Intrinsic Shape Stability of Equilibrium Motions in Poly(L-alanine)

Gustavo A. Arteca

Département de Chimie et Biochimie, Laurentian University, Ramsey Lake Road, Sudbury, Ontario, Canada P3E 2C6

Received May 8, 1996; Revised Manuscript Received August 8, 1996[⊗]

ABSTRACT: Under some conditions, an isolated macromolecular chain at a minimum energy configuration may be subject to either local or large-scale vibrational deformations that can preserve the *global molecular shape features of the given minimum*. We refer to this persistence of dynamical shape as *intrinsic shape stability*. Assessing shape stability is essential for studies of molecular recognition and binding. In this work, we present a novel approach to analyzing and quantifying shape stability. As an illustrative example, we apply the method in a quantitative study of stability in poly(L-alanine) helices of variable lengths. Our approach explores the interrelation between various *independent aspects* of macromolecular shape, namely, *size*, *compactness*, *anisometry*, and *chain entanglements* (*or degree of folding*). We associate a different shape descriptor to each aspect. In this work, we analyze the role of the chain's length on the fluctuations in shape descriptors. We show that the study of entanglements and anisometry can indicate the occurrence of conformational rearrangements that would otherwise remain hidden when monitoring only the radius of gyration. We estimate the *critical number of amino acid residues* for which a polypeptide, undergoing equilibrium motions, better conserves its helical shape for a given set of environmental conditions. The methodology can be applied to general configurational changes in other types of polymers.

Introduction

Consider a polypeptide with n amino acid residues, in an α-helical (*regular*) conformation. At room temperature and given environmental conditions, *short* chains unfold rapidly since the helix is not their global minimum.[1,2] Similarly, long chains also bend away from a regular helical conformation and produce coiled coils.[3] In this context, a relevant question is: For a given set of external conditions, what is the critical number of residues, n_c, of the polypeptide that ensures the maximum stability of the original *helical shape*? This is a particular case of the more general question relevant to polymer shape: what are the conditions (size, composition, temperature, solvent) under which a macromolecule *conserves the shape features* present at a conformational minimum? What is the time scale of the molecular vibrations which maintain these shape features? In the present work, I discuss and apply an approach where *all aspects of molecular shape* are used in the evaluation of *intrinsic* polypeptide shape stability. I consider a condition where external factors (temperature and solvation) are constant and the only variable is the number of residues.

It is important to understand that we deal with a definition of *geometrical and topological shape*; that is, we analyze features that may remain invariant during the motion and energy changes experienced by a molecular chain. Structural stability[4] is a generic concept used to characterize a system which conserves intrinsic properties when subject to small perturbations. In structurally unstable systems, a smooth change in control parameters (*e.g.*, the nuclear positions) causes a discontinuous change in the system's properties.[5] An analysis of structural stability, however, depends on which are the properties we consider important to define the notion of "structure". Curvature properties in electron density[6] and potential energy[7] provide criteria for establishing structural stability. In the present work, I deal with *stability* defined in terms of *global molecular shape features*. To distinguish from energetic (thermodynamic) and configurational (structural) stability, I shall reserve the term *shape stability*[8] for the conservation of molecular shape despite fluctuations in nuclear configurations, potential energy, and electron density.

A great deal of work has been devoted to the analysis of energetics and dynamics in two-state transitions in proteins, *e.g.*, helix/coil and native/unfolded transitions.[9−15] In the present work, I deal with a particular aspect of macromolecular flexibility: the response of the molecular shape of *regular polypeptide conformations* to deformations about the equilibrium positions. This is a better defined problem, which can serve to test the ability of a method to distinguish changes in molecular shape under small distortions.

The best studied cases of regular conformations are the secondary structures of polypeptides, in particular the α-helices. Analyses of stability on α-helices have been performed from different viewpoints. Computer experiments have probed their energetic and structural stability.[16−19] [Structural properties usually followed are, for instance, the fluctuations in internal hydrogen bonds.] The effects of solvation[20−24] and amino acid replacement[25−27] on helix stability are also known. In addition to classical rules for the tendencies of amino acids to occur in helices,[21,28,29] recent work has provided quantitative results on helix propensity in terms of polymer length[30,31] and composition.[32−35]

The above studies have focused on monitoring either the potential energy or some geometrical properties in order to establish degrees of stability. In this work, I deal with a measure of helical persistence that is based on global molecular shape features, instead of *ad hoc* local structural parameters. To this purpose, I probe the shape stability of a polypeptide conformation in terms of three independent global descriptors: *molecular size*,[36] *anisometry*,[37−39] and *entanglement complexity*.[40] Previous work in the literature has dealt with the *critical polymer length* required for the α-helix to become

[⊗] Abstract published in *Advance ACS Abstracts,* October 1, 1996.

S0024-9297(96)00686-9 CCC: $12.00

the global conformational minimum of a (short) polypeptide.[30,41−44] Here, I move one step further along this analysis. Our work presents a criterion to establish the *critical helical length* that maximizes (or *maintains*) the shape stability of the helix over possible equilibrium deformations. In order to evaluate the *intrinsic stability*, results are provided for a single isolated homopolymer chain at constant temperature. In order to test the single role of chain length, we work at ambient temperature conditions, where on would expect in principle the shape stability be maintained under most accessible motions. (By contrast, it is evident that shape stability will not be maintained at high temperature.) In the present work, I test whether all accessible motions at low temperature do indeed maintain the shape features of a minimum energy conformer.

The work is organized as follows. In the next sections, I present the methodology for shape characterization and the details of the molecular dynamics simulations of equilibrium motions. From the results, I discuss in detail the conditions for shape persistence, the value of the critical helical length, and an interpretation of the nature of the structural deformations that distort molecular shape in long chains. Finally, our results allow us to estimate the time scales for the motions which preserve the molecular shape of a desired conformer.

