ELSEVIER

Structural transitions in neutral and charged proteins *in vacuo*

Gustavo A. Arteca* and O. Tapia†

**Département de Chimie et Biochimie, Laurentian University, Sudbury, Ontario, Canada*
†Department of Physical Chemistry, Uppsala University, Uppsala, Sweden

In vacuo proteins provide a simple laboratory to explore the roles of sequence, temperature, charge state, and initial configuration in protein folding. Moreover, by the very absence of solvent, the study of anhydrous proteins *in vacuo* will also help us to understand specific environmental effects. From the experimental viewpoint, these systems are now beginning to be characterized at low resolution. Molecular dynamics (MD) simulations, in combination with tools for protein shape analysis, can complement experiments and provide further insights on the folding–unfolding transitions of these proteins. We review some aspects of this issue by using the results from a detailed MD study of hen egg-white lysozyme. For lysozyme ions, unfolding can be triggered by Coulombic repulsion. In neutral lysozyme, unfolding can be induced by centrifugal forces and also by weakening the monomer–monomer interaction. In both cases, the resulting unfolded transients can be used as initial configurations for relaxation dynamics. All trajectories are analyzed in terms of global molecular shape features of the backbone, including its anisometry and chain entanglement complexity. This strategy allows us to quantify separately the degree of polymer collapse and the evolution of large-scale folding features. Using these last two notions, we discuss some basic questions regarding the nature of the accessible paths associated with unfolding from, and refolding into, compact conformers.

1. INTRODUCTION

The realistic modeling of biomolecules has progressed enormously in the past 15 years. Early molecular dynamics (MD) simulations typically involved short trajectories (e.g., $t \leq 100$ ps) of small proteins that were embedded in a vacuum or a dielectric continuum.[1] Currently, nanosecond-long simulations of medium-sized proteins embedded in a microscopic solvent model are standard.[2–7] If the possibility of a detailed modelling of solvated proteins is readily available, why, then, do we study proteins *in vacuo*? We first consider this question to help understand the motivation for this work. Although the natural medium for proteins is an aqueous solution, advances in desorption and electrospray ionization now allow researchers to study anhydrous proteins experimentally.[8–18] These techniques allow researchers to place "intact" biomolecular ions in a gas phase or vacuum. In turn, these species can be characterized in flight by an array of modern techniques, including mass spectrometry,[15–18] rates of H/D exchange,[13–14] ion-drift mobility,[8,10,11,17] and surface imprinting.[19,20] Even though these techniques still produce low-resolution structural information, they provide valuable insights into the conformational properties of proteins. More importantly, these results suggest the possibility that anhydrous proteins, depending on their total charge and temperature, can appear in both unfolded and folded states. This possibility, already conjectured theoretically in 1995,[21] casts a different light on current notions regarding solvent effects in protein folding.[22–35] If unfolded proteins can refold *in vacuo*, is there a unique specific role that the solvent plays in this process? Does the solvent change the time scale for the relaxation from the unfolded state, or does it open up pathways that are different from those found *in vacuo*? These are legitimate questions, whose answers require realistic computer modeling of proteins in a vacuum. We regard proteins *in vacuo* as interesting systems in their own right. On the one hand, they can teach us a great deal about the intrinsic properties of polypeptide chains. On the other hand, they are more than a simplified version of solvated proteins. They exhibit particular behaviors that are not yet fully understood. In both cases, theoretical and computational modelling have an important role to play. The problem must be addressed with proper simulation protocols that are consistent with *in vacuo* boundary conditions. These conditions are not necessarily the same as those used in early simulations of unsolvated proteins. Short-time simulations with no solvent are useful approximations to study local motions and intrinsic flexibility of protein regions. In contrast, the study of folding-unfolding transitions *in vacuo* requires long-time simulations that approach the behavior of actual biomolecular ions in-flight, i.e., molecules that rotate and translate freely. This is the viewpoint that we adopt in this

Corresponding author: G.A. Arteca, Département de Chimie et Biochimie, Laurentian University, Ramsey Lake Road, Sudbury, Ontario P3E 2C6, Canada

E-mail address: gustavo@nickel.laurentian.ca (G.A. Arteca)

Journal of Molecular Graphics and Modelling 19, 102–118, 2001

655 Avenue of the Americas, New York, NY 10010

1093-3263/01/$–see front matter
PII S1093-3263(00)00130-3

work. In summary, we view *in vacuo* proteins as a convenient laboratory to study the specific roles of sequence, initial configuration, and coupling to a thermal bath on protein folding. For this reason, we present here a strategy to analyze a large ensemble of MD trajectories that simulate unfolding and refolding transitions for proteins under various conditions of charge, temperature, and composition. The simulation data, combined with specialized tools for the analysis of molecular shape, provide enough information for a first try at tackling important issues. This study addresses the following questions:

1. Under which conditions can we trigger unfolding in neutral or charged proteins *in vacuo*?
2. How do the corresponding continua of unfolded conformers compare with each other?
3. Is there a common pattern of structural transitions associated with the unfolding of species in various charged states?
4. How does refolding proceed once the unfolding conditions are switched off?
5. How does the relaxation of an unfolded protein compare with the nonspecific collapse of a homopolymer?
6. Is there evidence of a well-defined folding process *in vacuo*, starting from the unfolded state ensemble?

This study reviews our recent work in these areas. We discuss current research in the field and pose a number of open questions that still require additional research. We have not attempted to cover all issues, but rather to select representative problems that illustrate well the techniques and the results. To this end, this discussion is based on results obtained for one protein, hen egg-white lysozyme (PDB code *1hel*). In Section 2 we present the techniques used for characterizing backbone shape during folding-unfolding transitions. In Section 3 we discuss a scheme to interpret the dynamical information in terms of current ideas of protein folding (e.g., the degree of polymer collapse and the growth of tertiary structure). Section 4 and Section 5 examine the conditions for triggering unfolding at room temperature in neutral and charged proteins, respectively. The trajectories are analyzed with the tools described in Section 2 and Section 3. Section 6 deals with effects associated with the strength of the nonbonded (van der Waals) pair interaction between monomers. In Section 7 we present the relaxation behavior of conformers in the unfolded state. These conformers are selected among the transient structures found along the unfolding trajectories. The results show manifold paths leading to diverse persistent structures, the latter including both quasi-native conformers and folded denatured conformers. The pattern displayed by the resulting paths suggests the occurrence of a well-defined folding process *in vacuo*. We discuss these findings in terms of current notions of the factors that control protein folding. A summary of conclusions is given in Section 8.

2. DESCRIPTORS OF MOLECULAR SHAPE FOR PROTEIN BACKBONES

The proper choice of molecular shape descriptors is an important ingredient for a detailed analysis of folding-unfolding trajectories.[36] Even when the analysis is restricted to the properties of the backbone trace, the amount of information involved in a single MD run is still large. This information must be filtered by using a selected set of molecular shape properties that provide both a clear interpretation and useful insight. In our case, we focus on a priori independent molecular shape descriptors, which can easily be related to the notions of polymer collapse and the development of a tertiary structure. The current consensus is that molecular compactness, content of secondary structure, and the formation of tertiary contacts are key properties that characterize a folding process.[22–34,37] Results show that compactness can stabilize secondary structure if it is already present, but it does not necessarily lead to its formation.[22,23,29,31,38,39] Moreover, the relation between these factors is subtle, since compactization may trigger growth in secondary structure, depending on the strength of the monomer–monomer interaction.[24,26,40,41] With the present shape descriptors, analysis of *in vacuo* dynamics can be brought within the context of current notions on protein folding. A descriptor of chain compactness, e.g., molecular size, can measure the extent of polymer collapse. Similarly, features can be conveyed with relative descriptors, such as the root-mean square (RMS) deviation with respect to a target structure (e.g., the native fold).[30] Yet, such descriptors (e.g., the instantaneous radius of gyration, R_g) often do not discriminate well between chains in different spatial organizations. Instead, we use two shape descriptors differing in the information on which they are built. On the one hand, we employ a descriptor of chain anisometry[42–45] determined by the chain geometry (i.e., the α-carbon position vectors). On the other hand, we use a descriptor of chain entanglement[46–52] determined by both the geometry and the connectivity of the chain. The connectivity (i.e., the sequence of "bonds" between α-carbons) constitutes the simplest piece of topological information on the chain.

2.1. Chain Compactness

The descriptor of anisometry is the so-called asphericity Ω, given in terms of the principal moments of inertia $\{I_i\}$ as[42–45]:

$$\Omega = \frac{(I_2 - I_1)^2 + (I_3 - I_2)^2 + (I_3 - I_1)^2}{2(I_3 + I_2 + I_1)^2}. \qquad (1)$$

If a protein changes from a compact spheroid in the native state to an elongated structure in the unfolded state, we expect a change from $\Omega = 0$ to $\Omega = 1/4$. There are known properties for the asphericity of model polymers when averaged over all accessible conformers. We quote here a result that provides an important reference value for detecting folding-unfolding transitions. Let $\langle\Omega_D\rangle$ be the configurationally-averaged asphericity for D-dimensional chains. In the limit of infinitely long random walks (RWs), we have[42–44]:

$$\langle\Omega_D\rangle = \frac{2}{5(D-1)^2} - \frac{12}{175D(D-1)^2} + O(D^{-4}), \qquad (2)$$

which produces the estimate $\langle\Omega\rangle \approx 0.0943$ for three-dimensional chains ($D = 3$). This value is a good approximation for finite chains in the continuum, since Monte Carlo simulations with 10^3 uncorrelated conformers give $\langle\Omega\rangle_{RW} \approx 0.098 \pm 0.002$ for chains with $n > 50$ beads.[49] Since random walks contain no monomer–monomer interaction, their averaged asphericity represents an upper bound for the anisometry of realistic chains dominated by attractive interactions (e.g., protein native states). For random walks with excluded-volume interaction (i.e., self-avoiding walks, [SAWs]) in the continuum, the $\langle\Omega\rangle$ value increases with the degree of monomer

repulsion. For instance, in chains with $n \approx 100$, we have $\langle\Omega\rangle_{SAW} \approx 0.12 \pm 0.01$ for radii of excluded volume near one bond length.[49] This value represents a lower bound to the asphericity of realistic chains with specific repulsive interactions. That is, we expect protein chains in the unfolded state to have asphericity larger than $\langle\Omega\rangle_{SAW}$. As shown below, we can partition the configurational space into regions for compact and noncompact conformers by combining the $\langle\Omega\rangle$ values with those for the entanglement descriptor.

