvent in three dimensions, Flory mean-field theory gives $\nu \approx 0.6$, and for a polymer in bad solvent, $\nu \approx {}^1/_3$. Renormalization calculations[12] along with large scale simulations[38,39] have refined the good solvent exponent to approximately 0.5874.

The end to end distance probability distribution, $P(R_{ete})$, of a freely jointed ideal chain converges to a Gaussian even for chains as small as 10 residues.[7,40] However for polymer chains undergoing random walks in dimensions less than 4 the end-to-end distance probability distribution of a chain is no longer Gaussian when excluded volume effects are present.[41−44] The statistics of a polymer is then often approximated by that of a self-avoiding random walk (SAW) on a hypercubic lattice. The functional form of $P(R_{ete})$ for an excluded volume chain has been investigated using both Monte Carlo simulations[45−47] and analytical

Table 1. Values of Parameter A, and Exponents θ and δ, for (a) End to End Probability Distribution of an Ideal (Ghost Chain) Random Walk, (b) End to End Probability Distribution of an On-Lattice SAW,[47] (c) End to End Probability Distribution of an Off-Lattice, Finite N ($50 < N < 200$) SAW Using the Pivot Algorithm, (d) End to End Probability Distribution of an Off-Lattice SAW Using the Naive (Reflecting) Growth Algorithm, (e) Volume Probability Distribution of the Polymer from eq 10, Using Pivot Algorithm Simulations, (f) Volume Probability Distribution in the Absence of the Collapsed Globule Correction, by Using eq 10) without the Factor of V_o to Best Fit the Simulation Data

	(a) ideal RW (R_{ete})	(b) on-lattice SAW (R_{ete})	(c) off-lattice finite N SAW (R_{ete})	(d) reflecting off-latt SAW (R_{ete})	(e) principal box volume (V_{pb})	(f) principal box volume (V_{pb})
A	1.5	0.144	0.057 ± 0.002	0.46 ± 0.01	8.45	10.53
θ	0.0	2.269	2.40 ± 0.01	2.1 ± 0.01	10.44	14.67
δ	2.0	2.43	2.85 ± 0.05	1.7 ± 0.02	0.688	0.687

approaches.[11,48,49] Here the results from Monte Carlo simulations and Lagrangian theory approaches are in excellent agreement for two and three-dimensional chains. A scaling law for the distribution $P(R_{ete})$ has been derived by des Cloizeaux for an on-lattice SAW[11]

$$P(\mathbf{R}_{ete}, N) = \frac{1}{\xi^d} C\rho^\theta e^{-A\rho^\delta} \qquad (2)$$

where $\rho = R_{ete}/\xi$, $\xi^2 = \langle {R_{ete}}^2 \rangle/(2d)$ where d is dimension. The mean-squared end-to-end distance is given by $\langle {R_{ete}}^2 \rangle = B^2 \ell^2 N^{2\nu}$ where ℓ is the bond length (taken to be unity), the exponent $\nu = 0.5874$ as given above and the prefactor $B \approx 1.1$ is a number of order unity that is nevertheless important for quantifying return probabilities. We found that best-fit values for B varied from about 1.08 and 1.11 for both on- and off-lattice walks. This value is comparable to that obtained ($B = 1.2$) from end-to-end distance data in Monte Carlo simulations of off-lattice self-avoiding chains of various lengths.[50] The normalization constant C is a nonuniversal number of order that varied from 2 to 10 for off-lattice walks and 4 to 20 for on-lattice walks, as N varied from 50 to 200. Literature values for on-lattice SAWs are given in Table 1. For an ideal random walk, $\nu = 0.5$ and $B = 1.0$; values of A, θ, and δ are also given in Table 1 for comparison.

For off-lattice walks, we found that the parameters A and exponents θ and δ converged to the universal well-known on-lattice values in 3-dimensions for large N. However for the finite sizes in the range of typical protein chain lengths ($N \approx 100$), the end-to-end distribution was better fit by different exponents, which are given in Table 1. These exponents are obtained by best fit to off-lattice end-to-end distribution functions for $N = 50$, 100, and 200.

The configurations of real polymers are better described by off-lattice self-avoiding random walks rather than on-lattice models, which allow the angle between three consecutive monomers to have any value consistent with steric volume constraints. We have written a C++ algorithm to generate off-lattice SAW conformations by the well-known pivot algorithm, which has been shown to deal effectively with the attrition problem for SAWs.[38,51] Other generating algorithms which solve the attrition problem are also widely used,[52,53] The attrition problem corresponds to the fact that for a polymer to avoid a sterically excluded region, the appropriate boundary condition corresponds to that of an absorbing boundary. Any generated conformation which penetrates into the boundary must then be removed from the statistics. Thus, in growing a conformation of a self-avoiding polymer, any added monomer that by chance wanders into the sterically excluded region corresponding to the previously generated monomer positions must be eliminated, and the walk must be either reinitiated or appropriately reweighted.[54] Walks that survive this process become exponentially rare as the chain length increases.

The pivot algorithm is implemented for off-lattice SAWs by first generating a straight-line N-step walk, with residues placed at a distance ℓ apart. New conformations are then obtained by performing random symmetry operations at random positions i along the chain. For off-lattice walks these symmetry operations include rotations around the pivot point with arbitrary Euler angles. If the distance $r_{i+1,j}$ for any $j < i$ is less than 2σ, where σ is the monomer radius taken here to be $\sigma = \ell/2$, the walk is canceled and a new pivot operation is attempted until a viable SAW of N steps is obtained. Because the pivot algorithm generates radically different conformations, after $\sim N^{0.19}$ moves a globally different conformation is achieved.[38] The initial straight-line configuration was "equilibrated" typically for 50 pivot operations before sampling chain configurations.

A plot of the autocorrelation function for the root mean squared deviation (rmsd) between conformations as a function of pivot step t is given in the inset of Figure 2a, for a polymer with $N = 100$. The rmsd has decorrelated (such that the correlation function reaches $1/e$) after about 3 successful steps. The correlation function is not fit well by a simple exponential decay. The best fit was given by a stretched exponential $C(t) = \exp(-\alpha t^\beta)$ with $\alpha = 0.62$ and $\beta = 0.70$.

To see the effects of a self-avoiding vs a reflecting boundary condition on SAW statistics, we have also generated random walks using a naive growth algorithm that corresponds to a reflecting boundary condition as follows. Walks are generated as above, with the $i + 1$th residue placed a distance ℓ from the ith residue at random angle; but now if the distance $r_{i+1,j}$ for any $j < i$ is less than 2σ, only the last step is canceled and a new step is attempted, until a walk of N steps is generated. The process is then repeated from the first step to generate a new conformation. In a study by Dayantis and Sturm,[55] reflecting or absorbing boundary conditions on a spherical confining boundary resulted in significantly different concentration profiles as a function of radius, as well as different scaling laws for the entropy cost of confinement as a function of sphere radius.

Using the above generating methods, we have found that the end-to-end distribution for an off-lattice SAW has the same functional form as the on-lattice SAW in (eq 2), however in the range of protein chain lengths the universal exponents A, θ, and δ had significant finite-size corrections. The resulting values are given in Table 1.

Figure 1 compares the end to end probability distribution of an on-lattice SAW and an off-lattice SAW for a 200-mer. The

on-lattice constraints tend to contract the polymer relative to the off-lattice walk. On-lattice SAWs that we generated using the pivot-algorithm recovered the same set of universal parameters for an on-lattice SAW as derived in previous studies.[11]

Off-Lattice SAWs from Discontinuous Molecular Dynamics Simulations. Discontinuous molecular dynamics (DMD) is an efficient method that has been used to study protein folding and *ab initio* protein thermodynamics and kinetics,[6,56−59] structure prediction,[60] protein aggregation,[61,62] and the effects of osmolyte−protein interactions.[6] A brief description of the DMD method for freely jointed polymer chains is presented here (more complete descriptions of DMD methods for polymers and proteins are contained in the above references). We used DMD simulations for the purpose of generating polymer configurations. In our model the polymer is a freely jointed chain of N beads or monomers and $N - 1$ joints, wherein each monomer is represented as a hard sphere. Two bonded beads i and j are constrained to be within 10% of an average distance, ℓ, by an infinite square-well potential:

$$u_{i,i+1}^{bond} = \begin{cases} \infty, & r \le 0.9\ell \\ 0, & 0.9\ell < r \le 1.1\ell \\ \infty, & 1.1\ell < r \end{cases} \quad (3)$$

where $\ell = 1$. Two nonbonded atoms i,j may interact by hard-sphere potential (purely repulsive) with a hard-core radius

$$u_{ij}^{non-bond} = \begin{cases} \infty, & r \le \sigma_{HS} \\ 0, & \sigma_{HS} < r \end{cases} \quad (4)$$

where σ_{HS} is the hard-sphere diameter of a monomer. In this work we set $\sigma_{HS} = \ell$.

Using these parameters, we have performed DMD simulations on homopolymers of lengths 50, 100, and 200 with the above potentials, as a confirmation of SAW statistics generated by the pivot algorithm mentioned above. The system is simulated at a finite temperature (in practice this was $T = 10$ K; however, since there is no interaction energy-scale for a purely self-avoiding polymer, any finite temperature will generate the same equilibrium ensemble). Temperature equilibration is achieved by ghost particle methods. Moves are generated by integrating Newton's equations and conserving momentum and energy for interparticle collisions. Statistics such as the probability distribution of the end-to-end distance are calculated by sampling 1 000 000 conformations for each polymer chain length.

The autocorrelation function for the root mean squared deviation (rmsd) between conformations as a function of DMD simulation time-step t is plotted in Figure 2a, for a polymer with $N = 100$. The autocorrelation function reaches $1/e$ after about 31 steps. The function is not fit well by a simple

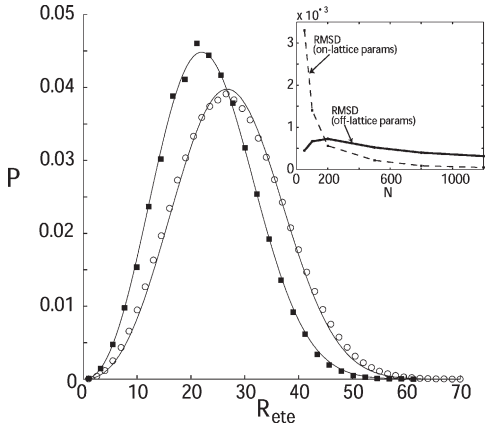


Figure 1. Comparison of the end to end probability distribution of an on-lattice SAW of $N = 200$, as generated by the on-lattice pivot algorithm described in the text (black filled squares), and an off-lattice SAW generated by the pivot algorithm (open circles). The curves passing through each set of data are fits using the des Cloizeaux functional form in (eq 2). For the on-lattice data the well-known exponents in Table 1 column 2 are used, and for the off-lattice data the exponents in Table 1 column 3 are used. Inset shows the root-mean-square deviation of the residuals in the best fit to the simulation data using the on-lattice des Cloizeaux parameters (dashed line), and the finite-size parameters obtained from best fit to the simulation data for sizes $N = 50$, $N = 100$, and $N = 200$.