Theoretical Background on Shape Stability

A minimum energy conformation whose *shape* is invariant under small perturbations will be considered *stable*.[4] [Small perturbations on a single molecule can be caused by vibrations, an external bath, or external fields such as those produced by other neighboring molecules. Note that the vibrations refer to all those accessible at a given temperature. These vibrations may either be mostly localized at some bond or globally extended over the entire molecule.] We say that a molecule undergoes a *shape transition* whenever the values of its shape descriptors present a sharp change during a *configurational rearrangement*. An initial conformation that undergoes shape transitions when allowed to evolve freely is *structurally unstable*. Highly flexible molecules belong to this class.

Characterizing the *dynamical shape of a flexible molecule* involves averaging over the values of shape descriptors for accessible conformations. These averages can be derived by sampling configurational space, or by following the time evolution in phase space with molecular dynamics (MD) trajectories. By analyzing fluctuations in averaged shape descriptors, we can express the flexibility and structural stability of a given minimum energy conformation.[8]

The above notions can be rendered quantitatively. Let Γ be some descriptor associated with a molecular model (*e.g.*, a descriptor of entanglements in a chain). Let $\langle\Gamma\rangle$ and σ_Γ be the configurational average and the broadness of the distribution ("fluctuation") of $\Gamma(t)$, respectively. Using an MD trajectory with time span t_{max}, we have

$$\langle\Gamma\rangle \approx \frac{1}{t_{max}}\int_{t_0}^{t_0+t_{max}}\Gamma(t)\,dt \qquad (1)$$

$$\sigma_\Gamma = [\langle\Gamma^2\rangle - \langle\Gamma\rangle^2]^{1/2} \qquad (2)$$

where t_0 is the initial point for sampling (*e.g.*, after equilibration) and $\Gamma(t)$ the value of the shape descriptor for a configuration $K(t)$, found at time t. [The integral in eq 1 is typically evaluated from discrete sampling along the trajectory. Note that the standard deviation in the mean $\langle\Gamma\rangle$ will be smaller than σ_Γ.] Let Γ_0 be the value of the descriptor at an initial (minimum energy) configuration K_0. A criterion for shape stability can now be expressed in terms of $\langle\Gamma\rangle$, σ_Γ, and Γ_0. We can state that K_0 is *structurally stable* if $\Gamma(t)$ fluctuates close to Γ_0 over a long period of time, that is

$$\langle(\Gamma(t) - \Gamma_0)^2\rangle^{1/2} \le p\sigma_\Gamma \qquad (3)$$

where the constant p can be chosen according to a desired confidence level.

The inequality in eq 3 establishes an *equivalence relationship* for configurations based on molecular shape: if eq 3 is satisfied along the trajectory, then *all configurations K(t) are equivalent to K_0*.[8] Whenever eq 3 is not satisfied *over a short-time range*, there is an indication that shape transitions take place along the trajectory. With a proper choice of a descriptor Γ, $\langle\Gamma\rangle$ and σ_Γ can serve as order parameters to follow these transitions.

The *simultaneous* use of several independent descriptors Γ can provide a detailed characterization of all essential features associated with polymer shape. In our case, three properties (and their corresponding shape descriptors) are monitored: molecular size and compactness (using the radius of gyration), anisometry (using the asphericity), and complexity of entanglements or "degree of folding" (using the probabilities of overcrossings). Comparing fluctuations in such descriptors along MD trajectories provides a measurement of persistence of a given shape feature. This approach is applied in the analysis of shape stability of equilibrium motions in polypeptides of variable length. The motions studied are single-molecule oscillations, with coupling to a thermal bath and a pseudosolvent (continuum) dielectric screening. This approach allows us to estimate the intrinsic shape stability of a given structural feature for a constant setting of the environment. Finally, the critical length n_c for shape stability can be expressed by the number of residues which minimizes the shape fluctuations (measured by the standard deviations) associated with at least two independent descriptors Γ_1 and Γ_2.

Molecular Shape Descriptors of α-Helical Chains

Let $\{\mathbf{r}_i\}$ be the set of nuclear coordinates associated with a conformation of a molecule with n nuclei ($i = 1, 2, ..., n$). We shall choose the origin of the coordinate system *at the center of mass* of the molecular backbone. All shape descriptors we use here are derived from $\{\mathbf{r}_i\}$. In addition, we will take into account the bond connectivity in oder to derive a more detailed characterization of molecular shape.

1. Descriptor of Size and Compactness. The standard descriptor for a distribution of n nuclei in space is the *instantaneous radius of gyration* R_G. For *molecular backbones* with particles of the same mass (*e.g.*, α-*carbon backbones*), the radius of gyration (written in center-of-mass coordinates) is simply[36]

$$R_G^2 = \frac{1}{n}\sum_{i=1}^{n} r_i^2 \qquad (4)$$

The "proper" radius of gyration is the configurational average of R_G (*i.e.*, $\langle R_G\rangle$) and is a well-known measurable

property for polymers.[36] The value of R_G measures size and compactness. There are other alternative descriptors of compactness which incorporate explicitly the dependence on polymer composition by taking into account different excluded volume contributions.[45−48] For our present goal, the standard radius of gyration is sufficient.

2. Descriptor of Anisometry. A distinct aspect of molecular shape is the deviation of the nuclear position distribution from a spherical one. We refer to this as *anisometry*. Early approaches used the so-called "shape factors", derived from the principal moments of inertia, to characterize anisometry.[49,50] A more convenient single descriptor is the so-called *asphericity* Ω.[37−39] It is defined in terms of the three principal moments of inertia (calculated with the origin at the center of mass) $\{\lambda_i\}$:

$$\Omega = \frac{1}{2}\{\sum_{i=1}^{2}\sum_{j=i+1}^{3}(\lambda_i - \lambda_j)^2\}\{\sum_{i=1}^{3}\lambda_i\}^{-2} \qquad (5)$$