2.2. Chain Entanglements

Consider now a dynamical process in which the tertiary fold changes without largely affecting the mean size or anisometry of the chain. Such a situation would be found in the internal reorganization of collapsed homopolymer by loop diffusion[53] or within a continuum of partially folded intermediates during the folding of a protein.[54–58] For these situations, we need a different shape descriptor to capture the changes in fold. We use the self-entanglement complexity of a single polymer chain.[46–52] The term self-entanglement refers to the occurrence of projected bond–bond crossings, or overcrossings, in a chain.[46–48] This notion has been used to characterize permanent (or topological) entanglements, such as those found in molecular chains that are knotted or linked.[59–65] In these cases, there is a minimum number of overcrossings that cannot be removed despite homeomorphically accessible molecular rearrangements (i.e., motions that exclude chain breaking or self-crossing). In other words, the minimum overcrossing number is a topological invariant for chains with permanent entanglements. Linear chains, in contrast, exhibit only transient entanglements, which can be removed by loop diffusion or chain reptation.[66] Nevertheless, the dynamical properties of a polymer are still affected by the physical impossibility of instantly removing these entanglements by actual intrachain (ghost) crossings. Thus, a measure of self-entanglement complexity provides a characterization of folding features in a single chain, whether knotted or not. We use the distribution of projected bond–bond crossings[46,67] to create simple geometrical descriptors of chain entanglement. For quantitative analysis, we proceed as follows. First, we represent an instantaneous protein configuration, K, by its α–carbon backbone. Suppose that, upon projecting this backbone along an arbitrary direction in space, a number of bond–bond overcrossings, N, are observed. When considering all directions, we define the probability distribution $\{A_N^{(K)}\}$ for observing a number $N \geq 0$ of overcrossings, or overcrossing spectrum.[48] The $\{A_N^{(K)}\}$ distribution provides a rotationally and translationally invariant descriptor for the K conformer. This descriptor takes into account both geometrical and topological information of the chain. An efficient algorithm to compute $\{A_N^{(K)}\}$ using a large number of random projections is discussed in an article by Arteca.[48] We use two simple descriptors of chain entanglement based on the $\{A_N^{(K)}\}$ distribution. First, one can use a relative descriptor of entanglements, where the folding features of a given K conformer are compared with those of a reference K^* conformation, e.g., the native state. To this end, we use the RMS deviation of overcrossing probabilities for the K and K^* conformers [68,69]:

$$\sigma_A(K,K^*) = [\sum_{N=0}^{N_{max}} (A_N^{(K)} - A_N^{(K^*)})^2]^{1/2}. \quad (3)$$

For an absolute descriptor of entanglements, we can use the first moment of the distribution, i.e., the mean number of overcrossings $\bar{N}$:

$$\bar{N} = \sum_{N=0}^{N_{max}} N\, A_N^{(K)}\,, \quad (4)$$

where N_{max} is the maximum number of overcrossings. For linear chains with n monomers, we have $N_{max} \leq (n-2)(n-3)/2$. The mean number of overcrossings can also be computed directly, without evaluating the probability distribution. In this case, one must introduce the parametrized representation of the backbone, in terms of the line segments $\gamma_i(s)$, connecting a pair of i and $i+1$ (consecutive) α-carbons, and their corresponding parametric derivative, $\dot{\gamma}(s) = \partial\gamma_i/\partial s$. As shown elsewhere, the mean overcrossing number can be represented as a sum of contributions from all possible pairs of overcrossing segments or "bonds" [70,71]:

$$\bar{N} = \frac{1}{2\pi} \sum_{i=1}^{n-3} \sum_{j=i+2}^{n-1} \int_0^1 \int_0^1 \frac{|(\dot{\gamma}_1(s) \times \dot{\gamma}_j(t)) \cdot (\gamma_i(s) - \gamma_i(t))|}{\|\gamma_i(s) - \gamma_i(t)\|^3}\, ds\, dt. \quad (5)$$

A pair contribution to the overcrossings depends on the relative orientation of the bonds, as well as their mean distance. For two nonparallel bonds, their contribution to the total mean overcrossing number decreases with the square of the mean distance between the segments [71]. In DNA knots, $\bar{N}$ characterizes topological complexity,[60–63] and correlates closely with gel-electrophoresis drift velocities and sedimentation rates for different knots. In the case of linear chains, $\bar{N}$ measures fluctuations in chain configurations and the persistence of transient chain self-entanglements. If the chain unfolds, $\bar{N}$ decreases up to a certain amount that depends on the extent of residual secondary structure. In the formal limit of a rod, we have $\bar{N} \to 0$. Similarly, the relaxation of a protein chain from the unfolded state leads to an increase in $\bar{N}$, the final value depending on the precise folding features present in the intermediates encountered.[57,58,70,72]

3. SHAPE SPACE FOR MONITORING FOLDING-UNFOLDING TRANSITIONS

The two main shape descriptors we use here, the chain asphericity and entanglement complexity, convey geometrical, rather than topological, properties. That is, these two descriptors change during conformational rearrangements that conserve the connectivity. (We use here the term "topology" in its usual mathematical sense, i.e., when refering to properties of the chain that are unaffected by homeomorphic deformations. This meaning should be distinguished from the use of the term "chain topology" when referring to "folding features" or "tertiary structure.") Our goal is to use the pair $(\bar{N},\Omega)$ as order parameters to follow configurational rearrangements during folding-unfolding transitions.[51,57,58,72] As stated in Section 2, these two parameters serve as useful coordinates to monitor the evolution of some key molecular shape features. However, one should not forget that two coordinates alone cannot capture all

important information contained in an MD trajectory. Inevitably, the ($\bar{N}$,Ω) will exhibit degeneracy, i.e., two conformers having similar ($\bar{N}$,Ω) values could have different folding characteristics. Low Ω-values are related to the occurrence of spheroidal (possibly compact and folded) conformers. Similarly, low $\bar{N}$ values are related to the formation of elongated (possibly noncompact and unfolded) conformers. As a result, the path traced by a MD trajectory on a two-dimensional ($\bar{N}$,Ω) map conveys information that can be interpreted in terms of familiar notions such as polymer collapse and the growth of tertiary folds. Using the hypothetical trajectories in Figure 1, we discuss briefly how to interpret the main patterns in an ($\bar{N}$,Ω) map. First, the configurational space of a protein can be partitioned into two conceptually different sets by using the behavior of simple model polymers within the ($\bar{N}$,Ω) map. Consider the configurationally-averaged descriptors $\langle\bar{N}\rangle$ and ⟨Ω⟩ for self-avoiding walks of a given chain length and excluded-volume interaction.[49,50] If this repulsive interaction is specified by a radius of excluded volume, r_{ex}, the descriptors $\langle\bar{N}\rangle$ and ⟨Ω⟩ will change with the r_{ex} value. As r_{ex} increases, the SAWs elongate, and thus decrease the $\langle\bar{N}\rangle$ value by increasing ⟨Ω⟩. In the opposite extreme, we find the limit of RWs, associated with $r_{ex} \rightarrow 0$, which corresponds to a maximum in $\langle\bar{N}\rangle$ and a minimum in ⟨Ω⟩. The relation between $\langle\bar{N}\rangle$ and ⟨Ω⟩ is a smooth curve, not far from a straight line with negative slope for not-too-large values of r_{ex}.[57] This approximate line, denoted in Figure 1 as "*SAW line*," divides the ($\bar{N}$,Ω) plane in two regions according to the strength of the monomer–monomer interaction. Since SAWs exhibit only a random, nondirectional repulsive interaction, the SAW line separates the shapes corresponding to collapsed and unfolded polymers. The former correspond to compact structures with globally attractive interactions, whereas the latter involve noncompact conformers dominated by directional repulsions. That is, structures above the SAW line are swollen, i.e., noncompact, whereas compact structures appear below the line. The line provides an unbiased assessment of compactness for realistic polymer chains with the same number of monomers as the SAWs. To some extent, the ($\bar{N}$,Ω) map distinguishes between folding processes initiated by nonspecific polymer collapse from those initiated by growth of local secondary structure. These two processes can be recognized qualitatively by the type of trajectory traced on the shape map. Below, we discuss briefly these limit relaxation trajectories on the ($\bar{N}$,Ω) plane. The limit trajectories are represented by the hypothetical paths in Figure 1, denoted by (1) and (2). These *gedanken* trajectories start from a noncompact conformer. The snapshot in Figure 1 (top left) is representative of a typical initial unfolded transient: an elongated protein, with a reduced content of secondary structure. Such a conformer exhibits a large asphericity and a low entanglement complexity. Let us supppose that the relaxation paths (1) and (2) lead to similar final structures (Figure 1, bottom right), which are compact and entangled. Along trajectory (1), the chain evolves horizontally in ($\bar{N}$,Ω) space during the initial stages of relaxation. This chain does not compactify globally, but rather grows local "pockets" of secondary structure, thereby contributing to an increase in entanglement complexity. [See snapshot along path (1).] Eventually, tertiary structure develops sufficiently to produce spheroidal conformers, located within the regime of compact chains (i.e., below the SAW line). Qualitatively, such a process is akin to the occurrence of a "pearling" transition in homopolymers.[73] The hypothetical path (2) represents an opposite relaxation mechanism, dominated by the initial onset of global compactization.[23] The snapshot along trajectory (2) in Figure 1 represents a typical conformer where the chain becomes spheroidal without major changes in secondary structure. As a result, the chain decreases in asphericity, while only marginally increasing its entanglements. Eventually, higher entanglement complexity develops once the chain becomes sufficiently compact. In this regime, the path evolves horizontally on the ($\bar{N}$,Ω)-space plane. In summary, the descriptors $\bar{N}$ and Ω serve as order parameters to follow the degree of polymer compactization and growth of tertiary fold. The above-limit trajectories provide a reference to understand how the latter factors interrelate during actual relaxation. In the next sections, we use this approach to gain insight into the structural rearrangements that accompany folding-unfolding transitions for proteins *in vacuo.*