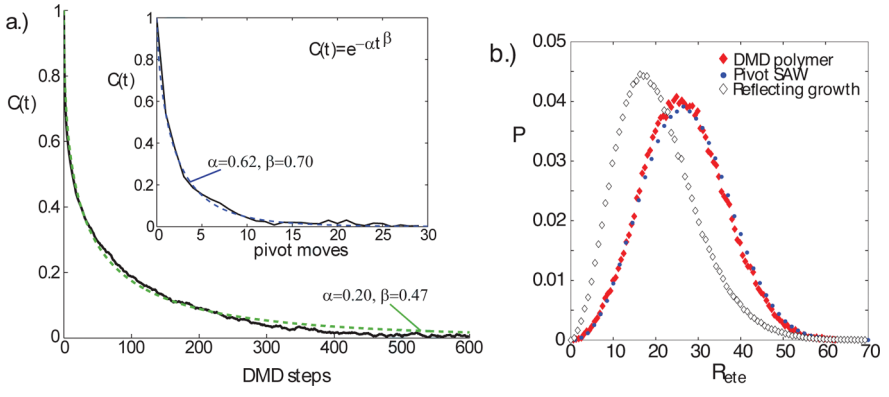


Figure 2. (a) Autocorrelation function of the root mean squared deviation $rmsd(t) = ((1/N)\Sigma_i(\mathbf{r}(t) - \mathbf{r}(0))^2)^{1/2}$ as a function of time step t in the DMD simulations, or number of successful moves t for the pivot algorithm. The pivot algorithm decorrelates over much fewer moves than the DMD simulations. Stretched exponentials were needed to fit both correlation functions, with exponenets given in the figure legends. (b) End to end probability distribution of a continuum SAW with $N = 200$ using a naive growth algorithm corresponding to a reflecting boundary condition (open black diamonds), the pivot algorithm (blue circles) and DMD simulations (red filled diamonds). The naive chain growth algorithm with effective reflecting boundary condition when dead-ends are encountered does not provide the correct end-to-end statistics of a SAW. The effective absorbing boundary condition of a self-avoiding walk shifts the distribution to larger values of end-to-end distance. The pivot algorithm and DMD simulations show the same statistics within sampling error.

dx.doi.org/10.1021/ma200454e |*Macromolecules* 2011, 44, 6182–6197

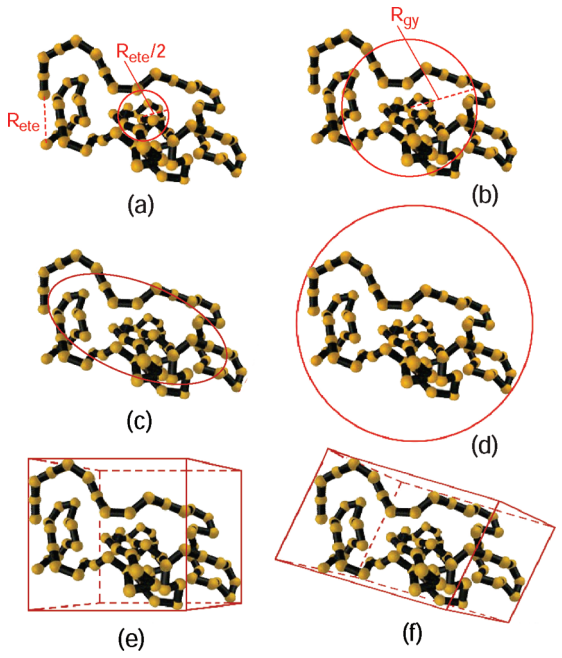


Figure 3. Comparison between different ways of measuring the size of a chain: (a) the end to end distance model, (b) the radius of gyration model, (c) radius of gyration tensor, (d) embedding sphere model, (e) Cartesian box model, and (f) principal box model. Schematic polymer figures are adapted from ref 63.

exponential decay. The best fit was given by a stretched exponential $C(t) = \exp(-\alpha t^{\beta})$ with $\alpha = 0.20$ and $\beta = 0.47$.

A plot of the end-to-end distribution for a walk of length $N = 200$, as generated by the pivot algorithm, naive growth algorithm, and DMD simulations, is shown in Figure 2b. As expected, the DMD simulations reproduce the same statistics as those of the pivot algorithm for a continuum SAW. However this distribution is significantly expanded relative to that generated by the naive growth algorithm with reflecting boundary conditions.

Measures of Polymer Volume and Their Properties. To determine an accurate measure of the volume occupied by a polymer, we considered six different ways to characterize a specific conformation of polymer. Figure 3 depicts the six different models:

(a) The end-to-end distance model, which approximates polymer size by a sphere centered at the center of mass of the polymer, having radius equal to the end-to-end distance R_{ete} over 2. This volume is also called the Flory volume.

(b) The volume of gyration model, where the volume measure is given by a sphere having radius equal to the radius of gyration R_{gy} of the polymer, defined through 6 below.

(c) A tensorial generalization of the radius of gyration where the radius of gyration tensor is defined analogously to the moments of inertia, then diagonalized to obtain radii a, b, and c along the principal axes.[14−20,22,23,64−68] Such a measure of polymer size has formed the basis of characterizing anisotropy through the asphericity.

(d) The embedding sphere model, which consists of a sphere with center located at the center of mass of the polymer, and radius $R = |\vec{r}_f - \vec{r}_{com}|$ where $\vec{r}_f$ is the position of the farthest monomer from center of mass; This measure is closely related to the miniball,[29,69] which is the smallest sphere containing all the residues in the polymer (and need not be centered at the center of mass).

(e) The Cartesian box model, which is a box oriented in a fixed Cartesian frame of reference, with volume v given by $|x_{max} - x_{min}| \times |y_{max} - y_{min}| \times |z_{max} - z_{min}|$ where x_{min} and x_{max} are the x-components of the position vector of the monomers with smallest and largest components along the x-axis correspondingly (the same definition applies in y and z directions). This measure was first introduced by Rubin and Mazur.[21]

(f) The principal box model, which determines polymer volume using a box aligned along the principal axes of the polymer in each conformation, having volume $v = |r_{1max} - r_{1min}| \times |r_{2max} - r_{2min}| \times |r_{3max} - r_{3min}|$ where r_{1min} and r_{1max} are the components of the position vector of the monomers with smallest and largest components along the first principal axis respectively (the same definition applies in second and third principal axis directions). This is a kind of "giftbox" of nearly minimal size that encloses the polymer. It was first introduced by Rubin and Mazur,[22] who also studied minimal spanning boxes constructed from a projection recipe by taking successive pairs of maximal distances. The projection recipe was also applied to knotted polymers by Millett and Rawdon,[29,37] where it is referred to as the "standard box". A more efficient construction of what is likely the minimal volume rectangular cuboid enclosing the residues is the so-called *skinny box* construction of Rawdon et al.,[29] which involves projections onto successive pairs of parallel orthogonal planes separated by minimal distance.

A scatter plot of the end-to-end radius R_{ete}, defined as the end-to-end distance over 2 (method (a) above), and the radius R_{sphere} of the smallest sphere which circumscribes the polymer configuration (method (d) above) is given in the Supporting materials. R_{sphere} is always larger than R_{ete}, since the distance between any pair of residues must always be less than the embedding sphere's diameter.

The radius of gyration tensor $\mathbb{S}$ is constructed from the dyadic of the position column vector $\mathbf{r}_i = (x_i, y_i, z_i)^T$ in the center of mass system (where $\sum_1^N \mathbf{r}_i = 0$) as

$$\mathbb{S} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{r}_i \mathbf{r}_i^T = \frac{1}{N}\begin{bmatrix} \sum x_i^2 & \sum x_i y_i & \sum x_i z_i \\ \sum x_i y_i & \sum y_i^2 & \sum y_i z_i \\ \sum x_i z_i & \sum y_i z_i & \sum z_i^2 \end{bmatrix} \quad (5)$$

Transforming to the principal axis diagonalizes $\mathbb{S}$ to $(1/N)diag(\sum x_i'^2, \sum y_i'^2, z_i'^2)$, with eigenvalues corresponding to the variances of the coordinates along the principal axes directions.

The moment of inertia tensor $\mathbb{I}$ is directly related to the radius of gyration tensor $\mathbb{S}$ by $\mathbb{I} = \mathrm{tr}(\mathbb{S})\mathbb{1} - \mathbb{S}$, where $\mathbb{1}$ is the unit tensor, and tr$\mathbb{S}$ is the trace of the radius of gyration tensor giving the sum of the eigenvalues. Thus, both $\mathbb{I}$ and $\mathbb{S}$ are diagonal in the same principal axis system, and tr $\mathbb{I}$ = 2tr $\mathbb{S}$. The trace of $\mathbb{S}$ is a scalar invariant that is independent of the coordinate basis chosen; this scalar invariant is defined as the radius of gyration squared of the polymer configuration:

$$\mathrm{tr}(\mathbb{S}) = \frac{1}{N}\sum_{i=1}^{N}(x_i'^2 + y_i'^2 + z_i'^2) = R_{gy}^2 \quad (6)$$

dx.doi.org/10.1021/ma200454e |*Macromolecules* 2011, 44, 6182–6197

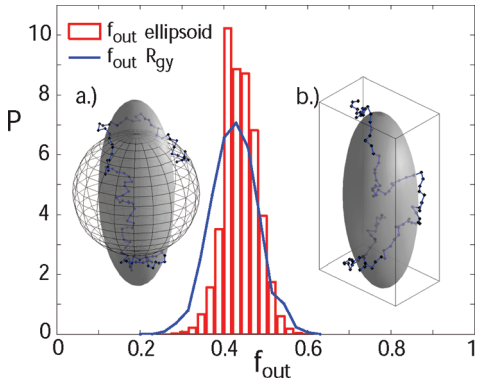


Figure 4. Fraction of the total residues of SAWs that lie outside of the volume of the radius of gyration ellipsoid defined in eq 7 (red histogram), and the fraction of excluded residues for a sphere with radius equal to the radius of gyration (blue curve). Volumes are centered at the center of mass. Histograms are generated for a 100-mer. Insets show two configurations taken from the right side tail of the distribution, with large amounts of polymer excluded from the corresponding volume. Inset a shows a configuration with 63 residues outside the radius of gyration sphere, and 48 outside of the radius of gyration tensor ellipsoid (sphere and ellipsoid are also shown). Inset b shows a configuration with 43 residues outside of the radius of gyration sphere, and 60 outside of the radius of gyration tensor ellipsoid (tensor ellipsoid is shown, along with the principal box as defined in the text).

Written in terms of average values $R_{gy}{}^2 = \overline{x^2} + \overline{y^2} + \overline{z^2}$. If the system were isotropic, the radius of the effective sphere determining the characteristic volume of the polymer conformation is $R_{gy} = (3)^{1/2}\Delta x$, where Δx is the standard deviation along the coordinate x. Thus, if the system is anisotropic, the measure of the effective radius along each principal axis is $(a,b,c) = (3^{1/2}\Delta x, 3^{1/2}\Delta y, 3^{1/2}\Delta z)$ respectively. The volume of the effective ellipsoid characterizing the polymer configuration is then

$$V_{ell} = \frac{4}{3}\pi abc = \frac{4}{3}\pi 3^{3/2}\Delta x\Delta y\Delta z = 4\pi\sqrt{3}\prod_{i=1}^{3}\sqrt{\lambda_i} \quad (7)$$

where λ_i are the eigenvalues of the radius of gyration tensor. The effective radius of a sphere with the same volume as eq 7 is given by the geometrical mean of the individual radii: $R_{gy}^{eff} = (abc)^{1/3} = (3)^{1/2}\Pi_{i=1}^{3}\lambda_i{}^{1/6}$. This is to be compared with the volume of the effective sphere defined by $(4/3)\pi R_{gy}{}^3$, with $R_{gy} = (\Sigma_{i=1}^{3}\lambda_i)^{1/2}$. A scatter plot of the radii determined by both of the above measures is shown in the Supporting Material. The radius of gyration R_{gy} is always larger than the effective radius R_{gy}^{eff}.