As a shape descriptor, Ω discriminates between geometrical shapes. For a spherical distribution, we find $\Omega = 0$. For *prolate* molecules (cigar-shaped with $\lambda_1 \approx \lambda_2 \gg \lambda_3$) we find $\Omega \approx {}^1/_4$, whereas *oblate* molecules (disk-shaped with $\lambda_1 \gg \lambda_2 \approx \lambda_3$) give $\Omega \approx 1$. It should be noticed that, in the case of an n α-carbon backbone, the moments of inertia are related to R_G as follows:

$$\sum_{i=1}^{3}\lambda_i = 2nR_G^2 \qquad (6)$$

3. Descriptors of Entanglement Complexity. Recently, we have introduced the *probability distribution of overcrossings* as an absolute, global shape descriptor of a macromolecular conformation.[40] This distribution, indicated as $\{A_N(n)\}$, gives the probability of observing N overcrossings by projecting into two dimensions the bonding pattern of a rigid n-atom macromolecular conformation. [The overcrossings are the "double points" where two bonds appear to "cross" when projected to a plane.] The projections considered are those of the backbone to planes tangent to a sphere whose radius is the span of the molecule (see below).[8,51] The distribution $\{A_N(n)\}$ can be defined for arbitrary architectures and not only for linear molecular chains.[52−54]

The computation of $\{A_N(n)\}$, as well its basic properties, is discussed in the literature.[40] Briefly, the algorithm for evaluating the probabilities is as follows:

(a) Determine the center-of-mass coordinates of the main-chain atoms defining the backbone and compute the span R. This is the radius of the smallest sphere, centered at the geometric center of the backbone, which encloses the backbone completely.

(b) Consider an arbitrary point **r** on the sphere with radius R and determine the plane tangent to the sphere at **r**.

(c) Project the backbone coordinates onto the plane in (b) and establish the number N of bond−bond crossings associated with this projection.

(d) Repeat steps (b) and (c) for a number m ($m \gg 1$) of *randomized* points **r** on the sphere. The probability A_N is computed as the ratio m_N/m, where m_N is the number of projections yielding N overcrossings.

The overcrossing probabilities for a molecule with at least three atoms are normalized as follows:

$$\sum_{N=0}^{max\ N} A_N(n) = 1, \qquad \forall n \geq 3 \qquad (7)$$

where the value of *max N* depends on the molecular architecture. For linear chains, it is $(n-2)(n-3)/2$. [Note that $n = 3$ corresponds to a planar configuration, where evidently no bonds will overcross with nonzero probability. In this case, we find $A_0(3) = 1$ and $A_N(3) = 0$, for $N \geq 1$.] The set $\{A_N(n)\}$ is a *global (absolute) shape descriptor*. It does not describe local features but conveys information on the folding of the entire backbone. *A priori*, the *instantaneous* overcrossing distribution of a polymer conformation does not depend much on its size or anisometry, *but rather on the twists, turns, and convolutions (self-entanglements) of the chain*. We refer to these features as complexity of entanglements.[52−56]

Two parameters can be used as shape descriptors for the distributions above: (1) A^*, the probability of the most probable number of overcrossings N^* (*i.e.*, the maxima of the distribution); and (2) $\overline{N}$, the mean number of overcrossings, given by

$$\overline{N} = \sum_{N=0}^{max\ N} NA_N(n) \qquad (8)$$

These descriptors capture essential folding features. Their basic qualitative properties are as follows: (a) for convoluted and entangled backbones, $\overline{N}$ and N^* take large values and A^* is small; (b) for swollen and disentangled chains, $\overline{N} \to 0$ and $A^* \to 1$. Analyzed in detail, the overcrossing distribution allows one to recognize global shape homologies between tertiary folding in proteins[40] and common features among entire families of proteins.[54,55] In addition, the configurational averages of these descriptors for self-avoiding random walks indicate a well-defined power-law scaling behavior.[54,55] Other properties of these descriptors found in the study polymer models with excluded volume interaction are discussed in refs 53 and 54.

The differences between the overcrossing spectra of two molecules can be illustrated with of our present example, the *n*-alanines. Consider the regular α-helical conformations of 10-alanine and 25-alanine. [See the next section regarding the details of their determination.] We can compare some of their particular values of overcrossing probabilities in the range of low overcrossing numbers. Up to two significant figures, the $A_N(n)$ values found for N from 0 to 8 are respectively 0.45, 0.13, 0.32, 0.02, 0.04, 0.00, 0.03, 0.00, and 0.01 in the case of 10-alanine. [The overcrossing probabilities extend, of course, beyond $N = 8$. Only the first nine $A_N(n)$ values are quoted for illustration.] The quantitative shift in the spectrum with the chain length can be noticed for the same A_N values in 25-alanine: 0.43, 0.01, 0.06, 0.01, 0.03, 0.03, 0.32, 0.01, 0.00. We observe that, whereas the most probable number of overcrossings is $N^* = 0$ in both cases, the second most probable value is $N = 2$ for 10-alanine, but $N = 6$ in 25-alanine. When the *complete* overcrossing spectra are taken into account, this translates in a shift in the mean number of overcrossings from $\overline{N} = 1.34 \pm 0.01$ in 10-alanine to 5.36 ± 0.04 in 25-alanine. By the position of the first overcrossing probability maximum (*i.e.*, A^*), we can recognize the occurrence of common shape features, in this case, the α-helical conformation. The mean number of overcrossings is linearly related to the length of the

helix. [However, note that the scaling takes a more complicated form, $\langle\overline{N}\rangle \sim an^{\beta}$ ($\beta > 1$), in the case of configurational averages over *random walk* (as opposed to *regular*) conformations.[54]

In conclusion, we propose that the simultaneous use of the shape descriptors R_G, Ω, A^*, and $\overline{N}$ provides a thorough picture of all relevant shape features. Recently, we have used these descriptors to study the role of temperature and composition on folding transitions in decapeptides.[57] In the present work, these descriptors and their fluctuations will be used as the parameters Γ and σ_Γ (eqs 2 and 3), employed to monitor the intrinsic shape stability of polypeptides with variable length.