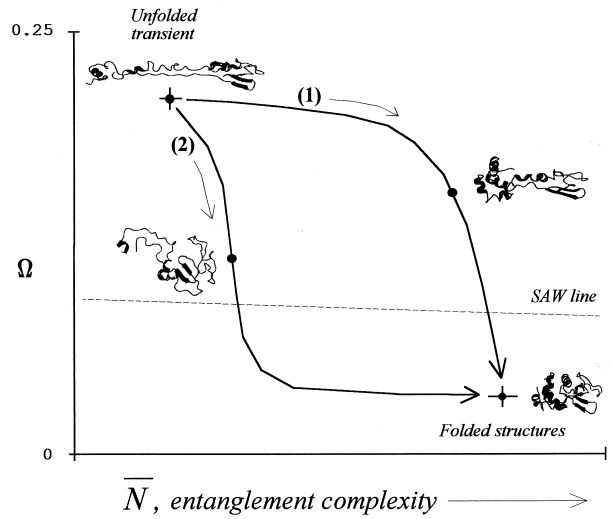

Figure 1. Two hypothetical relaxation paths in ($\bar{N}$,Ω)-space, starting from an unfolded transient and leading to a compact fold. Along path (1), the chain initially evolves by developing local secondary structure, thereby increasing the entanglement complexity, $\bar{N}$, without reducing the asphericity, Ω. Along path (2), the chain initially undergoes compactization without any development of local secondary structure. The "SAW line" represents the configurationally-averaged $\langle\bar{N}\rangle$ and ⟨Ω⟩ values for self-avoiding walks. Structures above this line are swollen with respect to SAWs by the occurrence of specific repulsions. Structures below the line are more compact than SAWs, therefore in a regime with dominant attractive interactions.

4. UNFOLDING OF NEUTRAL *1HEL*-LYSOZYME

The "unfolded state" of a protein is a continuum of conformers characterized by large configurational fluctuations. Despite large chain flexibility, this "state" does not include every conceivable unfolded structure. Some structures may not be reachable under realistic unfolding conditions, and unfolded states generated by distinct unfolding conditions may not even

be pathwise connected.[74] Accordingly, the proper choice of meaningful unfolded transients is a very important step in studies of protein relaxation and refolding. Various procedures have been used in the literature to produce unfolded structures, including the use of high temperature pulses,[75–77] freezing,[75,78] biasing by an external force,[79,80] solvent penetration into the core,[81] and residue charging[82] (see an article by Brooks[83] for a discussion of various techniques). An interesting alternative is the so-called "centrifugal" unfolding,[84] which we discuss and employ below. This unfolding technique involves the tranfer of energy from internal degrees of freedom to collective (large-scale) rotational and vibrational modes. Given that vibrations and rotations are coupled, a protein undergoing fast spinning about inertial axes becomes subject to weak, but systematic, "centrifugal forces." Although weak, this effect acts consistently in the same direction and can eventually trigger a large-scale vibrational mode that elicits the unfolding.[84] With this approach, it is possible to unfold a neutral protein at room temperature within a relatively short timescale. To this end, we consider a protein subject to *in vacuo* boundary conditions, whereby the molecule is allowed to rotate and translate freely.[82,84] In addition, the rotational activation of an unfolding mode can be facilitated by using a sufficiently strong coupling to a simulated thermal bath.[84] When applied to lysozyme, this procedure produces long unfolded structures with molecular sizes similar to those observed experimentally (approximately 120Å in length).[20] Also, the stepwise nature of the simulated transition is consistent with experimental observations.[85] In an article by Reimann et al.,[84] the practical conditions for unfolding neutral lysozyme (with intact disulfide bridges) are discussed in detail. Here, we present an expanded analysis to illustrate the nature of the transition. Lysozyme is studied with MD simulations using the GROMOS87 force field,[86] with the 37D4 parameter set that includes explicit polar hydrogens and electrostatic hydrogen bonding.[87] This parameter set is a proper force field for the simulation of neutral, anhydrous proteins. The GROMOS87 potential energy function includes the following terms:

$$V = V_{bond} + V_{bend} + V_{imp\text{-}dih} + V_{LJ} + V_{elec}\ , \qquad (6)$$

where V_{bond}, V_{bend}, and $V_{imp\text{-}dih}$ are the harmonic potentials for bond vibrations, bendings, and improper dihedral motions, respectively. The V_{tor} term corresponds to the torsional motion along proper dihedral angles defined by a sequence of three connected bonds. As usual, it is represented as a superposition of periodic, cosinusoidal wells. The remaining terms define the nonbonded pair potential, and they comprise a Lennard-Jones (6–12) function for the van der Waals interactions (i.e., V_{LJ}) and a Coulombic term for the interaction between locally charged groups (i.e., V_{elec}). All the parameters required to define the force field in Equation 6 depend only on the protein, and not on its environment. For this reason, Equation 6 describes properly the behavior of a single unsolvated protein *in vacuo*. Also, note that the dynamics of internal degrees of freedom can be affected by molecular rotations because the dihedral torsions in Equation 6 couple the bending modes. Thus, the force field allows us to simulate the behavior of a nonrigid molecular rotor, with energy flow between localized and globally extended motions. The simulation starts from the crystal structure of *1hel*-lysozyme, with intact disulfide bridges. The connectivity is maintained during the simulation. We have evaluated 2-ns MD trajectories with an integration step of Δt = 2fs (Verlet algorithm) and using a 13Å distance cutoff for nonbonded interactions. Independent trajectories are generated by changing the set of initial (randomized) velocities at T = 293K. (As discussed before, the resulting constant-energy trajectories are nonstationary, in the sense that they may feature a substantial transfer of energy from internal degrees of freedom to collective vibrorotations.) Unfolding is induced by using *in vacuo* boundary conditions, i.e., free rotations and translations, together with a strong coupling to a Berendsen thermostat[88] at T_B = 293K. The coupling to the bath is characterized by a relaxation constant, τ. For small τ values (i.e., stronger couplings), the rotational kinetic energy energy can become as large as a third of the total kinetic energy.[84] Unfolding can occur as a multistep transition that yields a continuum of elogated transients. Depending on the τ-value used, this transition may not be observed during stimulations of 2-ns. Below, we discuss briefly the typical results. When coupling the above lysozyme model to a Berendsen thermostat with τ = 5fs (a strong coupling), unfolding occurs during all stable trajectories (whose average temperatures are $T = 292 \pm 1$K). Moreover, unfolding appears as a very sharp transition between two states at a time, t^*, along each trajectory. A representative trajectory appears in Figure 2, which shows the change in mean overcrossing number as a function of time. (The displayed snapshots were generated with the program Molmol.[89]) For times $t < t^*$, we find a compressed version of the native state, the so-called *in vacuo* native structure (IVNS).[57,84] *In vacuo* compression is a well-known effect due to the lack of screening interactions.[84] After the transition, we find elongated structures. Figure 3 shows how the folding features evolve along the trajectory by displaying the relative RMS deviation-overcrossing descriptor $\sigma(K(t),K^*)$, where $K(t)$ is the instantaneous conformer along the trajectory and K^* the native state in the crystal. Along this trajectory, the protein adopts a compressed form, drifting toward $\sigma(K(t),K^*) \approx 0.07$ during the

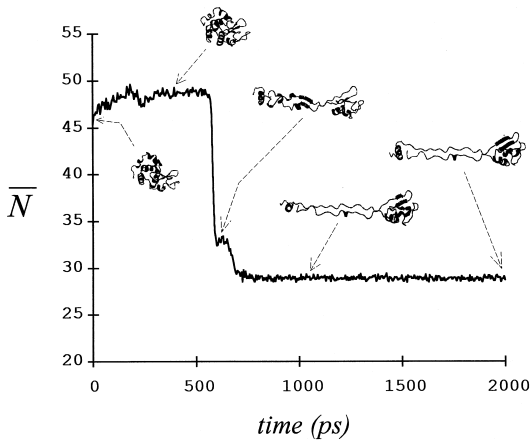

Figure 2. Evolution of chain entanglements along an unfolding trajectory for neutral (disulfide-intact) 1hel-*lysozyme. Centrifugal unfolding elicits under the presence of a strong coupling to a simulated Berendsen bath (τ = 5 fs) at T = 293K. A fast transition takes place at* $t^* \sim 600$ *ps. Selected snapshot configurations are depicted thoughout the trajectory.*