Both of the above measures result in spheres or ellipsoids that do not encompass all of the polymer, and so result generally in different packing fraction than would be obtained from a box which encloses the whole polymer. We propose using the principal box method in order to calculate polymer packing fraction. This method characterizes polymer shape through three scalars representing the edge lengths of the principal box, and so it accounts for the anisotropy which may be present in a given conformation. Anisotropy cannot be captured in measures characterizing polymer size by a single scalar. The principal box also contains all the monomers in the polymer by definition. In the limit of a long polymer, only a small error is introduced by neglecting the finite size of the monomers in the definition of the principal box. We thus use the geometric centers of the monomers in defining the boundaries of the principal box.

Figure 4 shows histograms of the fraction of residues outside of the volumes defined by the radius of gyration, and the radius of gyration tensor. The radius of gyration tensor generally excludes slightly more residues (44.1% on average for a 100mer) than the radius of gyration sphere, which excludes 42.5% on average. This is essentially due to the much larger volume of the radius of gyration sphere than the radius of gyration tensor ellipsoid (about 2.8 times larger on average for SAWs of a 100-mer).

Figure 5 shows scatter plots of various volume measures of a $N = 100$ polymer, for about 2×10^4 conformations. All volumes in the analysis are given in units of the volume of a cube with edge length equal to the link length (i.e., the monomer−monomer separation). Figure 5a shows the principal box volume V_{pb} (method e above), vs the Flory volume V_{ete} of the polymer (method a above). The two measures differ significantly and do not correlate well: correlations between all quantities considered are given in Table 2 for this data set. A particular outlying conformation for which V_{ete} overestimates the volume is shown in the inset. For such anisotropic conformations, V_{pb} (shown for this configuration) or the radius of gyration tensor ellipsoid volume V_{ell} are necessary to quantify the volume. Likewise V_{pb} does not show a strong correlation with the scalar volume of gyration V_{gy} (method b above) (Figure 5b and Table 2). Figure 5c shows that both V_{pb} and V_{ell} capture the anisotropy inherent in the ensemble of SAWs. The two measures correlate well, but as discussed above in the context of Figure 4, the radius of gyration tensor ellipsoid does not enclose a significant fraction of the monomers, thus the slope of the curve in Figure 5c is not unity, but ≈2.1. Figure 5e shows that alignment of the bounding box significantly alters the volume and is so important for quantifying the polymer density. Perhaps not surprisingly, the unaligned bounding box volume V_{box}, the Flory volume V_{ete} and the volume of gyration V_{gy} all correlate reasonably well with each other (Table 2).

The steric molecular volume is calculated using the *mol_volume* program,[70] which enumerates the number of vertices in a dense cubic lattice grid that are less than a probe radius in distance from all of a molecule's atoms. For this computation a probe with radius equal to the monomer radius of was used (here ⁄/2). The molecular volume is smaller than the polymer volume by any of the above metrics for essentially all configurations. As expected, none of the volume measures correlate particularly well with the molecular volume of the polymer, indicating that conformational volume has only an indirect connection with solvent-excluded volume. This can be seen in Figure 5f, where most points in the distribution lie in the northwest half of a rectangle, above the imaginary diagonal. If a configuration has large V_{pb}, the system then tends to be expanded with minimal overlap between solvent-excluded regions and thus have large molecular volume. Conversely if the system has small molecular volume with significant overlap between solvent-excluded zones, the principal box volume tends to be small. However, one cannot say that a small principal box volume guarantees a small molecular volume, as can be seen for the configuration in the inset of Figure 5a. The inset to Figure 5f shows the configuration with smallest solvent-excluded volume for this sampling. It is not particularly remarkable, except for having a higher than usual number of local contacts.

Figure 6 compares the effective diameter probability distributions using the various models of the size of a 100-mer. Employing the letter notation above describing the six volume measures, the effective linear size of the polymer is given by (a) R_{ete},

dx.doi.org/10.1021/ma200454e |*Macromolecules* 2011, 44, 6182–6197

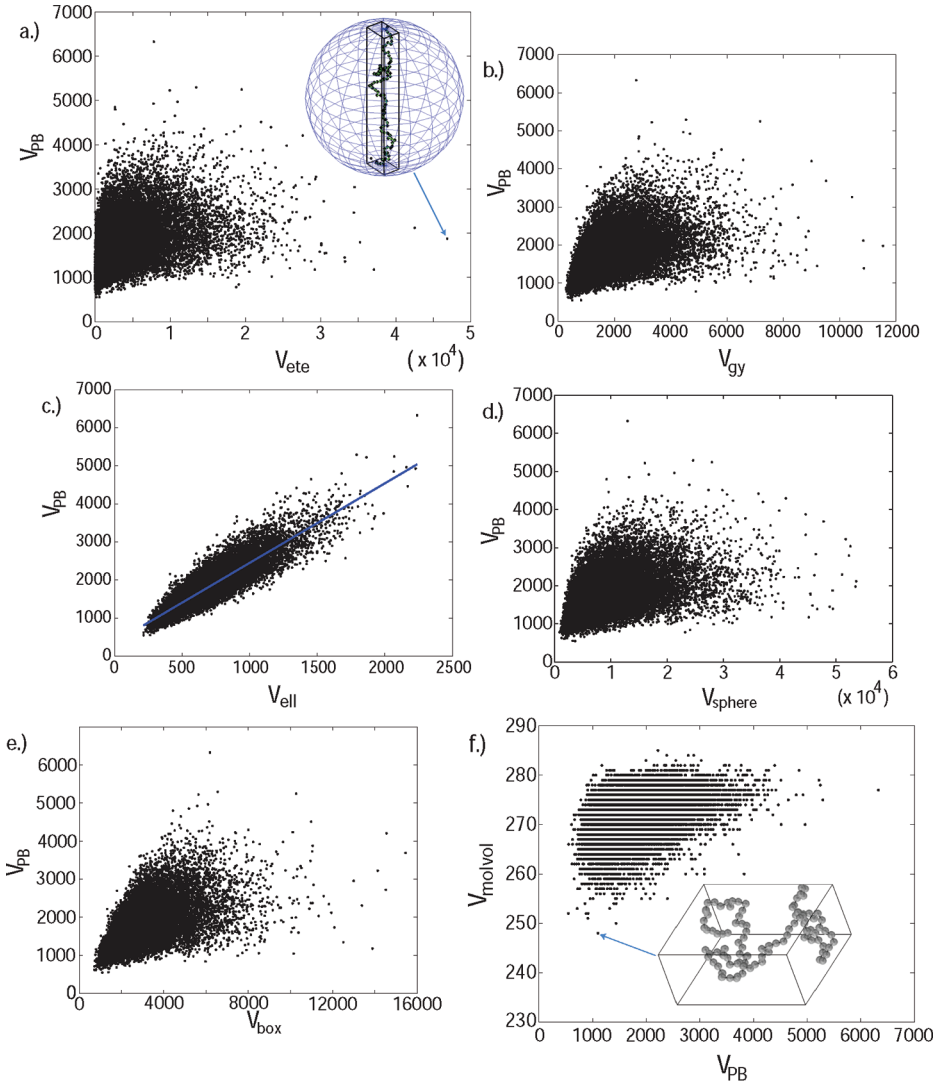


Figure 5. Scatter plots showing the relationship between the principal volume and other quantities characterizing the size of a polymer, for a SAW with $N = 100$. Correlation coefficients between all quantities are given in Table 2. (a) Principal box volume vs the Flory volume $V_{ete} = (4/3)\pi(R_{ete}/2)^3$. (b) Principal box volume vs the volume determined by the radius of gyration $V_{gy} = (4/3)\pi {R_{gy}}^3$. (c) Principal box volume vs the volume of the radius of gyration tensor ellipsoid V_{ell} in eq 7. (d) Principal box volume vs the embedding sphere volume V_{sphere}. (e) Principal box volume vs the unaligned box volume V_{box}. (f) Steric molecular volume vs the principal box volume. All volumes are given in units of the link length cubed (ℓ^3).

Table 2. Matrix of Correlation Coefficients for the Various Volume Measures shown in Figure 3 and Figure 5′[a]

	V_{ete}	V_{gy}	V_{ell}	V_{sphere}	V_{box}	V_{PB}	V_{molvol}
V_{ete}	1	0.813	0.308	0.860	0.703	0.379	0.254
V_{gy}	0.813	1	0.508	0.864	0.800	0.477	0.319
V_{ell}	0.308	0.508	1	0.380	0.564	0.891	0.398
V_{sphere}	0.860	0.864	0.380	1	0.750	0.446	0.295
V_{box}	0.703	0.800	0.564	0.750	1	0.571	0.371
V_{PB}	0.379	0.477	0.891	0.446	0.571	1	0.391
V_{molvol}	0.254	0.319	0.398	0.295	0.371	0.391	1

[a] All correlations are statistically significant ($p < 10^{-278}$).

(b) $d_{gy} = 2 \times R_{gy}$, (c) $d_{ell} = (6V_{ell}/\pi)^{1/3}$, (d) $d_{sphere} = 2 \times R_{sphere}$, (e) $d_{box} = {V_{box}}^{1/3}$, and (f) $d_{pb} = {V_{pb}}^{1/3}$.

For many conformations the end to end distance description does not accurately represent the size statistics of a polymer. It is the most broadly distributed quantity in Figure 6. It can be anomalously large, or can be zero, whereas even in the most collapsed conformation the real size of a polymer cannot be smaller than $\approx/N^{1/3}$. Using the size distribution of principal box as a benchmark, the embedding sphere largely overestimates the size of the polymer, as well as does the unaligned Cartesian box model. The diameter of gyration also overestimates the size of the polymer, which is somewhat surprising given the amount of polymer excluded by the radius of gyration sphere (Figure 4). For this reason we analyze the amount of anisotropy present in the polymer distribution below. As illustrated in Figure 4, the radius of gyration tensor ellipsoid underestimates the size of the polymer by excluding a significant fraction of monomers. Also plotted in Figure 6 is Lhuillier's functional form[13,32,71,72] for $P(d_{gy}) = P(d_{gy}{}^*)e^{-B(4\rho^{-15/4}/5+6\rho^{5/2}/5-2)}$, where $\rho = d_{gy}/d_{gy}{}^*$, $P(d_{gy}{}^*)$ is the peak value of the distribution which occurs at $d_{gy} = d_{gy}{}^*$, and $B = 1.26$ is obtained by best fit to the pivot simulation data.

dx.doi.org/10.1021/ma200454e |*Macromolecules* 2011, 44, 6182–6197

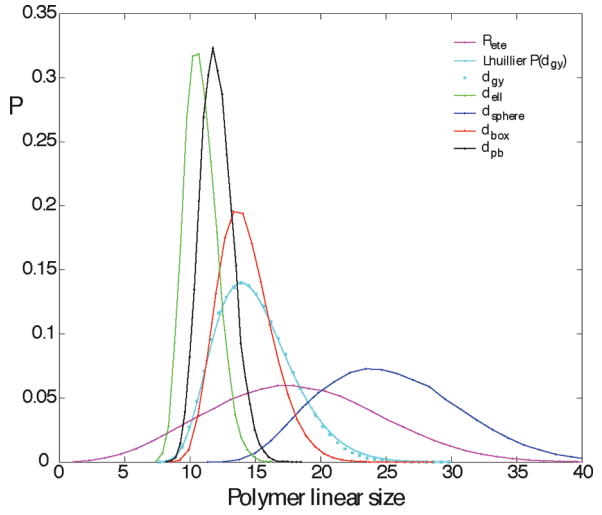


Figure 6. Probability distributions of various size measures of the polymer, for a 100mer SAW distribution obtained from the pivot algorithm. Magenta: the end-to-end distance probability distribution. Cyan squares: Probability distribution of 2 times the radius of gyration (i.e., the diameter of gyration). Cyan solid line: Plot of Lhuillier's functional form for the probability distribution of the diameter of gyration[13,32,71,72] ($d_{gy} = 2R_{gy}$). Green: Probability distribution of the mean diameter of the radius of gyration tensor ellipsoid, defined by $d_{ell} \equiv (6V_{ell}/\pi)^{1/3}$. Blue: Probability distribution of the embedding sphere diameter (the smallest sphere enclosing the polymer). Red: Probability distribution of the mean edge length d of the unaligned box enclosing the polymer, defined by $d_{box} \equiv V_{box}^{1/3}$. Black: Probability distribution of the mean edge length $d_{pb} \equiv V_{pb}^{1/3}$ of the box enclosing the polymer, after the principal axes of the radius of gyration tensor for the polymer have been aligned to the axes of the box.