Molecular Dynamics Simulations

The molecules considered are a series of poly(L-alanines), with residue numbers $5 \leq n \leq 35$. [For simplicity, we refer to them as polyalanines or *n*-alanines.] Polyalanine is the simplest polypeptide known to form α-helices easily.[20−27] For sufficiently long polyalanines, the global minimum is thought to be "helicoidal" in nature.[30] The *minimum* number of residues needed to reach this condition, n^*, is approximately 10.[30,43−48]

We are interested in following *equilibrium motions* about the α-helical minimum energy conformations. [Whereas all the work below has been restricted to *n*-alanines, the basic concepts are general and applicable to the study of shape stability in any other polymer.] In order to assess *intrinsic shape stability*, only one isolated molecule, coupled to a simulated thermal bath, has been considered. In order to simulate the effect of a solvent, we have included a variable dielectric screening of atomic charges. We want to determine the critical value n_c of the number of residues which conserve more strongly the helical shape. One would expect $n_c > n^*$.

For each molecule, we have computed constant-temperature trajectories and analyzed the shape behavior along them. For a proper comparison, the simulation has been carried using a strategy comparable to previous analyses of intrinsic stability in the literature.[17,18] Structures are analyzed at molecular mechanics level. [Electronic effects should play a small role in determining global macromolecular shape features.] The conditions include the thermal bath and dielectric screening mentioned above.

The details of the simulations are as follows: (a) The AMBER[58,59] force field (all-atom, version 3.0) was employed, as implemented in the molecular modeling program HyperChem.[60] (b) A distance-dependent dielectric constant ($\epsilon = r$) was used to simulate some degree of solvent screening. (c) No cutoffs for nonbonded interactions were included. (d) The initial structures were fully optimized with the force field, starting from standard helices (defined by dihedral angles $(\phi,\psi) = (-58°,-47°)$).[3] (e) Helices were finished at the peptide bonds, without charged or capped terminals. (f) Dynamic simulations were performed (after 10 ps equilibration) with a coupling to a thermal bath at 300 K. The scaled-velocity algorithm of Berendsen *et al.*,[61] with a bath-relaxation constant of 0.1 ps after equilibrium, was used to allow energy dissipation. (g) Configurational snapshots were retained at intervals of 0.25 ps over a 150 ps trajectory. (h) At each snapshot, the shape descriptors R_G, Ω, A^*, and $\overline{N}$ were evaluated. In the case of entanglement descriptors, A^* and $\overline{N}$ were evaluated along the trajectory with three series of 10 000 randomized points per snapshot.

It has been noticed that polyalanines with $n > 10$ exhibit a clustering of easily accessible conformational minima resembling a helix.[30] This clustering of minima suggests that the helical conformation is stable (or "persists") in configurational space. Velikson *et al.*[30] found that this distribution of minima depends little on a dielectric constant ϵ, taking values between 1 and 80. We expect therefore that our choice of a distance-dependent dielectric constant ($\epsilon = r$) for *n*-alanines ($n \leq 35$) should not affect the essential features of the potential energy landscape *near the helical minimum*.

Note that properties of the conformational surface can serve to characterize the persistence of a helical structure *only* if we have an acceptable definition of the "α-helical content" of a given conformation. However, this is not a uniquely established notion in the literature. There is no general agreement on how much a structure resembles a helix once the vibrations about the minimum energy conformer are taken into consideration. For this reason, we do not define a helical content but instead monitor the deviation of several shape descriptors from the values of a standard helix at equilibrium. As commented before, the shape descriptors we use are not dependent on each other by construction. Therefore, the occurrence of similarities or correlations between their configurational averages can be an important result toward understanding macromolecular dynamics.[57] In the next section, I discuss the behavior of configurationally averaged descriptors from the simulations with a variable number of amino acid residues.

Results and Discussion

1. Dependence of Configurationally Averaged Shape Descriptors on Helical Length. By following the shape descriptors along the MD trajectories, we have computed configurational averages and fluctuations (eqs 2 and 3). These trajectories represent equilibrium motions about the helical conformational minimum at $T = 300$ K. The helices considered have *n* amino acid residues $n = 5$, 10, 15, 20, 25, 30, and 35.

A first visual analysis indicates that the shorter helices, $n \leq 10$, exhibit some degree of "unfolding". This behavior for short helices coincides with various results in the literature, based on energetic and structural analysis. The lack of stability for short helices appears to be the same whether one uses a weak bath coupling[17,22] or long-time overdamped Langevin dynamics.[31] Our observation can be made quantitative in terms of all the shape descriptors. The averages of the size descriptor, $\langle R_G\rangle$, and the mean number of overcrossings, $\langle\overline{N}\rangle$, show a linear correlation with the number of residues. In the case of the radius of gyration (with $5 \leq n \leq 35$), we find

$$\langle R_G\rangle\,(\text{Å}) \approx (0.40 \pm 0.02)n + (1.0 \pm 0.4), \qquad C = 0.99922 \quad (9)$$

where C is the correlation coefficient and the quoted errors are 95% confidence intervals. The mean number of overcrossings shows a similar behavior:

$$\langle\overline{N}\rangle \approx (0.266 \pm 0.013)n + (-1.4 \pm 0.3), \qquad C = 0.99912 \quad (10)$$

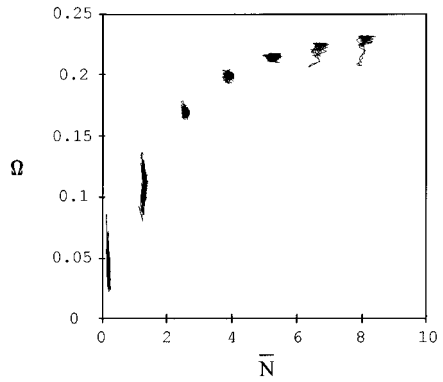

Figure 1. Instantaneous values of asphericity Ω and mean overcrossing numbers $\overline{N}$ for polyalanine helices with $n = 5$, 10, 15, 20, 25, 30, and 35, along constant-temperature molecular dynamics trajectories ($T = 300$ K).