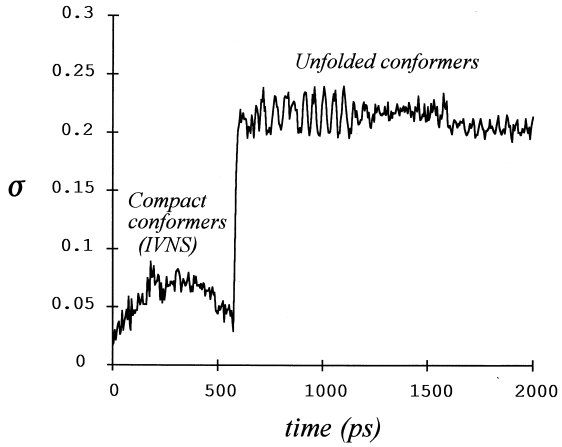

Figure 3. Changes in RMS deviation of overcrossing probability along an unfolding trajectory of neutral lysozyme. (The data correspond to the trajectory in Figure 2. Entanglements are measured relative to the native state of lysozyme in the crystal. Note the occurrence of maximum similarity to the native state just before unfolding.)

first 300 ps. (For comparison, the crossed RMD deviation value for all hen egg-white mutants in the PDB gives $\sigma(K(t),K^*) \approx 0.035 \pm 0.015$.[68]) Afterward, the protein relaxes and swells, adopting quasi-native folding features just before the unfolding transition. This feature is observed throughout all MD simulations. It strongly suggests that the native fold acts as an attractor, in the sense that the *in vacuo* compressed "state" can only relax back into the native structure and cannot unfold directly via a path that bypasses the latter.[69] When using the strong coupling (τ = 5fs), the onset of unfolding shows a small dispersion over a series of trajectories. Figure 4 illustrates this effect by collecting the results of ten independent unfolding trajectories (including the one highlighted in Figure 3). From this set, we estimate that onset of unfolding at $t^* \approx 600 \pm 100$ ps. Figure 4 shows also that the "unfolded state" is an ensemble of similarly elongated conformers, yet can exhibit a variety of folding features compared with the native state. The variations in entanglement complexity are associated with the amount of residual structure in the unfolded conformers. The above protocol for "centrifugal" unfolding is robust over variations in the initial distribution of velocities, as well as in the relaxation constant τ. Qualitatively, the same structural transitions are observed when the τ constant is changed. However, the distribution of partly unfolded intermediates, as well as the time for the onset of unfolding t^*, depends on τ. As τ increases, t^* is delayed (to the extent that it may not be observable with trajectories of 2 ns.). The persistence of partly unfolded intermediates increases with τ. Figure 5 illustrates this observation with some conformational snapshots selected during a MD trajectory with τ = 40 fs. The higher τ value entails a weaker coupling to the thermal bath, and thus a reduced possibility for fast "spinning" of the molecule. The initial velocity distribution is the same as that used in Figure 2, for τ = 5 fs. Figure 5 gives the time for each snapshot, accompanied by the corresponding $\bar{N}$ and Ω values. During this trajectory, lysozyme remains in quasi-native fold as late as t = 800 ps. At $t \approx 1000$ ps, the protein reaches structures that resemble to those observed

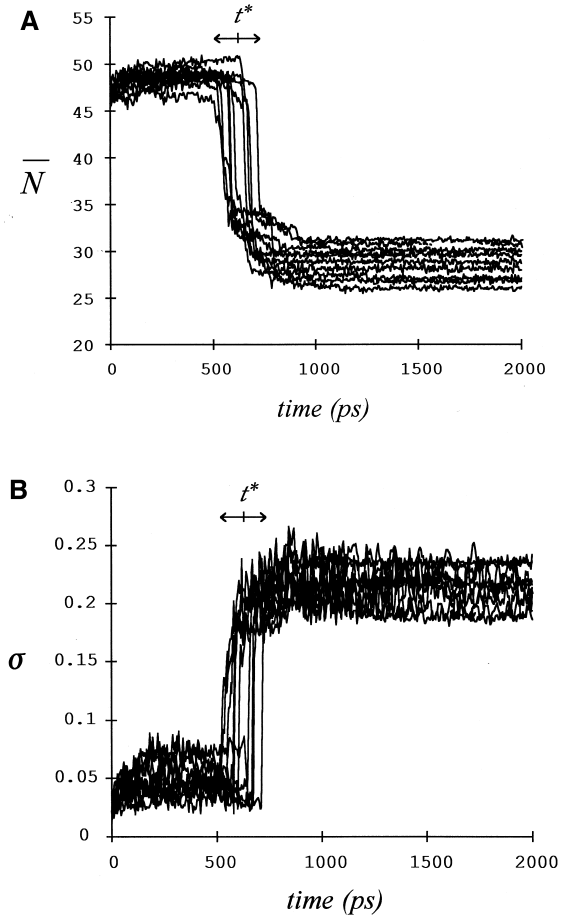

Figure 4. Ten MD trajectories of neutral lysozyme undergoing centrifugal unfolding (τ = 5fs). Each independent trajectory starts from the native state in the crystal, but differs in the ensemble of random initial velocities, from a distribution at T = 293K. The top diagram shows the evolution of the mean number of overcrossings. The lower diagram corresponds to the changes in the relative measure of entanglement. Note the consistency of a transition at t* ~ 600 ps.*

before the sharp unfolding transition for τ = 5 fs. However, when using τ = 40 fs, the protein remains in partly unfolded structures until $t \approx 1400$ ps. Full unfolding takes place only late during the trajectory, as indicated by the last snapshot in Figure 5 at t = 1800 ps. Despite differences in persistence in folding intermediates, the shape of the observed conformers appears to depend little on τ. This point is apparent in Figure 6, which compares the paths on the $(\bar{N},\Omega)$-map. The top diagram (Figure 6A) shows the 2D map corresponding to the ten trajectories in Figure 4 computed with τ = 5 fs. Despite their individual differences, the trajectories span a common swath of shape space, with relatively small dispersions in entanglement and asphericity. The lower diagram (Figure 6B) shows the shape map for seven trajectories differing only in τ value (i.e., having the same initial distribution of velocities). The MD trajectories correspond to τ values of 5, 10, 20, 40, 60, 80, and 100 ps. A comparison of Figures 6A and Figure 6B yields two important observations:

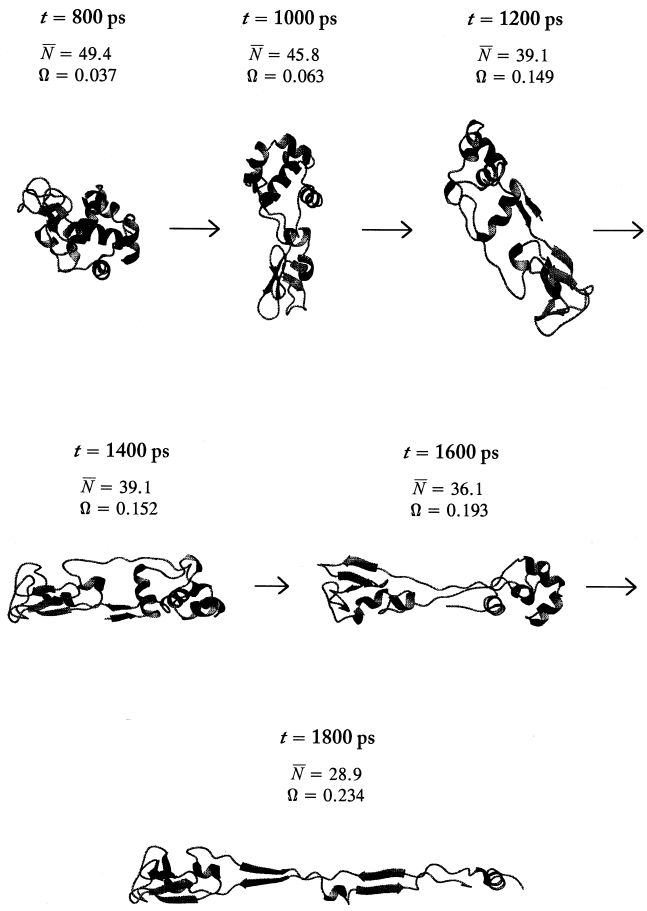

Figure 5. Selected configurational snapshots along an MD trajectory of lysozyme with $\tau = 40$ fs. The native state is stable until t $\sim$ *800 ps. Between 1000ps and 1600 ps, the protein is found in persistent intermediate structures, not found in Figure 4. The last snapshot (*t = *1800 ps) corresponds to a fully unfolded conformer.*

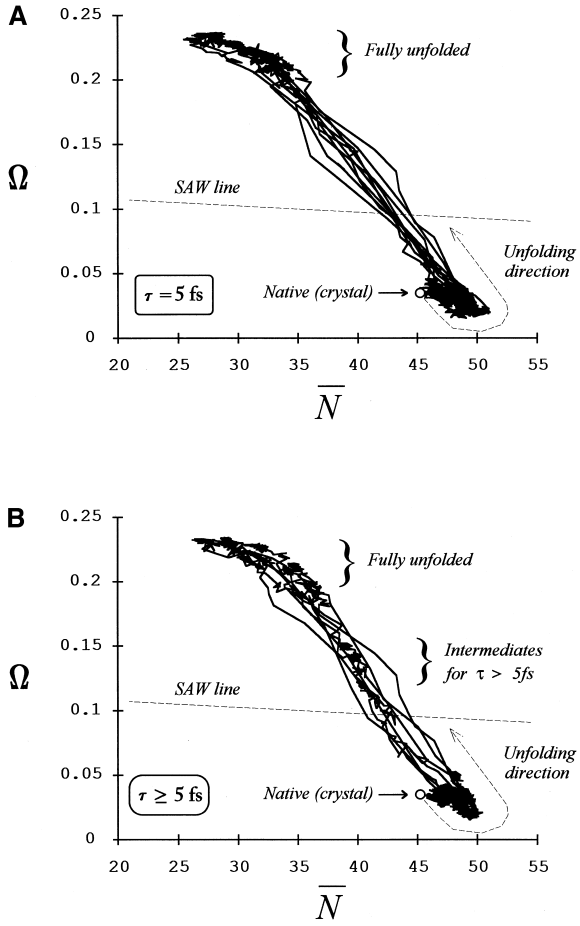

Figure 6. Variability in unfolding behavior for neutral lysozyme. (A) shows ten independent trajectories for $\tau = 5$ fs. (B) corresponds to seven MD trajectories with the same initial distribution of velocities, but with distinct τ values (see text). Different relaxation constants, τ, change the persistence of intermediates, but not the nature of the unfolding process.