It is observed in Figure 6 that the principal box size distribution is the sharpest peaked of the distributions relative to the mean. Table 3 compares the means and standard deviations for the volume (rather than size) distributions of the various measures. Also given is the "relative error" of associating the mean value of the distribution with any conformation. By this measure the principal box volume best characterizes the volume of the polymer. The improvement of principal box over the tensor ellipsoid volume is a 10% effect, however it is statistically significant, and only increases with larger N.[73]

For the rest of this work, we use the principal box volume V_{pb} as the representative volume of the polymer, and use V_{pb} or V interchangeably.

Anisotropy and Acubicity. As discussed in the introduction, there has been substantial interest historically in the anisotropy of polymer walks. The degree of anisotropy of polymer conformations can be quantified for the principal box measure by taking the edges of the principal box $a > b > c$, and forming the ratios $f = b/a$ and c/a. Ratios f less than unity indicate anisotropy. A ratio of b/a (or c/a) less than unity breaks cubic symmetry (O_h(*432)) to a square prism with D_h(*422) symmetry. A ratio of c/b less than unity breaks D_h(*422) symmetry to a rectangular cuboid with symmetry D_{2h}(*222). The probability distributions of these ratios for the SAW ensemble of a 100mer are plotted in Figure 7a. We also plot the aggregate distribution with b/a and c/a grouped together.

The distribution indicates substantial "square prism" anisotropy, with the aggregate f distribution broadly peaked at around

Table 3. Mean values of the volume measures shown in Figure 3 and Figure 5, standard deviations, and standard deviation over the mean, which gives an estimate as to how well the mean value accurately represents the conformational ensemble

	mean volume (10^3/^{3})	std deviation (10^3/^{3})	relative deviation
V_{ete}	4.30	4.24	0.99
V_{gy}	1.92	1.13	0.59
V_{ell}	0.689	0.240	0.35
V_{sphere}	9.73	6.22	0.64
V_{box}	3.07	1.40	0.46
V_{PB}	1.80	0.562	0.31
V_{molvol}	0.272	0.0041	0.015

$f = 0.34$. The "cuboid" anisotropy is somewhat milder, peaking around $f = 0.6$. Ensemble average values of the ratios of principal box edge lengths as well as standard deviations are given in Table 4.

As mentioned in the Introduction, the asphericity as given by eq 1 has been used as a measure of polymer asymmetry by many authors. Although the corresponding quantity for the principal box may not be (as) amenable to analytical approaches, it is instructive to define an analogous quantity to the asphericity in terms of the principal box edge lengths, rather than the eigenvalues of the radius of gyration tensor. We refer to this as the "acubicity" Δ_c:

$$\Delta_c = \frac{3}{2}\frac{\sum_{i=1}^{3}(L_i^2 - \overline{L^2})^2}{(3\overline{L^2})^2} \tag{8}$$

Here L_i are the principal box edge lengths of a configuration, and $\overline{L^2}$ is the average of their squares. The distribution of Δ_c is shown in Figure 7b. It is broadly distributed over much of its possible range ($0 < \Delta_c < 1$), with an average of $\langle\Delta_c\rangle = 0.348$ and standard deviation 0.172 as given in Table 4. By this measure the acubicity gives larger anisotropy than the asphericity (see e.g. Bishop and Saltiel[74] or Goldbart and co-workers[24] who have reported $\langle\Delta\rangle$ = 0.46, and 0.447, respectively).

From the definition in eq 8 it is clear that a more "reduced" measure of anisotropy can be made that involves the linear dimensions of the principal box rather than its dimensions squared:

$$\Delta_c' = \frac{3}{2}\frac{\sum_{i=1}^{3}(L_i - \overline{L})^2}{(3\overline{L})^2} \tag{9}$$

This measure also satisfies $0 < \Delta_c' < 1$, however its distribution is substantially more sharply peaked than Δ_c, as shown in Figure 7b, and the value of Δ_c' where it peaks is smaller than the corresponding distribution for Δ_c. The average $\langle\Delta_c'\rangle = 0.113$ and standard deviation 0.065 are also given in Table 4. For these reasons it is apparent that Δ_c' is a more sensitive measure of anisotropy than Δ_c.

Principal Box Volume Distribution Function. As described above, the effective volume of the polymer is taken to be the volume of a box aligned with the polymer's principal axes that contains all the monomers. Using both the pivot algorithm and

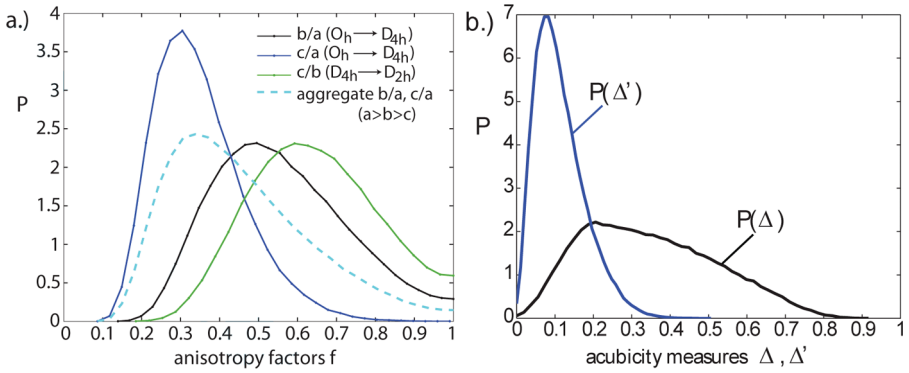


Figure 7. (a) Probability distribution of the anisotropy factors for the principal box enclosing 100mer polymer configurations, defined by taking the ratio of the smaller edge lengths to the largest edge length. Smaller values indicate higher anisotropy: a cubic box would have an anisotropy ratio $f = 1$. Cuboid edge lengths of the principal box are defined such that $a > b > c$. Ratio (black) of the second largest edge length to the largest edge length (b/a), (blue) ratio of the smallest to largest edge lengths (c/a), (green) ratio of the second smallest to smallest edge lengths (c/b), (dashed cyan) aggregate distribution of the ratios of either of the smaller edge lengths to the largest edge length. Symmetry groups corresponding to these anisotropies are indicated in the legend. (b) Distributions of acubicity measures, defined analogously to the asphericity in eq 1. The acubicity Δ' defined through a linear scaling with edge lengths is much more sharply peaked than the traditionally defined acubicity Δ. Both distributions show significant anisotropy.

Table 4. Mean Values and Standard Deviations of Various Anisotropy Measures of the Principal Box, Whose Distributions Are Shown in Figure 7, Parts a and b[a]

($N = 100$)	mean	std. deviation
b/a	0.557	0.168
c/a	0.349	0.114
c/b	0.643	0.159
$(b,c)/a$	0.453	0.178
Δ_c	0.348	0.172
Δ_c'	0.113	0.065

[a] Included are the ratios of the edge lengths of the principal box, and the acubicity measures described in the text.

DMD simulations of homopolymers, we have calculated the volume of the polymer in each conformation. Figure 8 displays the volume probability distributions of a 50-mer, 100-mer, and 200-mer. We have found that a semiempirical extension of the des Cloizeaux functional form

$$P(V,N) = \frac{C}{\delta V}\left(\frac{V - V_o}{\delta V}\right)^{\theta'} e^{-A'((V - V_o)/dV)\delta'}$$

$$\approx \frac{C}{\delta V}(V/\delta V)^{\theta'} e^{-A'(V/\delta V)^{\delta'}} \quad (10)$$

also applies well to the principal box volume of the polymer. In eq 10, δV is the standard deviation of the principal box volumes, and V_o is the approximate volume of a collapsed globule, taken to be equal to $b\sqrt[3]{N}$ with b an adjustable parameter. Subtraction of the quantity V_o accounts for the fact that the volume cannot be smaller than that of the collapsed globule. Consideration of a term such as V_o is appropriate for a metric accounting for polymer volume (as opposed to say end-to-end distance which can in principle be zero). Curve fitting to the pivot algorithm data for $N = 100$ gives the scaling exponents in Table 1 from the best fit solution. Several solutions were found, however, with variations in the exponents of approximately $(\Delta A'/A', \Delta\theta'/\theta', \Delta\delta'/\delta') = (0.06, 0.1, 0.004)$. The proportionality constant b in the collapsed globule volume was always of order unity, and for the best fit $b = 2.43$. The normalization constant C for the best fit solution was 240, and varied by about 5%.

Typical polymer configurations have substantially larger volumes than the collapsed globule, however, so we tested the importance of including the term V_o in eq 10. Removing the term V_o and finding best fit exponents gives the values in Table 1. Again several solutions were found however now with smaller variations $(\Delta A'/A', \Delta\theta'/\theta', \Delta\delta'/\delta') = (0.005, 0.003, 0.002)$. The best fits for both recipes are plotted in Figure 8b. The distributions have quite similar residuals with the distribution including the factor V_o having marginally better fit. An Ansari–Bradley test was used to find the statistical significance of the improvement in fit due to the additional parameter V_o, by choosing samples from the two different distributions and asking what the probability was that they would be obtained from the same distribution (the p-value). For a large enough sample size, any two nonidentical distributions eventually become distinguishable. The inset to Figure 8b plots the p-value as a function of sample size. The distributions are only statistically distinguishable once 5×10^6 or more polymer conformations are sampled. We thus conclude from these studies that the parameter V_o is not essential for characterizing the distribution, unless very high accuracy for large sample sizes is required.

Figure 8a shows both pivot and DMD results for the distribution of V_{pb} for a 50-mer. To test whether the same exponents as found for the 100-mer and given in the best fit solution of 10 to the distribution obtained from the DMD data, by either fixing A, θ, and δ to the 100-mer values, or allowing the values to be variable. In either case both the normalization constant C and the prefactor to the collapsed globule volume b were allowed to vary. Both methods fit the data very well, as shown in Figure 8a. The values of the exponents varied by the relative amounts $(\Delta A'/A', \Delta\theta'/\theta', \Delta\delta'/\delta') = (0.1, 0.05, 0.06)$ between the two methods. The constant b remained of order unity ($b = 1.847$). An Ansari–Bradley test on samples chosen from the two distributions finds them indistinguishable up to 2×10^6 conformations.