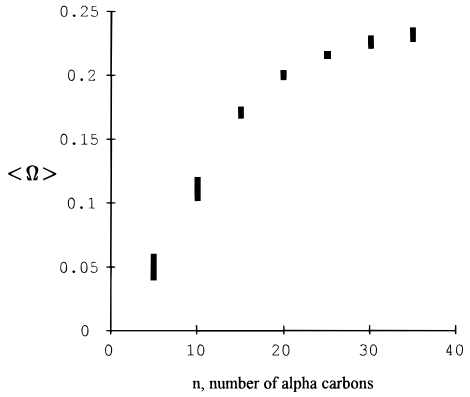

Figure 2. Configurational average of asphericity ⟨Ω⟩ for equilibrium motions of polyalanine helices with variable number of α-carbon atoms n. [The variable sizes of the bars correspond to the configurational fluctuations represented by standard deviations σ_Ω.]

The correlation between size and overcrossing numbers is even stronger:

$$\langle R_G \rangle \text{ (Å)} \approx (1.48 \pm 0.02)\langle \overline{N} \rangle + (3.0 \pm 0.1), \qquad C = 0.99994 \quad (11)$$

The fluctuations along the trajectories show a more complicated behavior. This behavior can be visualized more clearly if we study the other two shape descriptors, namely, the conformational asphericity Ω and the maximum overcrossing probability A^*.

Figure 1 shows the complete results for the asphericity and the mean overcrossing numbers along the trajectories for $5 \leq n \leq 35$. The "spots" correspond to the values of shape descriptors Ω and $\bar{N}$, linked by straight line segments consecutively as they appear along the trajectories. [Note that the behavior of Ω is nonlinear since it depends asymptotically on n as $\Omega \approx {}^1/_4 - \text{const} \times n^{-2}$, for $n \gg 1$.] The figure reveals two facts: (a) The fluctuations in the descriptors depend nonlinearly on n. (b) The trajectories for long helices explore shapes which are significantly different from the mean helical shapes.

The results in Figure 1 indicate that both short and long helices exhibit shape fluctuations. Figure 2 confirms this observation by displaying the mean value ⟨Ω⟩ and the fluctuations σ_Ω in the distribution (as error bars). The fluctuations appear to be smaller for the helices with 20 and 25 amino acid residues. Figure 3 complements this description with the results for the averages and fluctuations in the maximum overcrossing

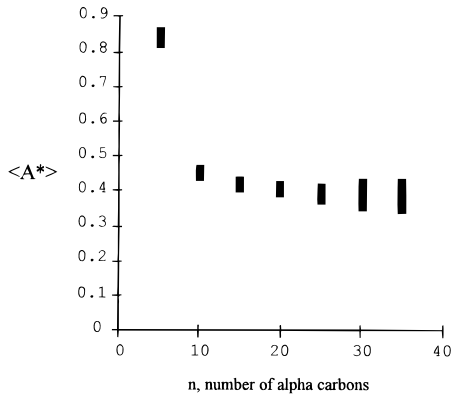

Figure 3. Configurational average of the maximum probability of overcrossings for equilibrium motions of ⟨A^*⟩ for polyalanine helices with variable number of α-carbon atoms n. [The variable sizes of the bars correspond to the configurational fluctuations represented by standard deviations σ_{A^*}.]

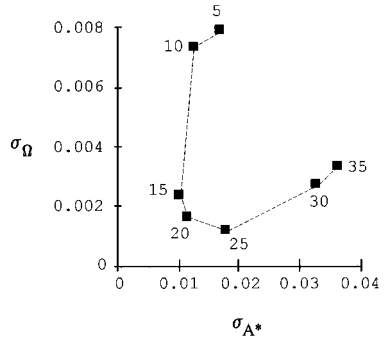

Figure 4. Interrelation between fluctuations in asphericity and maximum probability of overcrossings for polyalanine helices with variable lengths. [The numbers inside the diagram indicate n, the number of α-carbons. According to this criterion, the helices with $n \approx 20$ conserve more strongly their shape under the simulation conditions.]

probability, ⟨A^*⟩ and σ_{A^*} respectively. Note that, whereas the mean A^* values depend little on n for polyalanine helices with more than 20 residues, the fluctuations do depend on the number of residues.

2. Differential Shape Stability in Terms of Helical Length. The distinct behavior observed in the above descriptors suggests that more than one property is needed for a complete characterization of shape stability in helices. Figure 4 presents such a characterization by relating fluctuations in asphericity with fluctuations in overcrossing probabilities for various helices.

In Figure 4, there are *two domains of shape fluctuations*. Short helices ($n \leq 10$) exhibit small changes in entanglement complexity accompanied by large changes in asphericity. That is, short helices maintain a "loop entanglement" that resembles a helix during equilibrium motions, but they distort the original helical cylindrical geometry. Longer helices exhibit the opposite behavior: the configurational fluctuations maintain the overall cylindrical shape but the "loop entanglements" change with respect to those of a helix. The optimum shape stability according to the two criteria is found for $n \approx 20$.

Figure 5 provides an alternative description by comparing the fluctuations in asphericity with the *normalized* fluctuations in the mean number of overcrossings, $\sigma_{\overline{N}}/\langle \overline{N} \rangle$. [Note that both σ_{A^*} and σ_Ω are bound quantities which do *not* scale linearly with the number of residues.] Even though the quantitative result is dif-

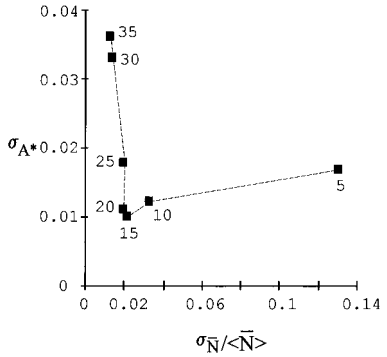

Figure 5. Interrelation between fluctuations in maximum probability of overcrossings and normalized fluctuations in mean number of overcrossings, $\sigma_{\bar{N}}/\langle\bar{N}\rangle$. [The numbers inside the diagram indicate n, the number of α-carbons. According to this criterion, the maximum shape stability appears at $15 < n < 20$.]

ferent from that in Figure 4, the qualitative conclusions are the same. Figure 5 indicates the existence of two stability regimes as a function of the number of residues. The maximum shape stability appears again in the range $15 < n < 20$.

Figures 4 and 5 represent some of the main results in this work. They convey shape stability quantitatively in terms of fluctuations in independent shape descriptors. In the case of α-helices, the analysis suggests a critical length for maximum helical shape stability of $n_c \approx 20 \pm 5$.