1. There are no persistent intermediates between the IVNS and the fully unfolded conformers for $\tau = 5$ fs (Figure 6A). In contrast, these structures appear for $\tau > 5$ fs, and they are located above the limit of compact conformers (Figure 6B). Note that the latter intermediates stay "on track," i.e., they are found along the same paths spanned by the trajectories for $\tau = 5$ fs value, which leads to the unfolded state with no persistent intermediates structures.
2. The swath of molecular shapes spanned by a manifold of MD trajectories with strong coupling to the thermal bath is essentially identical to the one observed for trajectories with other weaker couplings. The occurrence of a common pattern on the $(\bar{N},\Omega)$ map suggests that centrifugal unfolding of lysozyme is characterized by a well-defined, uniform mechanism, regardless of the strength of the bath relaxation constant. In addition, comparing Figure 1 and Figure 6 suggests that the unfolding mechanism follows a path somewhat intermediate to the hypothetical trajectories (1) and (2) (with a reversed time arrow). In other words, centrifugal unfolding of lysozyme at room temperature takes place as a concerted process where the protein loses compactness at the same time as it disentangles its local secondary structure. With its reversed direction, such a process also resembles the conjectured "optimal" folding paths where the changes in compactness and in folding features are believed to be synchronous.[34,90]

5. UNFOLDING OF LYSOZYME IONS

The occurrence of unfolding and refolding processes for protein ions in gas phase is well characterized experimentally.[17,18] Unfolding is believed to be driven essentially by Coulombic repulsion between charged amino acids, although the process may be compounded with the occurrence of heating under experimental conditions of electrospray. Lysozyme ions constitute a convenient system to explore the role of Coulombic repulsions on the onset of unfolding. Their behavior can then be contrasted with that of neutral lysozyme discussed previously. Experiments on charged proteins are consistent with the occurrence of a critical charge that triggers unfolding, q^*, a process that is also known to take place in simpler polyelectrolytes.[91–95] Within a narrow range of charge values, $q^* \pm 1$,

proteins can appear in both folded and partly unfolded configurations.[17,96,97] Fully-denatured conformations appear as high-charge states, where $q > q^* + 1$. In contrast, low-charge states, where $q < q^* - 1$, remain in compact quasi-native folds. When lysozyme is electrosprayed from room-temperature acidic solutions under experimental conditions, the observed value is $q^* \approx 9 \pm 1$.[9,12] Recent MD simulations provide a framework to understand the origin of this value.[98,99] Below, we describe these simulations and their associated patterns of molecular shape transitions. The protocol for the simulation of *in vacuo* unfolding of lysozyme ions is as follows.[82] As in Section 4, the initial structure for all MD simulations is the (energy-minimized) native form of hen egg-white lysozyme, with intact disulfide bridges. In this case, the MD trajectories are generated with the GROMOS87 37C4 force field, designed for the simulation of biomolecular ions regardless of their environment.[86] When combined with *in vacuo* boundary conditions for free rotation and translation, this force field represents the state of an anhydrous protein ion. The simulation specifications include no periodic conditions, and Coulombic interactions between charged groups are computed for all distances.[82] After the first 5 ps, the molecule is weakly coupled to a Berendsen thermostat at a temperature T_B (with $\tau = 0.1$ ps), and data are collected during 3 ns. Unfolding is controlled by the temperature and the charge distribution on the molecule; thermal coupling alone does not induce unfolding under the stated simulation conditions. For a given charge state and temperature, we have computed several MD runs, differing in the distribution of initial random velocities at T_B. Modeling the behavior of protein ions has an added complication: the large number of possible species that can share the same total charge q. In principle, different distributions of charges may elicit distinct unfolding patterns. An approximate, average unfolding behavior can be modelled with a distribution of charges where all possible chargeable sites contribute equally to the total q value.[82] We refer to this model as a statistical charging scheme. For clarity, the symbol q_s indicates when the total charge has been computed with this model. The distribution of individual charges (adding up to q_s) is determined with respect to a state of nominal maximum charge $q = 19$. This charge state is a reasonable approximation to the condition of lysozyme at pH $\sim$ 3, where all basic sites are protonated (eleven arginines, six lysines, one histidine, and the N-terminus).[100] In the statistical model, a state with total charge $q_s < 19$ is derived by assigning a fractional charge $q_s/19$ to each of the charged sites.[82] The result is an "effective" ion, whose dynamical behavior mimics that of an average over all possible "realistic ions" (i.e., those with integer charges distributed among 19 sites). The statistical model constitutes a first approximation to account for proton mobility, believed to occur in protein ions under the experimental conditions of migration in a low-pressure gas.[17] When using a statistical approach to the charge distribution, lysozyme undergoes fast unfolding at $T_B = 293$K for charges $q_s \geq 16$ [82,99]. This behavior gives a critical charge for unfolding that is larger than the experimental one ($q^* \approx 9$). The result is not surprising given that, under experimental conditions, protein ions probably emerge at higher termperatures due to heating at the electrospray needle.[98] This is consistent with ion-drift mobility experiments on cytochrome *c*, showing a decrease in q^* as the injection energy of the jet increases.[17] As well, MD simulations on the cytochrome *c* ions with $q = 7$ indicate that it persists in quasi-native form at 300K, but it is fully unfolded at 600K.[17] In our case, we discuss below the change in critical charge for lysozyme unfolding when the protein is simulated at $T_B = 500$K.[101] Independent trajectories for lysozyme ions in various charge states were computed by MD simulations with different distributions of initial velocities at 500K.[98,101] The typical behavior obtained is illustrated by Figure 7, which shows the evolution of chain entanglements in single trajectories for lysozyme ions with $q_s = 5$, 7, 8, and 11, started with the same initial velocities. As the figure shows, the native fold is conserved (albeit in a somewhat more compact form) at $q_s = 5$, but complete unfolding into elongated structures takes place rapidly for $q_s = 11$. (The structures displayed on the left-hand side correspond to snapshots found at the end of the trajectories.) The ions with charges $q_s = 7$ and 8 exhibit an intermediate behavior, whereby full unfolding does not take place but the protein remains in partly unfolded conformations. These differences are clearly evident in Figure 8, which collects the data for several MD runs of lysozyme ions. The diagram of Figure 8A shows the results for the trajectories in high- and low-charge state. The simulations indicate that unfolding takes place rapidly for ions with $q_s \geq 9$, whereas ions with $q_s \leq 6$ remain in slightly compressed native folds. No persistent partly-unfolded intermediates are observed at these total charge values. This behavior can be contrasted with that for the $q_s = 7$ and 8 charge states (cf. Figure 8B). Depending on the individual trajectory, lysozyme adopts a continuum of structures for $q_s = 7$ and 8, ranging from quasi-native to the fully unfolded. These protein ions are in a state where folded and unfolded conformer populations coexist. This behavior is a signature of the critical charge for unfolding. These results are summarized in Figure 9, which provides the configurational "phase" diagram for lysozyme ions in variable charge q_s at 500K. This figure indicates the region of $(\bar{N},\Omega)$ space associated with long-lived molecular shape features. (That is, Figure 9 excludes regions spanned by short-lived transients in $\bar{N}$ and Ω values.) The diagram reveals the interrelation between total charge and accessible folding features. Compact conformers in quasi-native fold are found only for $q_s \leq 8$. As discussed before, the native fold is the only accessible molecular shape for $q_s \leq 6$, but it is also compatible with $q_s = 7$ and $q_s = 8$. Similarly, noncompact structures are

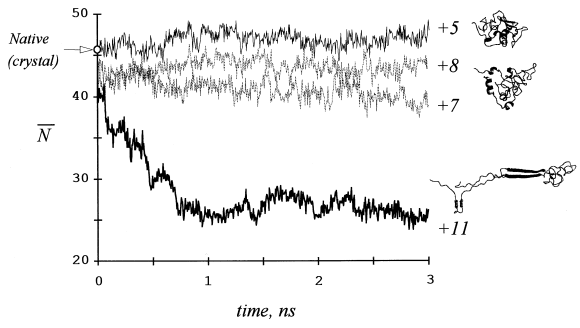

Figure 7. Changes in chain entanglement for lysozyme ions at 500K. The lines are single MD trajectories generated with identical initial conditions but different charged states, $q_s = 5, 7, 8,$ *and 11. Snapshots illustrate representative conformers observed at the end of the trajectories for* $q_s = 5, 8,$ *and 11. Coulombic repulsion causes full unfolding for* $q_s = 11$.