The 200-mer shows the same robustness in the exponents. Figure 8c shows best fits to the pivot algorithm data using eq 10

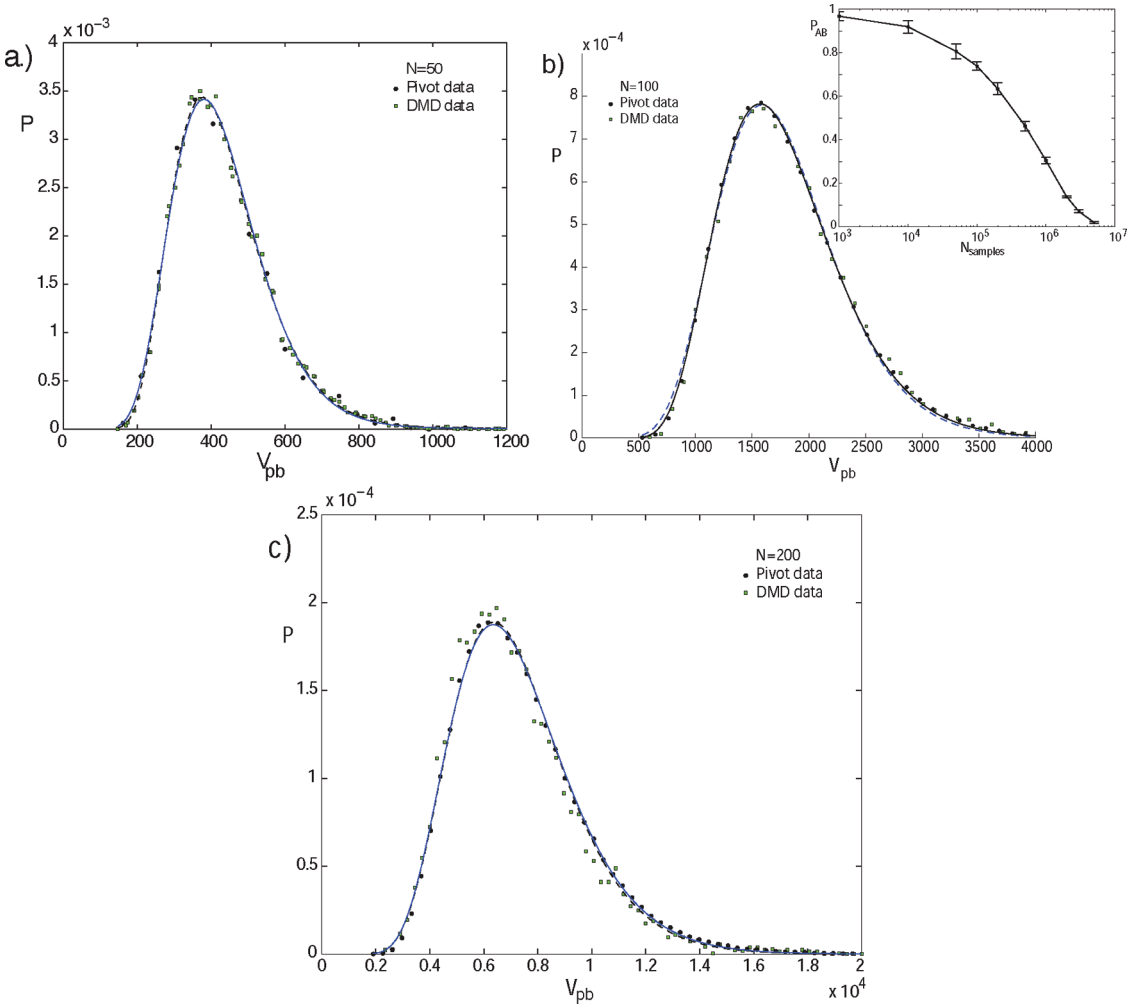


Figure 8. Probability distributions of the principal box volume V_{pb}, for a (a) 50-mer, (b) 100-mer, and (c) 200-mer. Both pivot data and DMD simulation results are shown. Panel (b) Also shows fits to the pivot algorithm data including (solid) and not including (dashed) the factor of V_o in eq 10. The statistical significance of the difference between these two very similar distributions is discussed in the text. Panel a also shows a fit to the DMD data with eq 10 using the parameters A, θ, δ derived from the 100-mer fit (solid line), to test the independence of these exponents on the chain length N. The dashed line shows the best fit allowing these parameters to vary. Likewise, in panel c, fits to the pivot algorithm data with fixed (100-mer) exponents (solid line) and variable exponents (dashed line) are shown. The results of the fitting do not depend significantly on whether the exponents are fixed or vary, so the exponents in eq 10 can be treated as essentially N-independent for the range of N used in this study.

with fixed and variable exponents as described above for the 50-mer. The values of the exponents varied by the relative amounts $(\Delta A'/A', \Delta\theta'/\theta', \Delta\delta'/\delta') = (0.02, 0.06, 0.01)$ between the two methods. The parameter b varied from 3.3 (variable exponents) to 3.5 (fixed exponents). An Ansari−Bradley test on samples chosen from the two distributions gives p-values of 90% for 10^5 samples and 72% for 10^6 samples, i.e., the distributions are indistinguishable at least up to millions of conformations. We thus conclude that the exponents A, θ, δ do not show significant N-dependence over the range of chain lengths characterizing typical proteins.

THERMODYNAMICS OF A MODEL POLYMER

Entropic considerations are very important in investigating the effects of the solvent on protein folding and polymer collapse. The folding or collapse dynamics involve a conformational search, guided by energetic bias, of all allowed states. The conformational entropy as a function of the polymer volume V is given by

$$S(V|N) = S_o + k_B \ln(P(V|N)) = S_o + k_B \ln \int d\Delta x d\Delta y d\Delta z\, \delta\,(\Delta x \Delta y \Delta z - V) P(\Delta x, \Delta y, \Delta z|N) \quad (11)$$

where S_o is the total polymer entropy and k_B is Boltzmann's constant. In eq 11, specifying volume V implies an integration over all possible boxes consistent with the constraint of a given volume. Now consider a polymer in sufficiently "poor" solvent such that the monomers have attractive interactions. Two nonbonded (nonconsecutive) monomers within a cutoff distance are given contact energy $-\varepsilon$, i.e. monomers i and j are in contact if $|i - j| \geq 3$ and $|\vec{r}_i - \vec{r}_j| < r_c$ where r_c is an interaction cutoff distance.

The packing fraction for a polymer is defined by the number of monomers contained within its characteristic volume, times

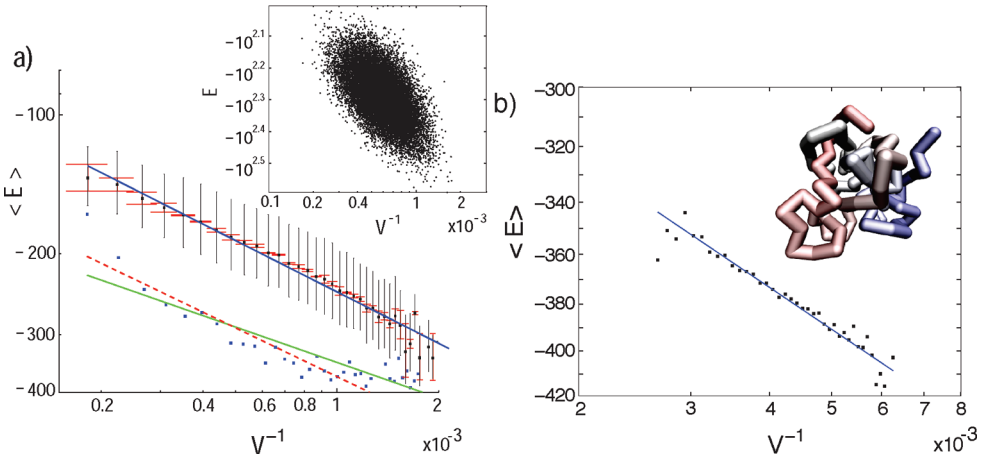


Figure 9. (a, Inset) Scatter plot on a log−log scale, of the energy defined through (minus) the number of contacts in an ensemble of 20 000 conformations of the 100mer, as a function of the reciprocal of the volume V_{pb}^{-1} for that conformation. The cutoff distance r_c used for this plot is 2.5 × the link-length. There is a significant correlation of about 0.6 between the number of contacts and V_{pb}^{-1}, however scatter indicates fluctuations conformation to conformation. (a) Log−log plot of the averaged number of contacts for configurations within a bin of 1/40th the width of the total distribution. Black vertical bars indicate the standard deviation of the distribution of points within each bin, while red error bars represent true statistical error of the mean values of energy within each bin. The linear fit to this data does not have an exponent of unity, but an anomalous exponent 0.37. (Blue squares) Boltzmann reweighted distribution of the number of contacts from eq 15 at a temperature such that $\beta\varepsilon = -0.2$. A linear fit to this data has slope 0.25 (green line). A linear fit to the first 2/3 of the data (red-dashed line) yields a slope of 0.35 which is still substantially less than unity. (b) Log−log plot of the mean energy vs reciprocal volume V^{-1}, for conformations resulting from DMD simulations of a self-interacting polymer with $\beta\varepsilon = -0.2$. Conformations are again sorted in bins having 1/40th the width of the total distribution of V^{-1}. The linear fit to this data has an anomalous exponent of 0.2, roughly consistent with the shallower values of slope observed in the reweighted pivot simulations for configurations having smaller values of volume. (b, Inset) Representative polymer configuration with values of energy and V^{-1} near the mean of the distribution (figure generated using VMD[82]).

the volume per monomer v_m, divided by the characteristic volume, i.e.

$$\eta(V) = Nv_m/V \tag{12}$$

The packing fraction depends on which measure is used for the protein volume; we take this volume to be the principal box volume as found previously.

In a random-mixing approximation,[7,75−81] the number of contacts n of an N-mer with volume V is given by

$$\langle n(V)\rangle = zN\eta(V) \tag{13}$$

where the constant z is the coordination number in a fully collapsed compact configuration (with $\eta = 1$). So the mean number of pairwise contacts in a polymer depends on the packing fraction η, which is itself determined by the volume that the polymer takes up.

In the mean-field approximation, the internal thermal energy of the polymer is determined by the average number of contacts in eq 13, and so is given by:

$$E(V) = -\varepsilon\langle \sum_{i<j}^{N} \delta(|\vec{r}_i - \vec{r}_j| < r_c)\rangle = -\varepsilon zN\eta(V)$$

$$= -\varepsilon zN^2 v_m/V \tag{14}$$

The validity of a volume measure can thus be tested for a given chain from plots of the mean number of pairwise contacts vs the reciprocal of the volume measure.

The inset of Figure 9a shows a scatter plot of (minus) the number of contacts in an ensemble of conformations of the 100mer, as a function of the reciprocal of the volume for that conformation as defined by the principal box metric. The large scatter indicates fluctuations from the mean trend conformation to conformation; the mean-field equation should be interpreted through the average over conformations with a given value of V_{pb}^{-1}. Figure 9a thus plots (minus) the averaged number of contacts for configurations within a bin of 1/40th the width of the total distribution. Like the scatter plot, this plot is shown on a log−log scale. The black vertical bars represent the standard deviation of the distribution of points within each bin, and the red error bars represent the statistical error in the average values due to the finite size of the data set.

Intriguingly, the slope of the best fit line to the $\ln(\langle E\rangle)$ vs $\ln(V^{-1})$ curve is not unity, but rather the unusual exponent of 0.37. One possible reason for this is that in the principal box formulation, a given conformation with small volume can either be semicollapsed, or can be largely extended and highly anisotropic. These two scenarios have very different numbers of contacts. However, parts a and b of Figure 10, which plot the slope of the energy vs reciprocal volume best-fit line for variable cutoff distance r_c, show that *all* volume measures lead to scalings which deviate from the predictions of mean-field theory. The deviations from mean field theory for the principal box volume and R_{gy} tensor ellipsoid volume are, in fact, the smallest.