3. Damped Periodic Shape Oscillations in Long Helices. By analyzing only the mean values of various shape descriptors, it is difficult to gain much insight into the actual processes leading to shape instability. In contrast, Figure 1 shows that the shape instability in long helices must be due to their accessibility to conformations which are less prolate than a helix and slightly less entangled. Therefore, we expect that a more detailed analysis of MD trajectories can clarify the nature of these motions.

As a first illustration, we consider the case of 30-polyalanine. Figure 6 shows the dynamical changes in shape descriptors along the 150 ps trajectory, after equilibration is achieved at 300 K. We note distinct differences in behavior for the shape descriptors during the initial 50 ps. Whereas R_G exhibits uniform oscillations about the equilibrium value $R_G \approx 12.92$ Å, the other shape descriptors indicate the presence of more entangled, spheroidal shapes before 50 ps. An analysis restricted to the radius of gyration would have indicated no special shape differences during the initial period (except, perhaps, for a larger configurational flexibility, as suggested by the larger amplitude oscillations). In contrast, the remaining descriptors ($\bar{N}$, Ω, and A^*) reveal the richness of features in the actual shape fluctuations taking place.

Figure 6 indicates that the asphericity Ω and maximum overcrossing probability A^* have periodic oscillations in shapes before 50 ps. These are accompanied by values of $\bar{N}$ below the mean $\langle\bar{N}\rangle$. The regular oscillations take place between configurations which resemble helices (the maxima in A^* and Ω), on the one hand, and more entangled and less prolate configurations (the minima in A^* and Ω), on the other hand.

The nature of these configurational oscillations is illustrated in Figure 7. The diagram shows four consecutive extrema in A^*. From a minimum to a maximum in A^*, the conformation passes from a *bent* helix

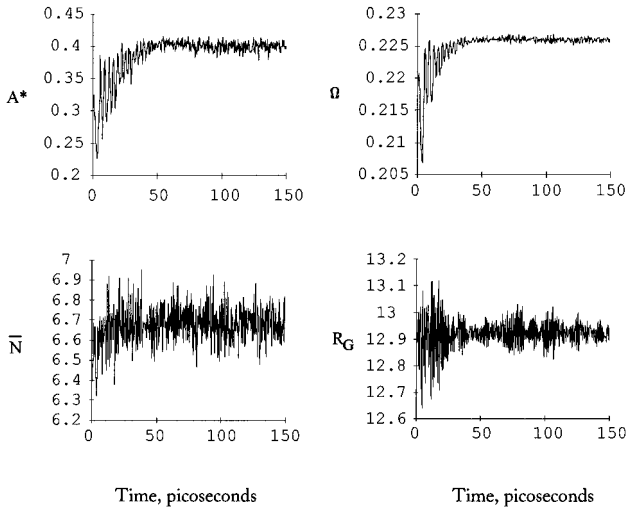

Figure 6. Dynamical changes in shape descriptors along the trajectory for 30-polyalanine at 300 K. [Radius of gyration values are in angstroms. Configurations are collected once equilibration is achieved ($t \equiv 0$). Note the differences in behavior for the shape descriptors for $t < 50$ ps. Whereas R_G exhibits uniform oscillations about the equilibrium value $R_G \approx 12.92$ Å, the other shape descriptors indicate the presence of more entangled, spheroidal shapes before 50 ps. The asphericity Ω and maximum overcrossing probability A^* indicate periodic oscillations in shapes before 50 ps.]

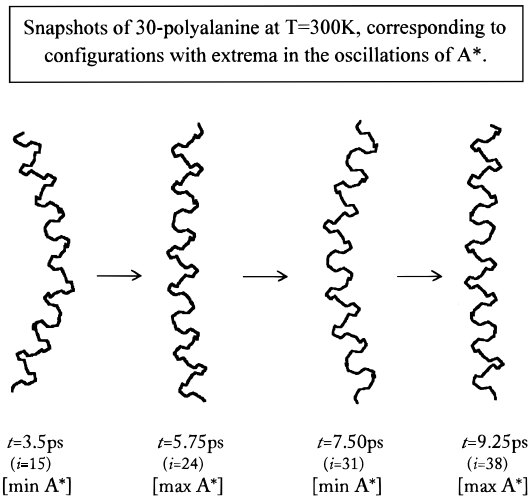

Figure 7. Configurations of 30-polyalanine corresponding approximately to the first maxima and minima in A^* found in Figure 6. [The numbers below indicate the number of snapshot (i) and the simulation time (t). The bent conformations are "more entangled" and less prolate than the stretched conformations. Note that configurational change from $i = 15$ to $i = 31$ (or $i = 24$ to $i = 38$) represents *half a period* for the oscillation.]

to a *normal* helix. The fluctuations in the shape descriptors correlate with a damped "hinge bending oscillation" in the helices. Along this motion, the helix moves away regularly from the perfect cylinder, yet with decreasing amplitudes. Note that the two consecutive minima in A^* correspond to the two consecutive *turning points of one period of oscillation*.

The oscillations appear to be damped in amplitude but regular in frequency. Using the first 10 vibrations, we can estimate the period τ_{30}, frequency ν_{30}, and wavenumber $\bar{\nu}_{30}$ of this motion in 30-alanine:

$$\tau_{30} \approx 6.8 \pm 0.4 \text{ ps}, \quad \nu_{30} \approx 0.147 \pm 0.009 \text{ ps}^{-1}, \quad \bar{\nu}_{30} \approx 4.9 \pm 0.3 \text{ cm}^{-1} \quad (12)$$

The order of magnitude of this motion is comparable with the typical time scales for elastic vibrations in globular regions of proteins, and global stretchings and bendings in proteins and nucleic acids.[9] These motions can be extracted by standard normal mode analysis of protein dynamics.[62,63] However, in our case, the result is derived from oscillations in shape descriptors of entanglement and anisometry. As we noted above, this behavior cannot be extracted from the changes in molecular size.