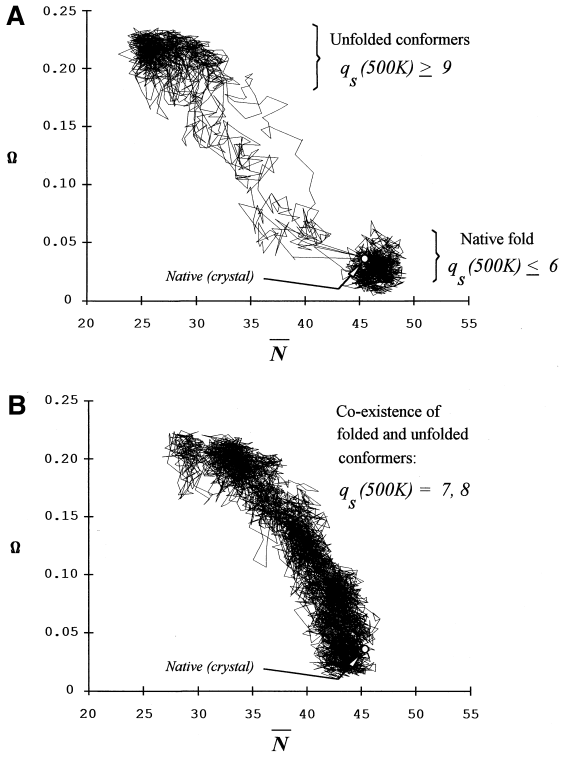

Figure 8. Molecular shape map for the in vacuo *dynamics of various lysozyme ions at 500K. (A) gives the behavior for high-charge ($q_s \geq 9$) and low-charge ($q_s \leq 6$) states. (B) diagram depicts the results for $q_s = 7$ and 8. These latter states mimic the critical charge for unfolding, q_s*, since they show co-existence of folded (compact) and unfolded intermediates.*

compatible with $q_s = 7$ and 8, but are compulsory for $q_s \geq 9$. Partly unfolded conformers (found near the SAW line) are found only for $q_s = 7$ and 8. These latter charges are close to the experimental value of critical charge for lysozyme unfolding. The agreement suggests that, under experimental conditions, lysozyme ions have effective temperatures over 293K.

The MD simulations are also consistent with two other pieces of experimental data:

1. Folded and unfolded conformers co-exist within a narrow range about the critical charge, $q^* \pm 1$. Outside this range, the protein is either folded or unfolded.
2. The unfolded conformers are elongated structures with characteristic maximum lengths of 120Å. This value agrees with the results of the simulations, which indicate maximum radial spans of ~65Å.

It is interesting to contrast the unfolding behavior of highly-charged lysozyme ions with the centrifugal unfolding of neutral lysozyme. Figure 10 shows two representative MD trajectories that illustrate the difference. The dashed line allows us to observe the unfolding of lysozyme ions with $q_s = 11$ at $T_B = 500K$. The solid line shows the unfolding of neutral lysozyme at $T_B = 293K$, using a strong coupling to the thermal bath ($\tau = 5$ fs). Some important differences are apparent:

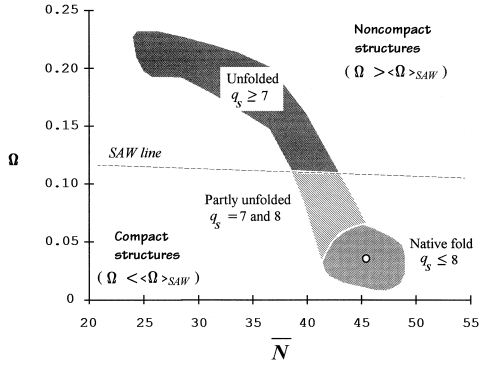

Figure 9. Configurational "phase" diagram for conformers of lysozyme ions at 500K. Conformers are classified as compact or noncompact according to their position with respect to the "SAW line." Compact conformers with quasi-native fold have fluctuations of 4Å in RMS deviations. The overall form of the diagram is the same as that for simulations conducted at T = 293K. However, the critical change for unfolding increases to +15 at room temperature.

1. Although the unfolded structures formed in both cases are qualitatively similar in molecular shape, the unfolded ions are "floppier" structures with larger configurational fluctuations.
2. The initial stages for unfolding are different. In neutral lysozyme, the initial compactization *in vacuo* is followed by a transition where a loss of sphericity is accompanied by a reduction in loop entanglements. In contrast, the unfolding of high-charge lysozyme ions is initiated by a fast reduction in entanglements, with little effect on the mean anisometry of the chain.
3. In lysozyme ions, full unfolding (after the initial stage, 2) is usually accompanied by structural transitions involving rapid increases in asphericity, with little effect on chain

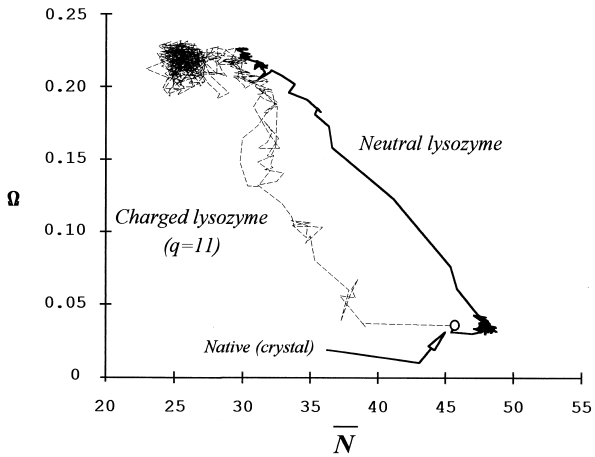

Figure 10. Comparison between the unfolding mechanisms for neutral lysozyme (thick line) and a lysozyme ion in a high-charge state (dashed line). Note that the final unfolded conformers are similar, despite differences in the path and the extent of the configurational fluctuations.

entanglement. Such a process is never observed during the unfolding of neutral lysozyme.

These differences suggest that the unfolding of lysozyme ions follows a different mechanism from that observed during the centrifugal unfolding of the neutral protein. For ions, the initial reduction in $\bar{N}$ (at constant Ω) suggests a process controlled by the reorganization of local folding features. The initial effect of the Coulombic repulsion is akin to a "meltdown" of secondary structure (represented as a decrease in $\bar{N}$), followed by a repulsion of charged loops (associated with an increase in Ω). With a reversed time arrow, such a process resembles qualitatively the hypothetical two-step path (2) in Figure 1. These characteristics are absent in the centrifugal unfolding of neutral lysozyme. Finally, we note that the observed differences in flexibility for the unfolded transients are consistent with the notion that the "unfolded state" is a diverse ensemble of structures, the properties of which can vary according to the denaturing conditions.[74,76,83,101,102]

6. UNFOLDING TRIGGERED BY WEAKER NONBONDED PAIR INTERACTIONS

As seen in the previous section, there is a charge threshold at which Coulombic repulsion can trigger unfolding in a protein ion. For a given protein, the value of the critical charge q^* depends not only on the temperature and the model distribution of charges, but also on the monomer–monomer interactions incorporated into the force field. In principle, the attractive interaction between nonbonded pairs of atoms plays an opposite role to the Coulombic repulsion. Below, we discuss briefly the interplay between these two interactions and their effect on the onset of the unfolding transition.[103] Force fields incorporate nonbonded pair potentials to account for the weak van der Waals interactions between monomers. Commonly, this potential is represented by a Lennard-Jones function, exhibiting repulsive and attractive interactions. Variations in the relative strength of these two "forces" can be used as a simple model to account for shielding and environmental effects. In our case, we explore the effects associated with changes in the attractive (nonbonded) pair potential, while keeping constant all other energy terms (i.e., stretching, bending, torsions, and electrostatics).[57] We modulate the Lennard-Jones interaction in GROMOS87[86] by introducing a scaling factor, f, into the attractive branch of the potential (cf. Equation 6):

$$V_{LJ}(r_{ij}) = S(r_{ij}) \left\{ \frac{A_{ij}}{r_{ij}^{12}} - f \frac{B_{ij}}{r_{ij}^{6}} \right\}, \qquad f > 0, \qquad (7)$$

where r_{ij} is the distance between the i-th and j-th atom, and $S(r_{ij})$ is a cutoff function. Neither $S(r_{ij})$ nor A and B change during the MD simulations. When $f = 1$, we find the standard GROMOS87 potential. By changing f, we modify the protein's ability to keep its native fold. In a first approximation, the variation of the attractive interaction provides a crude model for the environmental effects associated with solvents that either favor folding or denaturation. This approach also provides a simple test to gauge the qualitative dependence of the folding behavior with respect to the parameters included in the model force field. Results in the literature suggest that, depending on the strength of the monomer–monomer interaction, a change in chain compactness may be coupled to a change in secondary structure.[24] Below, we discuss that a related effect can be seen in the unfolding behavior of protein ions. In particular, we show that unfolding is possible at subcritical charges ($q < q^* - 1$) by introducing a less attractive nonbonded pair potential (i.e., using $f < 1$ in Equation 6). As an example, consider the case of lysozyme ions at T = 293K. At this temperature, the critical charge for unfolding is $q_s^* \approx 15$, within the statistical charging scheme. $q_s = 13$ represents a protein ion with subcritical charge, i.e., it exhibits no unfolding under the simulation conditions discussed in Section 6. Yet, unfolding can take place if the simulations are repeated with $q_s = 13$ and a weaker van der Waals interaction. Figure 11 illustrates this behavior by contrasting the accessible shape