The data (with error bars) in Figure 9a is taken from the pivot algorithm conformational sample, so it does not account for Boltzmann reweighting of the ensemble by the contact energies themselves. The mean number of contacts in the ensemble accounting for contact energies is

$$\langle n\rangle_\varepsilon = \frac{\sum_i n_i e^{-\beta(E_o + E_i(n))}}{\sum_i e^{-\beta(E_o + E_i(n))}} = \frac{\langle n e^{-\beta E(n)}\rangle_o}{\langle e^{-\beta E(n)}\rangle_o} \tag{15}$$

where $\beta = 1/k_BT$, n_i is the number of contacts in configuration i, E_o is the hard-wall potential accounting for self-avoidance, $E(n) =$

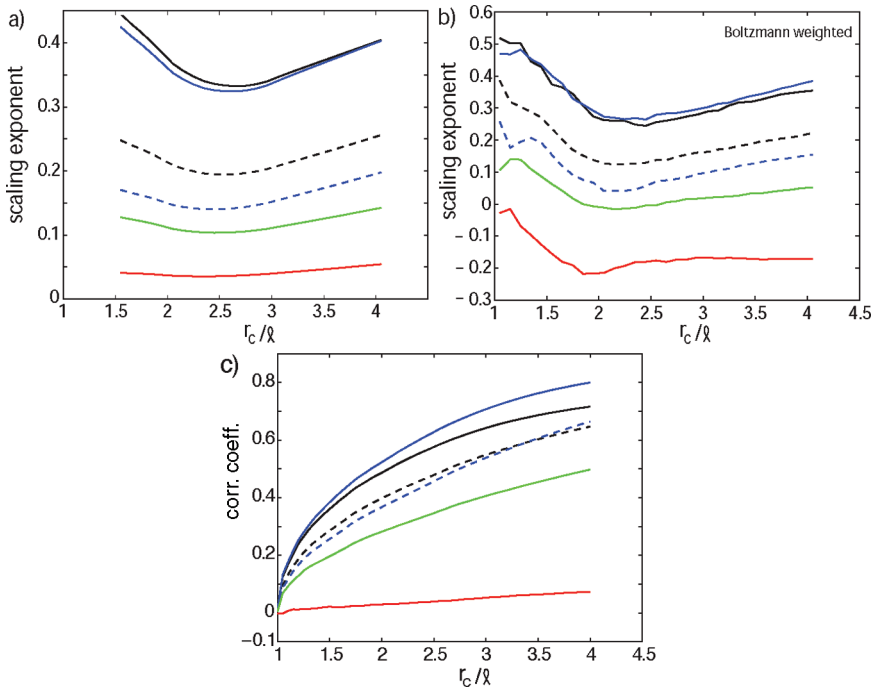


Figure 10. (a) Slope of the best fit line to the scatter data of the number of contacts vs the reciprocal volume V^{-1}, for all volume measures in Figure 3. The color code for panels a, b, and c is as follows: *solid black*, principal box volume; *dashed black*, unaligned box volume; *solid blue*, radius of gyration tensor ellipsoid volume; *dashed blue*, radius of gyration scalar volume; *solid green*, embedding sphere volume; *solid red*, end-to-end distance (Flory) volume. The principal volume and radius of gyration tensor ellipsoid have the largest scaling exponents, yet still show substantial deviation from mean-field behavior. (b) Scaling exponent of mean Energy with volume, for the Boltzmann-reweighted ensemble. Note the scale on the ordinate is larger for this panel: the scaling power law for the Flory volume and embedding sphere is reduced, but the scaling power law for the principal box and R_{gy} tensor ellipsoid are not significantly changed. (c) Correlation coefficient between the mean energy and reciprocal volume as a function of contact cutoff-length r_c, for the $\beta = 0$ ensemble. As r_c increases, the correlation indeed increases, however the scaling exponent never reaches mean-field theory (panels a, b).

$-n|\varepsilon|$ is the added energy of a conformation when n (attractive) contacts are accounted for, $\langle\cdots\rangle_\varepsilon$ indicates an average over the ensemble accounting for contact energetics, and $\langle\cdots\rangle_o$ indicates an average over the unperturbed SAW ensemble, as sampled by the pivot algorithm for example.

The contact data sampled directly from the pivot algorithm thus corresponds to the ensemble at $T = \infty$. Lowering the temperature so that $k_BT/\varepsilon = 5$ gives the bin-averaged data in Figure 9a indicated by blue-squares. Directly fitting this data gives an even shallower slope of 0.25. The data also do not follow a simple linear trend. Picking a subset of the ensemble with the largest 2/3 of the volume gives an exponent of 0.35 (dashed red line in Figure 9a); no particular range of volume shows an exponent near unity. As well, the data at smallest volume do not show the largest difference in energy between the $T = \infty$ and finite-T ensembles, which might be expected if these configurations had the largest number of contacts. As mentioned above, this is likely due to the large anisotropy present in small-volume conformations using the principal box metric. However, the data sets at the smallest values of volume do show the largest deviations from the mean field theory slope, and these ensembles will contain the most collapsed configurations, so they may most accurately represent the configurations and corresponding scaling relation that result when attractive energies are actually present in the conformational sampling.

To directly test the resulting scaling when contact energies are present, we simulated an $N = 100$ polymer chain with explicit self-interactions using the DMD method described above. Nonbonded monomers i and j with $|i - j| \geq 3$ and within $r_c = 2.5 \times$ the link length are given an attractive energy such that $\beta\varepsilon = -0.2$. The mean energy as a function of V^{-1} is given in Figure 9b. Like the infinite temperature ensemble above, the ensemble with mean energy at $\beta\varepsilon = -0.2$ also shows anomalous scaling $E \propto V^{-\alpha}$ with $\alpha = 0.2$. However the volumes sampled are smaller than those in the SAW: the polymer is semicollapsed, as indicated by the representative polymer configuration shown in the inset of Figure 9b. The range of energies and volumes sampled is correspondingly reduced. The scaling exponent is comparable to the slope in the small-volume regime of the pivot algorithm simulations, and as was the case there, also indicates significant deviations from mean field theory.

The above analysis was applied to a system with cutoff distance $r_c = 2.5\ell$ (where ℓ is the link length). Other values of r_c give different scaling exponents, as obtained from the slope of the best fit line to the number of contacts vs V^{-1} in the pivot simulations of Figure 9a. Figure 10a plots this slope for all of the volume measures in Figure 3. The principal box volume (solid black curve) and radius of gyration (solid blue curve) volume measures have the largest values of the slope; the end-to-end distance volume (solid red) has the smallest and is also the most weakly correlated as described below. Interestingly the slope is non-monotonic *vs* r_c, with a minimum around $r_c = 2.5\ell$. The general trend is not significantly modified by the above-described recipe of grouping states by volume and Boltzmann reweighting the states by energy, as shown in Figure 10b.

As r_c continues to increase past $\approx 7\ell$ for the 100mer, more and more configurations begin to have the maximal number of contacts, $(N - \Delta_{ij} + 1)(N - \Delta_{ij})/2$, where $\Delta_{ij} = |i - j|_{min}$,

dx.doi.org/10.1021/ma200454e |*Macromolecules* 2011, 44, 6182–6197

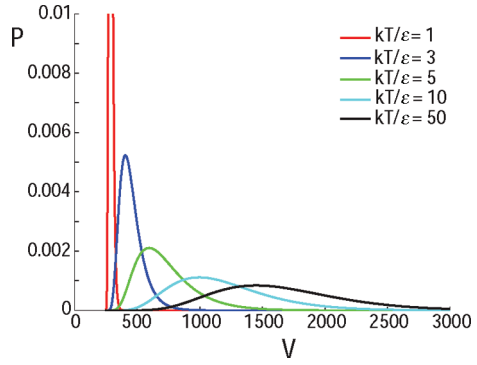


Figure 11. Probability distributions defined through the free energy in eq 17, for several temperatures as shown in the legend, for the 100-mer. At low temperature the system is in a collapsed phase, while at very high temperature the distribution converges to the conformational entropy distribution function in Figure 8b.

regardless of the volume of the configuration. This progressively decreases the correlation, and begins to decrease the slope, so that the slope never converges to unity for any r_c.

On the other hand, as the cutoff distance between contacts increases, more contacts are counted for a given configuration, and the mean-field approximation would be expected to improve. Figure 10c plots the correlation coefficient between the energy (number of contacts) and r_c for the all of the volume measures in Figure 3. Interestingly the end-to-end distance (red curve) shows very little correlation between energy and volume, while the embedding sphere (green curve) shows only modest correlation. The radius of gyration (dashed blue) and unaligned Cartesian box (dashed black) show stronger correlation, and the principal box volume (black) and radius of gyration tensor ellipsoid volume (blue) show the strongest correlation. One possible reason that the tensor ellipsoid correlates better than the principal box is that the contact density may be better represented inside the tensor ellipsoid than outside of it but still inside of the principal box, because relative fluctuations in contact-density configuration to configuration, are larger in the diffuse "halo" outside of the tensor ellipsoid (see also Figure 4 inset b).

Analogous to the treatment of the free energy of a polymer as a function of density or packing fraction, we can use the volume V as an order parameter measuring degree of collapse. The mean-field free energy $F(V,T)$ of a homopolymer as a function of the polymer volume V at a given temperature is given by $E(V) - TS(V)$, where the entropy $S(V)$ is given in eq 11. In our model, the internal energy of the polymer $E(V)$ is given by the mean number of contacts times the energy per contact, thus:

$$\begin{aligned} F(V,T) &= -\varepsilon\langle n(r_c,V)\rangle - TS(V) \\ &= -\varepsilon\frac{C}{V^{\alpha}} - k_B T \ln P(V) - TS_o \end{aligned} \tag{16}$$

Here the quantity $\langle n(r_c,V)\rangle$ is the average number of residue pairs within a cutoff distance r_c, given the polymer takes up a volume V, which was plotted in Figure 9a. Even though there is significant scatter between conformations, the mean energy correlates well with $V^{-\alpha}$. The exponent α is dependent on r_c. We take $r_c \approx 2$, which corresponds to $\alpha \approx 0.33$ in the Boltzmann reweighted pivot simulations in Figure 10b. Different exponents do not change the qualitative picture of collapse but will affect the collapse temperature. We let α be fixed with respect to contact energy even though in principle it does vary. The constant C is the hypothetical energy when the volume is extrapolated to unity, which can be determined from the fits analogous to that in Figure 9a for $r_c = 2$, as $C \approx 1320$. This gives an energy of $\varepsilon C/V_o^{\alpha} \approx 215\varepsilon$ in the fully collapsed state.

Up to volume-independent constants, the free energy can then be written using eq 10 as

$$F(V,T) = -\varepsilon\frac{C}{V^{\alpha}} + k_B T\left[A'\left(\frac{V-V_o}{\delta V}\right)^{\delta'} - \theta' \ln\left(\frac{V-V_o}{\delta V}\right)\right] \tag{17}$$

The probability distribution of volume V at temperature T is then given by $P \sim \exp(-\beta F(V,T))$, with free energy given in eq 17. Distributions for several temperatures are shown in Figure 11. The distributions show a second-order-like transition to a collapsed globule phase at low temperatures, consistent with previous analyses.[78,83−90] At high temperature, the distribution converges to that shown in Figure 8b.

DISCUSSION AND CONCLUSIONS

Motivated by the effects of crowding particles on the characteristic volume of a polymer or protein, as well as the opportunity to test mean-field theories by enumerating disordered ensembles, we proposed several candidate measures of polymer volume. We compared traditional measures of polymer volume, such as the Flory volume, and volume corresponding to the radius of gyration or radius of gyration tensor ellipsoid, to other less-common measures. These measures included the embedding sphere volume which is closely related to the miniball volume,[29,69] the Cartesian box volume[21] which is that of a box oriented in a fixed "lab" frame of reference that encloses all the monomers in the polymer, and a principal box measure,[22] which is the volume of the smallest box oriented along the principal axes of the polymer defined by the radius of gyration tensor, and which encloses all the monomers in the polymer. The principal box is closely related to minimal spanning boxes that enclose the polymer.[22,29,37]

We generated off-lattice self-avoiding random walks (SAWs) by using both Lal's pivot algorithm[38,51] and using discrete molecular dynamics simulations. We first showed that the exponents appearing in the des Cloizeaux functional form of the end-to-end distance probability distribution for the off-lattice SAW had significant finite-size modifications in the regime of small globular protein chain lengths, but converged to the well-known on-lattice exponents in the limit of large chain lengths, as expected. However, chain lengths corresponding to several hundred persistence lengths of a real polymer are needed to see such a crossover.