Longer polyalanine helices show a very similar behavior. Figure 1 indicates that 35-alanine also present oscillations toward less prolate and entangled conformations. Upon analyzing the oscillations in shape descriptors for $t < 50$ ps, we find fluctuations similar to those in 30-alanine. Averaging over the first eight periods, we obtain

$$\tau_{35} \approx 7.8 \pm 0.4 \text{ ps}, \quad \nu_{35} \approx 0.128 \pm 0.007 \text{ ps}^{-1}, \quad \bar{\nu}_{35} \approx 4.3 \pm 0.2 \text{ cm}^{-1} \quad (13)$$

The damped helical oscillations are not observed for short helices. The former are barely discernible at $n = 25$. We can only make the following approximate assessment:

$$\tau_{25} \approx 4.4 \pm 0.7 \text{ ps}, \quad \nu_{25} \approx 0.23 \pm 0.04 \text{ ps}^{-1}, \quad \bar{\nu}_{25} \approx 7.6 \pm 1.3 \text{ cm}^{-1} \quad (14)$$

These latter values agree well with a recent study of vibrations in a polyalanine helix of length $n = 21$.[64] These authors find that the dominant motions in the range of 11−20 cm^{-1} (for $n = 21$) correspond to global twisting and "bowstring bending" of the helix.

The clear differences between the values of τ_n for $n = 25$, on one hand, and $n = 30$ and 35 on the other, suggest that the corresponding equilibrium motions are not entirely comparable. Our results indicate that the "bending" motions which affect the molecular shape either do not take place for $n < 25$, or they become too fast to be detected. This is consistent with the fact that we find the strongest persistence of helical shape features for $n \approx 20$ residues.

In summary, our results indicate that global motions about the equilibrium conformation modify significantly the helical shape only in helices longer than $n = 25$. These motions become slower in longer helices, even though their "hinge" (or rather "bowstring") bending nature appears not to change. The associated fluctuations in shape can be related to the fact that many conformational minima are easily accessible from the helical minimum in longer polypeptides. Many of these minima will have molecular shapes distorted with respect to the helix. The fluctuations observed may represent the tendency in long structures to move away from a cylindrical helix to a coiled helix.

Conclusions and Closing Remarks

In this work, we have presented an approach to quantify the stability of intrinsic molecular shape features under equilibrium motions about regular conformations. The equilibrium motions involve *all* energies and conformations accessible to an isolated molecule in equilibrium with a simulated bath. These motions are not constrained to be either local or global vibrations. The present methodology involves a simultaneous analysis of a series of molecular shape descriptors especially designed to account for independent aspects of shape. We have shown that, by using shape descriptors of anisometry and entanglement, it is possible to recognize motions which modify the molecular shape. These motions would otherwise not be detected by using standard descriptors, *e.g.*, the radius of gyration. By using our approach, it is possible to quantify the role of various properties on shape, including composition, number of monomers, and the temperature of a thermal bath. In this context, we have evaluated the critical number of residues that minimize the overall shape fluctuations in a family of isolated *n*-alanines, at constant ambient temperature. The working conditions of relatively low temperature make polymer length the most important variable to affect shape stability.

The analysis of the *persistence of a molecular shape feature over time* is important in molecular recognition. For example, a flexible molecule will normally exhibit poor binding at a receptor site. Yet, it is possible for a floppy molecule to be biochemically active if its shape transitions take place within a time scale *larger* than the half-times for the recognition and binding reaction steps. In this work, we have presented a methodology to determine such critical time scales.

The analysis of persistence can also be relevant to study not only equilibrium but also global motions in proteins. The existence of structural transitions involving correlated displacements of many atoms over large distances is believed to play an essential role in protein functionality.[65,66] The present methodology can also be used in this context. By monitoring global folding features, we can establish whether such correlated motions lead to any essential changes in molecular shape.

Acknowledgment. This work has been supported by the Fonds de Recherche de l'Université Laurentienne (FRUL) and an operating grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References and Notes

(1) Poland, D.; Scheraga, H. A. *Theory of Helix−Coil Transitions in Biopolymers*; Academic Press: New York, 1970.
(2) Vázquez, C.; Némethy, G.; Scheraga, H. A. *Chem. Rev.* **1994**, *94*, 2183.
(3) Brändén, C.; Tooze, J. *Introduction to Protein Structure*; Garland: New York, 1991.
(4) Thom, R. *Structural Stability and Morphogenesis: An Outline of a General Theory of Models*; Addison-Wesley: Redwood City, CA, 1989.
(5) Gilmore, R. *Catastrophe Theory for Scientists and Engineers*; Wiley: New York, 1985.
(6) Bader, R. F. W. *Atoms in Molecules*; Oxford University Press: Oxford, 1990.
(7) Mezey, P. G. *Potential Energy Hypersurfaces*; Elsevier: Amsterdam, 1987.
(8) Arteca, G. A. In *Advances in Computational Biology*; Villar, H. O., Ed.; JAI Press: Greenwich, CT, 1994; Vol. 1, 1−67.
(9) McCammon, J. A.; Harvey, S. C. *Dynamics of Proteins and Nucleic Acids*; Cambridge University Press: Cambridge, 1987.
(10) Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *Adv. Chem. Phys.* **1988**, *71*, 1.
(11) Chan, H. S.; Dill, K. A. *Annu. Rev. Biophys. Biophys. Chem.* **1991**, *20*, 447.
(12) Daggett, V.; Levitt, M. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 353.
(13) Brooks, C. L., III; Case, D. A. *Chem. Rev.* **1993**, *93*, 2487.
(14) Daggett, V.; Levitt, M. *Curr. Opinion Struct. Biol.* **1994**, *4*, 291.
(15) Creighton, T. E., Ed. *Protein Folding*; Freeman: New York, 1992.