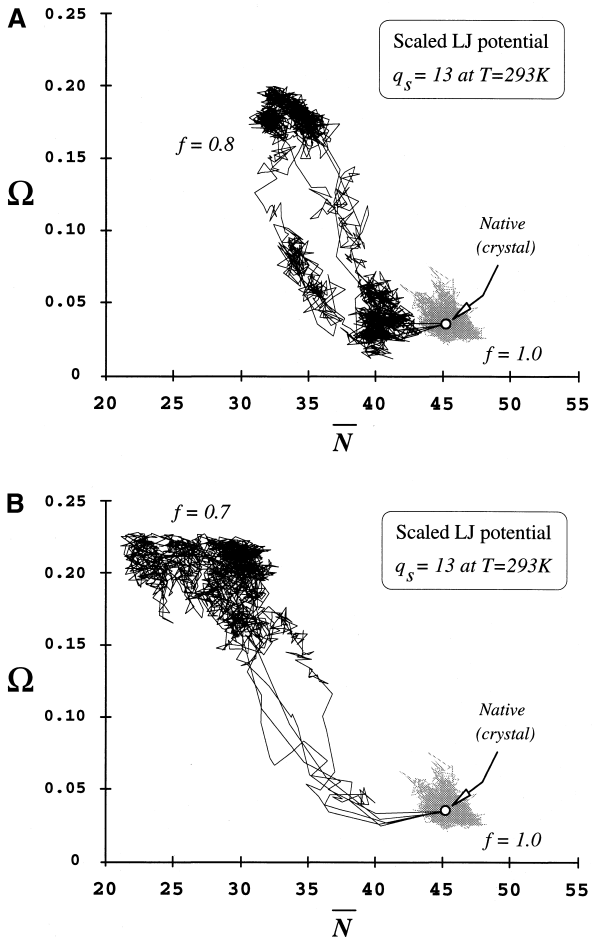

Figure 11. Effect of a reduction in the attractive Lennard-Jones nonbonded pair interaction. (A) compares the results for f = *1 (standard potential) and for* f = *0.8 (a potential with weaker attraction). The charge* q = *13 is subcritical and does not unfold for* f = *1. However, a population of both partly-unfolded and unfolded conformers appear for* f = *0.8. (B) shows a similar comparison between* f = *1 and* f = *0.7 (very weak attraction). In this latter case, unfolding proceeds fast without any intermediates. The unfolding path resembles those for ions in high-charge state and* f = *1. Results depicted include five MD trajectories for each* f *value.*

features observed MD runs of 3 ns with $q_s = 13$ and various f values. Figure 11A compares the results of five independent MD runs with $f = 1$ and $f = 0.8$. There is no unfolding for $f = 1$. Since $q_s = 13$ is just below the critical charge, this protein ion is precisely stabilized around the native fold, with no compression due to the *in vacuo* condition. (This latter effect is found for $q_s < 13$ and $f \geq 1$.) In contrast, the $q_s = 13$ ion is unstable toward unfolding if the monomer interaction is weaker. Figure 11 shows that a 20% reduction in attraction (i.e., $f = 0.8$) yields a population of partly unfolded and unfolded conformers. When using $f = 0.8$, the protein does not maintain the native fold. Initially, it drifts toward less entangled, yet spheroidal objects. Depending on the individual trajectory, these intermediates may persist for at least 3 ns or lead eventually to unfolded conformers (characterized by $\bar{N} \approx 35$ and $\Omega = 0.18$). When using a slightly stronger attraction ($f = 0.9$, data not shown), we find an even population of intermediates, stretching from the native state up to unfolded conformers. This example suggests that a less attractive interaction initially relaxes the native state toward conformers with similar Ω, but smaller $\bar{N}$ values. This effect resembles the unfolding of protein ions with $q_s > q_s^*$ and $f = 1$. The analogy is clearer with the results in Figure 11B for $q_s = 13$. The diagram contrasts the native state ($f = 1$) with the relaxation associated with $f = 0.7$ (five trajectories). A 30% reduction in attraction leads to fast unfolding through paths that resemble those for high-charge states (cf. Figure 10, dashed line). If we compare $f = 0.7$ and $f = 0.8$, it is clear that the main differences are in the persistence of intermediates along the unfolding path. However, the overall pattern of molecular shapes found along these paths is similar. The above results are representative of the behavior for other q_s and f values. Our main observations are:

1. A protein ion at subcritical charge value, $q \approx q^* - 1$, has the stability of an ion with charge q^* if its van der Waals attraction is reduced by ca. 10%.
2. Further reduction in the attraction leads to unfolding by paths initially controlled by a reduction in entanglement.
3. The population of unfolded and partly unfolded intermediates depends on the interaction. A 30% reduction in attraction is incompatible with partly-unfolded conformers at subcritical q.

These observations suggest that the unfolding mechanism is still controlled by the effects of Coulombic repulsion discussed in Section 5. Weakening the monomer–monomer interaction plays a role similar to raising the temperature of the simulation: unfolding can proceed at lower q values, yet by essentially the same reaction paths.[103] These results are consistent with the notion that the main features of protein dynamics are determined by broad properties, rather than the detailed structure, of the potential energy.[24–34,101,104] In other words, the results are not strongly dependent on the precise values of the key parameters in the model potential energy function. Small variations in the force field parameters appear to affect only the relative population of intermediates and their persistence over time, but not the nature of the conformers visited during the trajectory.

7. RELAXATION DYNAMICS FROM THE UNFOLDED STATE

We have shown that disulfide-intact lysozyme can unfold by following mechanisms that, though different, lead to qualitatively similar structures. The MD-generated unfolded conformers are elongated, with mean cross sections comparable to the experimental ones.[20] For this reason, we conjecture that these structures provide a reasonable representation for the "unfolded state" of this protein. Accordingly, we can use these conformers as initial "seeds" to perform studies of relaxation (and possibly refolding) dynamics. Proteins are known to relax from the unfolding state by avoiding a random search of configurational space.[105–109] Instead, folding dynamics can be better described as a nonrandom process involving a manifold of possible paths over a free-energy surface with a bias toward the native state.[24–32,107–109] In principle, this ensemble can include qualitatively different paths, with some of them leading rapidly to quasi-native states,[110,111] while others trap the protein at configurationally frustrated local minima.[31,32,109,112] Depending on composition and chain length, rapid folding can either take place along paths with long-lived intermediates, or no detectable intermediates.[56,61,101,102,113] Despite their variability, it is believed that all folding paths can be characterized in terms of a few dynamical properties. As discussed before, two main features are the formation of a compact globule and the development of the correct chain "topology." These notions were originally developed within the context of solvated protein dynamics, where solvent quality is related to the onset of the collapse transition.[22–34] However, recent work on anhydrous proteins has shown that refolding transitions can also occur *in vacuo*.[17,18,57,58,114] Thus, *in vacuo* protein relaxation becomes a key phenomenon in understanding the specific role of solvent in driving configurational transitions. Below, we discuss briefly the insights provided by MD simulations of unfolded lysozyme.[57,58,114] These simulations shed light on the interplay between chain compactization and the formation of tertiary structure during *in vacuo* folding.

7.1. Relaxation of Disulfe-Intact Lysozyme

We consider here the relaxation behavior of unfolded, disulfide-intact neutral lysozyme. To this end, structures found along MD trajectories with centrifugal unfolding (Section 4) are taken as initial conformers for relaxation studies. (Unfolded lysozyme ions remain as such unless their total charge is reduced below the critical charge for unfolding.) From these structures, relaxation starts by removing the unfolding conditions.[57,58,114] This implies reassigning random initial velocities to the unfolded transient, and reapplying *in vacuo* boundary conditions with a weak coupling to a Berendsen thermostat. This latter condition reduces the possibility of reunfolding. Below, we summarize the main findings from a large ensemble of trajectories.[115] In all cases, we use MD runs of 1.2 ns; this is sufficient time to detect the formation of persistent folds during *in vacuo* relaxation of proteins.[57,58,114] Figure 12 illustrates how chain entanglements evolve during relaxation.[58] The figure shows the original unfolding run, corresponding to an MD trajectory with centrifugal unfolding for $\tau = 10$ fs (thick line).[84] On this run, three initial transients selected for relaxation are indicated by A, B, and C, together with their times along the original trajectory. These transients, differing in 5 ps, are consecutive snapshots extracted during the unfolding transition. The thin lines starting from these structures correspond to the relaxation runs. The latter are indicated as *Rxxx*, denoting a relaxation initiated from the unfolded transient found at $t = xxx$ (in ps) along the unfolding run. The three examples show

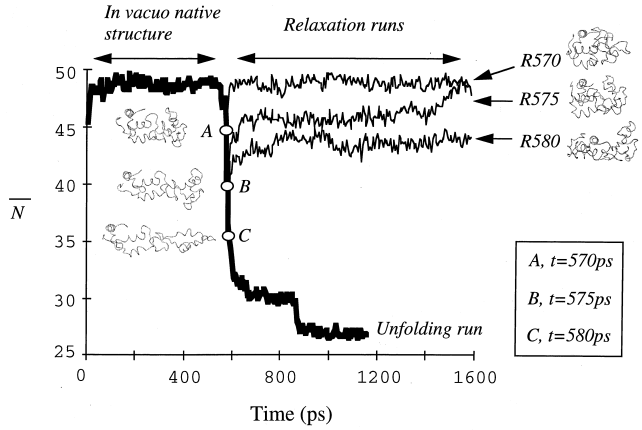

Figure 12. Variation in mean overcrossing number along the unfolding trajectory (thick line) and selected relaxation trajectories (thin lines) for disulfide-intact lysozyme. A, B, and C represent initial transients found along the unfolding run and used for relaxation. (Reprinted with permission from Arteca et al.[58] Copyright of the American Physical Society.)