The lack of correlation between the either the radius of gyration tensor ellipsoid or the principal box and other measures analyzed indicates the importance of anisotropy in characterizing SAWs. Moreover, the radius of gyration tensor ellipsoid excludes a significant fraction of polymer, approximately 44% for a 100mer SAW. This might be expected based on the arguments made in the introduction. One solution may be a simple rescaling of the ellipsoid, but to systematically find the appropriate scale factor might involve a best fit to the overall density profile. A simpler rescaling was used for the radius of gyration scalar sphere to determine the depth of ion-pairs in proteins,[91] by taking the

radius of the protein as that of a sphere with the same radius of gyration, or $(5/2)^{1/2} r_{gy}$.

The probability distribution of the end-to-end distance was much broader than all other volume measures considered, even the embedding sphere diameter. The distribution of the mean principal box edge length was the most sharply peaked, slightly sharper than the mean diameter of the R_{gy} tensor ellipsoid. This supports the notion that the principal box can be thought of as giving "the characteristic volume" of a polymer without reference to a specific conformation, although nearly any such measure of the SAW ensemble must be distributed. As an accurate characteristic volume, the principal box may be useful as an order parameter governing coil to globule collapse as solvent conditions are varied. For example the principal box volume should be much more useful than say the radius of gyration or end-to-end distance in capturing osmolyte or crowding induced collapse, or folding in crowded milieu representing the interior of the cell. Perhaps even more useful in this regard may be minimal convex spanning hull measures that enclose the polymer such as *qhull*,[92] which algorithmically constructs a spatial polyhedron that "shrink-wraps" all of the monomers.

The radius of gyration tensor is however more amenable to analytic treatment than the principal box, which has led historically to the analysis of asphericity to capture polymer anisotropy.[16−19] In spite of this, we can still introduce an analogous measure to the asphericity for the principal box, which we called the acubicity. The acubicity satisfies the same inequality relationships as the asphericity, but shows higher anisotropy. We also saw that the acubicity involves averages of box edge lengths squared, which was not necessary to capture the anisotropy and in fact reduces the sensitivity of acubicity as a measure to detect anisotropy. Using a measure of acubicity that scaled linearly in the box edge lengths enhanced the capability of the measure to capture polymer anisotropy.

For energy functions that give larger Boltzmann weight to anisotropic conformations, the anisotropic volume measures will become increasingly important. As an example, a polyelectrolyte that nevertheless folds (by counterion condensation) will have unfolded configurations that are significantly elongated due to charge−charge repulsion, as in the inset to Figure 5a, until the system folds or condenses. For such polymers, isotropic volume measures such as the radius of gyration could lead to large errors in characterizing collapse.

It is nontrivial to find a distribution function for the principal box volume analytically. Instead we found that the distribution could be well-fit heuristically by a des Cloizeaux function with modified exponents that were essentially N-independent over the range of polymer sizes used in this study. A correction to the des Cloizeaux function that accounted for a nonzero collapsed polymer volume was inconsequential for the SAW ensemble, but it becomes relevant in describing coil to globule collapse.

Random mixing approximations for the mean-field energy of an isolated polymer utilize the packing fraction of polymer configurations. However a problem with packing fraction is that a characteristic volume must be defined to use it. Put another way, in the mean-field approximation, the thermal energy per volume of a self-interacting polymer will scale as the square of the concentration of monomers. However the concentration of monomers is itself determined by the effective volume characterizing the polymer. Once a volume measure is introduced, the mean-field approximation can then be directly tested from a generated ensemble of conformations. Our tests showed that for no measure of volume was the mean-field approximation satisfied. In particular, the energy showed essentially no correlation with reciprocal volume when the Flory volume was used, indicating that the Flory volume would be a poor order parameter to characterize mean energies in a polymer system.

Instead we found anomalous exponents for the scaling of average energy with volume. The nonmean-field exponents were further investigated using direct simulations of a self-attractive homopolymer, which confirmed the anomalous scaling of energy with volume. One possible reason for the anomalous scaling is the nontrivial density distribution for SAWs in the principal axis frame,[22,23,90,93,94] so that a more accurate measure of the mean energy would involve generalizing the packing fraction, which is essentially the mean monomer density, to a local quantity $\eta \rightarrow \eta(\mathbf{r})\ \mathrm{d}^3\mathbf{r}$. The mean energy would then be given by $Nz\varepsilon \int \mathrm{d}^3\mathbf{r}\ \eta(\mathbf{r})$. However, one finding from studies of the density distribution is that the density profile is nearly flat along the major principal axis for a SAW, which would seem to support the mean-field approximation. Perhaps a more likely explanation for the anomalous scaling involves correlations in the density distribution, at least at the pairwise level. Analysis of the pair correlation function $g_n(\mathbf{r}|\mathbf{r}_o)$, giving essentially the probability density for a monomer to be at position $\mathbf{r}$, given it is separated by n residues along the chain from a residue at position $\mathbf{r}_o$, may allow for a treatment of fluctuation-dominated energy density. The role of density fluctuations in governing contact probabilities is an interesting topic of future research.

The density isocontours may provide another way to define effective polymer volume for a SAW ensemble through a density isocontour. A similar strategy was used to derive the volume and characterize the shape of small molecule candidate pharmacophores through the electron density.[95]

The accuracy of terms in a free energy function that scale as $\sim c^2$ is likely contingent on dilute monomer concentrations and thus small packing fraction $\eta \ll 1$. The origins of the c^2-dependence in mean-field theories of polymer solutions dates back to Flory−Huggins theory,[7,75−77] which was itself based on observations of corrections to ideal gas behavior of the osmotic pressure in dilute solutions of whole polymers.[96] The extension of the mixing approximation from dilute polymer solutions to effective monomer concentration in a single isolated polymer, even for a density distribution as "dilute" as a self-avoiding random walk, may simply never be quantitatively valid.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information. Measures of polymer volume and their properties. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: steve@physics.ubc.ca.

Author Contributions

S.H. generated plots, analysis, and preliminary text. A.L. provided DMD code and analysis. S.S.P. generated the data, figures, and analysis used in the final manuscript and wrote and revised the manuscript.

■ ACKNOWLEDGMENT

S.S.P. acknowledges funding from the Canada Research Chairs program, the A.P. Sloan Foundation, and the Natural Sciences and Engineering Research Council, as well as the use of computing resources provided by WestGrid and Compute/Calcul Canada.