(16) Rojewska, D.; Elber, R. *Proteins: Struct. Funct. Genet.* **1990**, *7*, 265.
(17) Daggett, V.; Kollman, P. A.; Kuntz, I. D. *Biopolymers* **1991**, *31*, 1115.
(18) Wang, C. Y.; Yang, C. F.; Lai, M. C.; Lee, Y. H.; Lee, T. L.; Lin, T. H. *Biopolymers* **1994**, *34*, 1027.
(19) Di Capua, F. M.; Swaminathan, S.; Beveridge, D. L. *J. Am. Chem. Soc.* **1990**, *112*, 6768.
(20) Marqusee, S.; Robbins, V. H.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 5286.
(21) Padmanabhan, S.; Marqusee, S.; Ridgeway, T.; Laue, T. M.; Baldwin, R. L. *Nature* **1989**, *344*, 268.
(22) Daggett, V.; Kollman, P. A.; Kuntz, I. D. *Biopolymers* **1991**, *31*, 285.
(23) Van Buuren, A. R.; Berendsen, H. J. C. *Biopolymers* **1993**, *33*, 1159.
(24) Okamoto, Y. *Biopolymers* **1994**, *34*, 529.
(25) Serrano, L.; Sancho, J.; Hirshberg, M.; Fersht, A. R. *J. Mol. Biol.* **1992**, *227*, 544.
(26) Horovitz, A.; Matthews, J. M.; Fersht, A. R. *J. Mol. Biol.* **1992**, *227*, 560.
(27) Serrano, L.; Neira, J.-L.; Sancho, J.; Fersht, A. R. *Nature* **1992**, *356*, 453.
(28) Chou, P. Y.; Fassman, G. D. *Annu. Rev. Biochem.* **1978**, *47*, 251.
(29) Fraga, S.; Parker, J. M. R. *Amino Acids* **1994**, *7*, 175.
(30) Velikson, B.; Bascle, J.; Garel, T.; Orland, H. *Macromolecules* **1993**, *26*, 4791.
(31) Grønbech-Jensen, N.; Doniach, S. *J. Comput. Chem.* **1994**, *15*, 997.
(32) Shalongo, W.; Dugad, L.; Stellwagen, E. *J. Am. Chem. Soc.* **1994**, *116*, 8288.
(33) Alemán, C. *Biopolymers* **1994**, *34*, 841.
(34) Toumadje, A.; Johnson, W. C., Jr. *Biopolymers* **1994**, *34*, 969.
(35) Jiménez, M. A.; Carreño, C.; Andreu, D.; Blanco, F. J.; Herranz, J.; Rico, M.; Nieto, J. L. *Biopolymers* **1994**, *34*, 647.
(36) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Interscience: New York, 1969.
(37) (a) Rudnick, J.; Gaspari, G. *J. Phys. A* **1986**, *19*, L191. (b) Rudnick, J.; Gaspari, G. *Science* **1987**, *237*, 384.
(38) Aronovitz, J. A.; Nelson, D. R. *J. Phys. (France)* **1986**, *47*, 1445.
(39) Baumgärtner, A. *J. Chem. Phys.* **1993**, *98*, 7496.
(40) Arteca, G. A. *Biopolymers* **1993**, *33*, 1829.
(41) Shoemaker, K. R.; Kim, P. S.; Brems, D. N.; Marqusee, S.; York, E. J.; Chaiken, I. M.; Stewart, J. M.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **1985**, *82*, 2349.
(42) Wojcik, J.; Altmann, K. H.; Scheraga, H. A. *Biopolymers* **1990**, *30*, 121.
(43) Rapaport, D. C.; Scheraga, H. A. *Macromolecules* **1981**, *14*, 1238.
(44) Ripoll, D. R.; Scheraga, H. A. *Biopolymers* **1988**, *27*, 1283.
(45) Gregoret, L. M.; Cohen, F. E. *J. Mol. Biol.* **1991**, *219*, 109.
(46) Chan, H. S.; Dill, K. A. *J. Chem. Phys.* **1991**, *95*, 3775.
(47) Maiorov, V. N.; Crippen, G. M. *J. Mol. Biol.* **1992**, *227*, 876.
(48) Yee, D. P.; Chan, H. S.; Havel, T. F.; Dill, K. A. *J. Mol. Biol.* **1994**, *241*, 557.
(49) Šolc, K. *Macromolecules* **1973**, *6*, 378.
(50) Gobush, W.; Šolc, K.; Stockmayer, W. H. *J. Chem. Phys.* **1974**, *60*, 12.
(51) Arteca, G. A.; Mezey, P. G. *Biopolymers* **1992**, *32*, 1609.
(52) Arteca, G. A. *J. Comput. Chem.* **1994**, *15*, 633.
(53) Arteca, G. A. *Int. J. Quantum Chem. QCS* **1994**, *28*, 433.
(54) Arteca, G. A. *Phys. Rev. E* **1994**, *49*, 2417.
(55) Arteca, G. A. *Phys. Rev. E* **1995**, *51*, 2600.
(56) Orlandini, E.; Tesi, M. C.; Whittington, S. G.; Sumners, D. W.; Janse van Rensburg, E. J. *J. Phys. A* **1994**, *27*, L333−338.
(57) Arteca, G. A. *Biopolymers* **1995**, *35*, 393.
(58) Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.
(59) Singh, U. C.; Weiner, P. K.; Caldwell, J.; Kollman, P. A. *AMBER 3.0*; University of California, San Francisco, 1986.
(60) *HyperChem for Windows* 4.0; Hypercube Inc., Waterloo, ON, Canada, 1994.
(61) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
(62) Tapia, O.; Nilsson, O.; Campillo, M.; Åqvist, J.; Horjales, E. In *DNA Protein Complexes and Proteins: Structure and Methods*; Sarma, R. H., Sarma, M. H., Eds.; Adenine Press: New York, 1990; Vol. 2, pp 147−170.
(63) Case, D. A.; *Curr. Opinion Struct. Biol.* **1994**, *4*, 285.
(64) Furois-Corbin, S.; Smith, J. C.; Lavery, R. *Biopolymers* **1995**, *35*, 555.
(65) Richard, L.; Genberg, L.; Deak, J.; Chiu, H.-L.; Miller, R. J. D. *Biochemistry* **1992**, *31*, 10703.
(66) Miller, R. J. D. *Acc. Chem. Res.* **1994**, *27*, 145.

MA9606869