■ REFERENCES

(1) Saunders, A.; Davis-Searles, P. R.; Allen, D. L.; Pielak, G. J.; Erie, D. A. *Biopolymers* **2000**, *53*, 293–307.
(2) Schellman, J. A. *Biophys. J.* **2003**, *85*, 108–125.
(3) Cheung, M. S.; Klimov, D.; Thirumalai, D. *Proc. Natl. Acad. Sci.* **2005**, *102*, 4753–4758.
(4) Zhou, H.-X.; Rivas, G.; Minton, A. P. *Annu. Rev. Biophys.* **2008**, *37*, 375–397.
(5) Elcock, A. H. *Curr. Opin. Struct. Biol.* **2010**, *20*, 196–206.
(6) Linhananta, A.; Hadizadeh, S.; Plotkin, S. S. *Biophys. J.* **2011**, *100*, 459–468.
(7) Flory, P. J. *Principles of Polymer Chemistry*; Cornell University Press: Ithaca, NY, 1953.
(8) Flory, P. J. *Statistical mechanics of chain molecules*; Wiley Interscience Publishers: New York, 1969.
(9) de Gennes, P.-G. *Scaling concepts in polymer physics*; Cornell University Press: Ithaca, NY, 1979.
(10) Kuhn, W. *Colloid Z.* **1934**, *68*, 2–15.
(11) des Cloizeaux, J. *Phys. Rev. A* **1974**, *10*, 1665–1669.
(12) des Cloizeaux, J. *J. Phys. (Paris)* **1981**, *42*, 635–652.
(13) Lhuillier, D. *J. Phys. (Paris)* **1988**, *49*, 705–710.
(14) Šolc, K.; Stockmayer, W. H. *J. Chem. Phys.* **1971**, *54*, 2756–2757.
(15) Šolc, K. *J. Chem. Phys.* **1971**, *55*, 335–344.
(16) Aronovitz, J. A.; Nelson, D. R. *J. Phys. (Paris)* **1986**, *47*, 1445–1456.
(17) Rudnick, J.; Gaspari, G. *Science* **1987**, *237*, 384–389.
(18) Rudnick, J.; Beldjenna, A.; Gaspari, G. *J. Phys. A: Math. Gen.* **1987**, *20*, 971–984.
(19) Diehl, H. W.; Eisenriegler, E. *J. Phys. A: Math. Gen.* **1989**, *22*, L87–L91.
(20) Kranbuehl, D. E.; Verdier, P. H. *J. Chem. Phys.* **1977**, *67*, 361–365.
(21) Rubin, R. J.; Mazur, J. *J. Chem. Phys.* **1975**, *63*, 5362–5374.
(22) Rubin, R. J.; Mazur, J. *Macromolecules* **1977**, *10*, 139–149.
(23) Theodorou, D. N.; Suter, U. W. *Macromolecules* **1985**, *18*, 1206–1214.
(24) Cannon, J. W.; Aronovitz, J. A.; Goldbart, P. *J. Phys. I* **1991**, *1*, 629–645.
(25) Honeycutt, J. D.; Thirumalai, D. *J. Chem. Phys.* **1989**, *90*, 4542–4559.
(26) Deutsch, J. M. *Phys. Rev. E* **1999**, *59*, R2539–R2541.
(27) Grosberg, A. Y. *Phys. Rev. Lett.* **2000**, *85*, 3858–3861.
(28) Alim, K.; Frey, E. *Phys. Rev. Lett.* **2007**, *99*, 198102.
(29) Rawdon, E.; Dobay, A.; Kern, J. C.; Millett, K. C.; Piatek, M.; Plunkett, P.; Stasiak, A. *Macromolecules* **2008**, *41*, 4444–4451.
(30) Rawdon, E. J.; Kern, J. C.; Piatek, M.; Plunkett, P.; Stasiak, A.; Millett, K. C. *Macromolecules* **2008**, *41*, 8281–8287.
(31) Goldenberg, D. P. *J. Mol. Biol.* **2003**, *326*, 1615–1633.
(32) Minton, A. P. *Biophys. J.* **2005**, *88*, 971–985.
(33) Tran, H. T.; Pappu, R. V. *Biophys. J.* **2006**, *91*, 1868–1886.
(34) Vitalis, A.; Wang, X.; Pappu, R. V. *Biophys. J.* **2007**, *93*, 1923–1937.
(35) Foteinopoulou, K.; Karayiannis, N. C.; Laso, M.; Kröger, M.; Mansfield, M. L. *Phys. Rev. Lett.* **2008**, *101*, 265702.
(36) Karayiannis, N. C.; Foteinopoulou, K.; Laso, M. *J. Chem. Phys.* **2009**, *130*, 074704.
(37) Millett, K. C.; Rawdon, E. J. *J. Comput. Phys.* **2003**, *186*, 426–456.
(38) Madras, N.; Sokal, A. D. *J. Stat. Phys.* **1988**, *50*, 109–186.
(39) Prellberg, T. *J. Phys. A: Math. Gen.* **2001**, *34*, L599–602.
(40) Yamakawa, H. *Modern Theory of Polymer Solutions*; Harper and Row: New York, 1971.
(41) Teramoto, E.; Kurata, K.; Chujo, R.; Suzuki, C.; Tani, K.; Kajikawa, T. *J. Phys. Soc. Jpn.* **1955**, *10*, 953–959.
(42) Verdier, P. H.; Stockmayer, W. H. *J. Chem. Phys.* **1962**, *36*, 227–235.
(43) Schatzki, T. F. *J. Polym. Sci.* **1962**, *57*, 337–356.
(44) Wall, F. T.; Windwer, S.; Gans, P. J. *J. Chem. Phys.* **1963**, *38*, 2220–2227.
(45) McKenzie, D. S. *J. Phys. A: Math., Nucl. Gen.* **1973**, *6*, 338–352.
(46) Bishop, M.; Clarke, J. H. R.; Rey, A.; Freire, J. J. *J. Chem. Phys.* **1991**, *95*, 4589–4592.
(47) Caracciolo, S.; Causo, M. S.; Pelissetto, A. *J. Chem. Phys.* **2000**, *112*, 7693–7710.
(48) Fisher, M. E. *J. Chem. Phys.* **1966**, *44*, 616–622.
(49) McKenzie, D. S.; Moore, M. A. *J. Phys. A: Gen. Phys.* **1971**, *4*, L82–L86.
(50) Barrett, A. J.; Mansfield, M.; Benesch, B. C. *Macromolecules* **1991**, *24*, 1615–1621.
(51) Lal, M. *Mol. Phys.* **1969**, *17*, 57–64.
(52) Kremer, K.; Binder, K. *Phys. Rep.* **1988**, *7*, 259–319.
(53) Grassberger, P. *Phys. Rev. E* **1997**, *56*, 3682–3693.
(54) Rosenbluth, M. N.; Rosenbluth, A. W. *J. Chem. Phys.* **1995**, *23*, 356–359.
(55) Dayantis, J.; Sturm, J. *Polymer* **1985**, *26*, 1631–1637.
(56) Zhou, Y.; Karplus, M. *J. Mol. Biol.* **1999**, *293*, 917–951.
(57) Linhananta, A.; Boer, J.; MacKay, I. *J. Chem. Phys.* **2005**, *122*, 114901.
(58) Zhou, Y.; Zhang, C.; Stell, G.; Wang, J. *J. Am. Chem. Soc.* **2003**, *125*, 6300–6305.
(59) Borreguero, J. M.; Dokholyan, N. V.; Buldyrev, S. V.; Shakhnovich, E. I.; Stanley, H. E. *J. Mol. Biol.* **2002**, *318*, 863–876.
(60) Ding, F.; Tsao, D.; Nie, H.; Dokholyan, N. V. *Structure* **2008**, *16*, 1010–1018.
(61) Jang, H.; Hall, C. K.; Zhou, Y. *Biophys. J.* **2004**, *86*, 31–49.
(62) Ding, F.; LaRocque, J. J.; Dokholyan, N. V. *J. Biol. Chem.* **2005**, *280*, 40235–40240.
(63) Seaton, D. T.; Mitchell, S. J.; Landau, D. P. *Braz. J. Phys.* **2008**, *38*, 48–53.
(64) Wei, G.; Eichinger, B. E. *Macromolecules* **1990**, *23*, 4845–4855.
(65) Zifferer, G.; Olaj, O. F. *J. Chem. Phys.* **1994**, *100*, 636–639.
(66) Haber, C.; Ruiz, S. A.; Wirtz, D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10792–10795.
(67) Luo, M.; Huang, J.; Chen, Y.; Xu, J. *Eur. Polym. J.* **2001**, *37*, 1587–1590.
(68) Eurich, F.; Maass, P. *J. Chem. Phys.* **2001**, *114*, 7655–7668.
(69) Fischer, K.; Gartner, B. The smallest enclosing ball of balls: combinatorial structure and algorithms. In *Proceedings of the nineteenth annual symposium on Computational geometry; SCG '03*; ACM: New York, 2003; pp 292−301.
(70) Balaeff, A. Copyright 2001, The Theoretical Biophysics Group, Beckman Institute, and The Board of Trustees of the University of Illinois, http://www.ks.uiuc.edu/Development/MDTools/molvolume.
(71) Victor, J. M.; Lhuillier, D. *J. Chem. Phys.* **1990**, *92*, 1362–1364.
(72) Bishop, M.; Saltiel, C. J. *J. Chem. Phys.* **1991**, *95*, 606–607.
(73) Values of the relative deviation in Table 3 are in fact accurate to at least a part in 10^3, with more accurate values for the relative deviation for $N = 100$ being 0.348(1) for V_{ell} and 0.3123(5) for V_{pb}. The number in parenthesis indicates the uncertainty in the last digit. For larger N the discrepancy only increases, e.g. for $N = 200$ the corresponding values for the relative deviation are 0.376(3) for V_{ell} and 0.328(4) for V_{pb}.
(74) Bishop, M.; Saltiel, C. J. *J. Chem. Phys.* **1988**, *88*, 3976–3980.
(75) Flory, P. J. *J. Chem. Phys.* **1942**, *10*, 51–61.
(76) Huggins, M. L. *J. Phys. Chem.* **1942**, *46*, 151–158.
(77) Huggins, M. L. *J. Am. Chem. Soc.* **1942**, *64*, 1712–1719.
(78) Bryngelson, J. D.; Wolynes, P. G. *Biopolymers* **1990**, *30*, 177–188.

(79) Chan, H. S.; Dill, K. A. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 447–490.
(80) Plotkin, S. S.; Wang, J.; Wolynes, P. G. *J. Chem. Phys.* **1997**, *106*, 2932–2948.
(81) Plotkin, S. S.; Onuchic, J. N. *Q. Rev. Biophys.* **2002**, *35*, 205–286.
(82) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
(83) deGennes, P. G. *J Phys.-Lett.* **1975**, *36*, L-55–L-57.
(84) deGennes, P. G. *J Phys.-Lett.* **1978**, *39*, L299–L301.
(85) Lifshitz, I. M.; Grosberg, A. Y.; Khokhlov, A. R. *Rev. Mod. Phys.* **1978**, *50*, 683–713.
(86) Grosberg, A. Y.; Kuznetsov, D. V. *Macromolecules* **1992**, *25*, 1996–2003.
(87) Grosberg, A. Y.; Khokhlov, A. R. *Statistical physics of macromolecules*; AIP series in polymers and complex materials; AIP Press: New York, 1994.
(88) Zhou, Y.; Hall, C.; Karplus, M. *Phys. Rev. Lett.* **1996**, *77*, 2822–2825.
(89) Noguchi, H.; Yoshikawa, K. *J. Chem. Phys.* **1998**, *109*, 5070–5077.
(90) Rissanou, A. N.; Anastasiadis, S. H.; Bitsanis, I. A. *J. Polym. Sci., Part B: Polym. Phys.* **2006**, *44*, 3651–3666.
(91) Guest, W.; Cashman, N. R.; Plotkin, S. S. *A Phys. Chem. Chem. Phys.* **2011**, 13, 6286−6295.
(92) Barber, C.; Dobkin, D.; Huhdanpaa, H. *ACM Trans. Math. Software* **1996**, *22*, 469–483, http://www.qhull.org.
(93) Janszen, H. W. H. M.; Tervoort, T. A.; Cifra, P. *Macromolecules* **1996**, *29*, 5678–5687.
(94) Millett, K. C.; Rawdon, E. J.; Tran, V. T.; Stasiak, A. *J. Chem. Phys.* **2010**, *133*, 154113.
(95) Zyrianov, Y. *J. Chem. Inf. Model.* **2005**, *45*, 657–672.
(96) Staudinger, H.; Schulz, G. V. *Ber. Dtsch. Chem. Ges. (A and B Ser.)* **1935**, *68*, 2320–2335.

Supporting material for:

Improved measures for the shape of a disordered polymer to test a mean-field theory of collapse

Shirin Hadizadeh,[†] Apichart Linhananta,[‡] and Steven S. Plotkin[*,†]

Department of Physics and Astronomy, University of British Columbia, Vancouver, Canada, and Physics Department, Lakehead University, Thunder Bay, Ontario

E-mail: steve@physics.ubc.ca

Measures of polymer volume and their properties

As a consistency check on the volume measures described in the main text, Figure 1a shows a scatter plot of the end-to-end radius R_{ete}, defined as the end-to-end distance over 2, and the radius R_{sphere} of the smallest sphere which circumscribes the polymer configuration (the embedding sphere). R_{sphere} is always larger than R_{ete}, since the distance between any pair of residues must always be less than the embedding sphere's diameter.

As described in the main text, the volume of the effective ellipsoid V_{ell} characterizing the polymer configuration is given by $4\pi\sqrt{3}\prod_{i=1}^{3}\sqrt{\lambda_i}$, where λ_i are the eigenvalues of the radius of gyration tensor. The effective radius of a sphere with the same volume as that of the ellipsoid (V_{ell}) is given by the geometrical mean of the individual radii: $R_{gy}^{eff} = (abc)^{1/3} = \sqrt{3}\prod_{i=1}^{3}\lambda_i^{1/6}$. This is to be compared with the volume of the effective sphere defined by $(4/3)\pi R_{gy}^3$, with $R_{gy} = \sqrt{\sum_{i=1}^{3}\lambda_i}$. A scatter plot of the radii determined by both of the above measures is shown in Figure 1b.

[*]To whom correspondence should be addressed
[†]University of British Columbia
[‡]Lakehead University

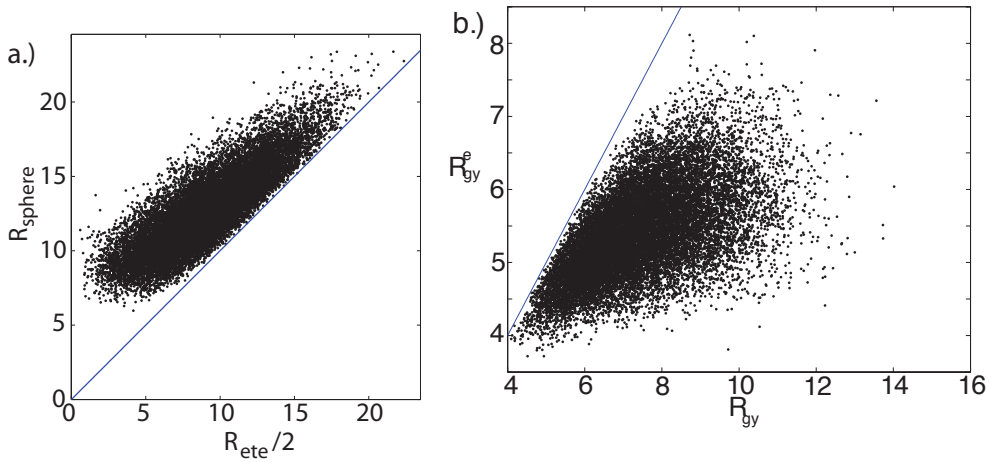


Figure 1: (a) Scatter plot of the end-to-end radius, defined as the end-to-end distance R_{ete} over 2, and the radius R_{sphere} of the smallest sphere which circumscribes the polymer configuration. R_{sphere} is always larger than $R_{ete}/2$, since the distance between any pair of residues must always be less than the embedding sphere's diameter: the thin blue line indicates $y = x$. (b) Scatter plot of the radius of gyration R_{gy} for a 100mer given by the trace of the radius of gyration tensor, and effective radius R_{gy}^{eff} defined above through the volume of the radius of gyration tensor ellipsoid V_{ell}. The data correlate ($r = 0.58$) but with significant scatter. The radius of gyration is always larger than the effective radius: the thin blue line indicates $y = x$. All distances are in units of the link length on the polymer chain. Approximately 2×10^5 configurations are shown.