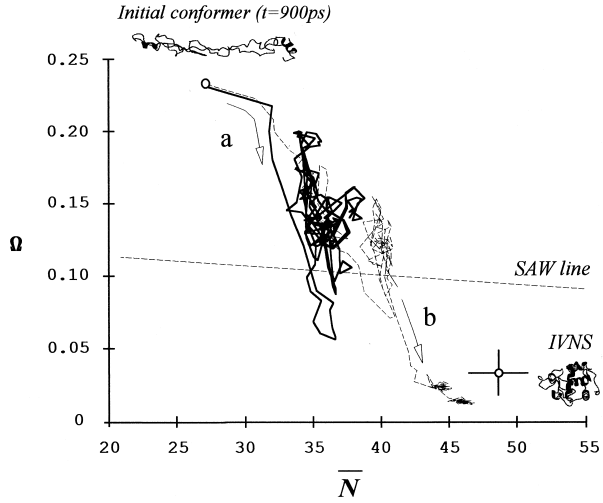

Figure 13. Comparison between two relaxation runs for neutral lysozyme, starting from the same fully unfolded conformer. Note the occurrence of partial chain collapse up to regions near the SAW line. One trajectory exhibits configurational frustration at noncompact conformers (a, thick line), whereas the second one succeeds in refolding toward compact conformers (b, dashed line).

that one can expect a wide range of behaviors from the ensemble of accessible relaxation paths. We find refolding to the native structure (*R570*); relaxation to an intermediate, which then drifts toward a quasi-native structure (*R575*); and trapping at a partly folded, denatured intermediate (*R580*). In all cases, the trajectories show an increase in chain entanglement during relaxation. This is accompanied by the formation of a compact conformer as indicated by an increase in asphericity. This is not an unexpected result when starting the relaxation from an elongated unfolded transient. A more important issue is whether these trajectories reflect generic relaxation dynamics, i.e., a cascade of events that is accessible to most polymer sequences, or a sequence-specific, well-defined protein-folding process. The pattern emerging from an ensemble of relaxation trajectories suggests that the latter process does also take place *in vacuo*.[57,58,114] Let us consider in detail the relaxation of fully unfolded conformers in terms of the familiar two-dimensional map of asphericity and chain entanglement. Figure 13 shows two independent *R900* trajectories, starting from the elongated conformer found at t = 900 ps along the unfolding run in Figure 12. These trajectories exhibit some characteristic features:

1. Along the "a" run (thick line), the asphericity decreases initially with little change in chain entanglements. During this step, conformers reach Ω values below the SAW line, but do not persist there. After the initial partial compactization, the protein relaxes and undergoes large configurational rearrangements as a nonspheroidal object. On this trajectory, the protein remains as nonspheroidal during the 1.2 ns of the run without recompactization.
2. Along the "b" run (dashed line), the initial behavior also involves partial compactization toward objects more compact than SAWs, followed by relaxation to less compact objects. In contrast to case "a," the protein in case "b" eventually finds a path that leads to the formation of very compact objects within 1.2 ns. The final conformers are characterized by small configurational fluctuations and by molecular shape features comparable to those of the IVNS on the ($\overline{N}$,Ω) map. (Their three-dimensional fold is, however, not necessarily native.)

These results are representative of a behavior observed throughout all simulations of lysozyme. Figure 14 collects twenty independent relaxation trajectories, including the two highlighted in Figure 13, generated with different initial velocities. This ensemble of trajectories provides a good representation of the path variability accessible to a fully unfolded conformer. Figure 14 indicates a high population of early transients with asphericity near the SAW line, with only small increases in entanglement complexity ($\overline{N} \sim 35$). In no case do

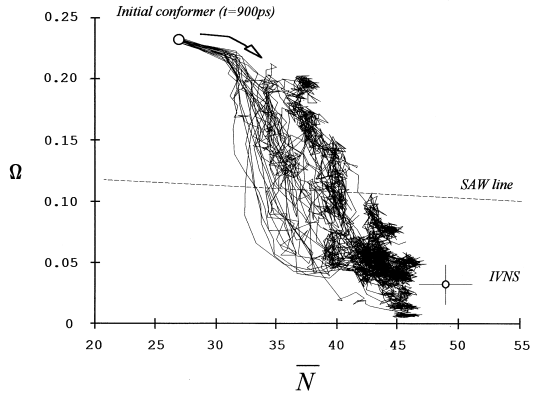

Figure 14. Twenty MD relaxation trajectories for the R900 unfolded transient of neutral disulfide-intact lysozyme. Note the incidence of noncompact persistent intermediates. However, most of the persistent structures have compactness similar to the IVNS. The behavior suggests an overall folding process initiated by partial protein compactization.

we observe the rapid formation of fully collapsed structures. Most of the persistent structures are indeed compact (i.e., found below the SAW line), but they are formed after at least 200 ps of dynamics. On the other hand, there is a substantial population of noncompact intermediates that can persist for as long as 1 ns. In the particular case of *R900* relaxations, most of the conformers found near the end of the trajectories have Ω values comparable to those of the IVNS, with $\bar{N}$-values similar to those of the native state of lysozyme in the crystal, $\bar{N} \sim 45$.[68] The pattern emerging from MD trajectories starting from fully unfolded conformers suggests the following principle[57,58]: relaxation is initiated by an incomplete collapse transition, i.e., one leading to structures characterized by large configurational flexibility and intermediate degree of compactness. From these conformers, a protein can reaccomodate its tertiary structure by undergoing several recompactization "trials," some of which can lead to quasi-native folding. The existence of a common relaxation pattern, despite the configurational noise in the trajectory ensemble, is suggestive of a well-defined folding process *in vacuo*. Our results indicate that the initial stages of this process can be characterized, in a first approximation, as qualitatively similar to the hypothetical path (2) in Figure 1, dominated by the onset of partial polymer collapse. This observation is consistent with the notion that protein collapse is the favored first folding step whenever the effective interaction between residues is strongly attractive.[24,40,41] This condition is compatible with *in vacuo* simulations, which involve no environmental screening. In addition, our results exhibit other features that are consistent with observations in the literature:

1. A variety of relaxation paths are accessible to lysozyme, some of which lead to faster folding than others. A similar behavior has been noted for hen egg-white lysozyme in solution.[110,111]
2. Some relaxation paths feature partial reunfolding, after the initial formation of a more compact structure. A similar observation has been made in the folding of ubiquitin.[76]
3. All relaxation paths exhibit some degree of coupling between chain collapse and partial refolding, in agreement with conjectures on "optimal" folding paths.[34,90]
4. The relaxation of lysozyme *in vacuo* involves both compact and partially compact intermediates. This behavior resembles a "hierarchical" folding process, believed to describe well the relaxation of multidomain proteins with more than 100 residues.[37,61,104]

A comparison between Figure 14 and Figure 6A shows the strong differences between the unfolding and relaxation, respectively, of disulfide-intact lysozyme. In the former case (Figure 6A), there is strong correlation between asphericity and chain entanglement. This relationship suggests that conformational elongation is achieved at the expense of reducing the content of secondary structure. Such a correlation does not occur during the relaxation of fully unfolded conformers. In other words, unfolding-refolding processes do not necessarily exhibit microscopic reversibility. Results in the literature show that reversibility is in fact possible, but only when relaxing from initial structures that are partly unfolded.[57,58,114] Partly unfolded conformers found near the unfolding transition (e.g., *R570* and R575 in Figure 12) can refold toward quasi-native conformers by retracing backwards the unfolding path. Initial configurations selected among conformers found later along the unfolding trajectory lose eventually this capability [58].

7.2. Dependence of Relaxation on Primary Sequence

The results discussed in Section 7.1 support the notion of a well-defined folding process. One important point that remains open is whether such a behavior is intrinsic to the protein, i.e., determined by its primary sequence as well as its initial point in phase space. It is therefore important to estimate how the specific effects of primary sequence are reflected on the refolding transition. One simple approach is to compare the relaxation behavior of "chimeric" proteins, consisting of different primary sequences threaded onto the same unfolded backbone configuration. Below, we provide a simple illustrative example in the case of an unfolded lysozyme conformer. A more systematic study appears elsewhere.[115] Figure 15 compares the relaxation behavior of the same unfolded conformer *R990*, but threaded by two different sequences. In one case, we have the original sequence of disulfide-intact lysozyme, for which *R990* is a proper unfolded transient. This conformer is an elongated structure, similar in shape to the *R900* discussed before. The corresponding relaxation behavior is depicted in Figure 15, using twenty trajectories of 1.2 ns (dashed lines). Qualitatively, the relaxation pattern resembles the one for *R900* (cf. Figure 14), characterized by early noncompact intermediates and persistent compact structures. This behavior can be compared with the relaxation of a polyglycine backbone with the same initial *R990* configuration (full line). The threading was performed by "mutating" the original sequence to glycine residues (i.e., by eliminating all side chains),[115] using the molecular modelling program TOM/mdfrodo.[116] The resulting structure was reoptimized with a "frozen" backbone geometry. In fact, adapting the original unfolded conformer to alternative

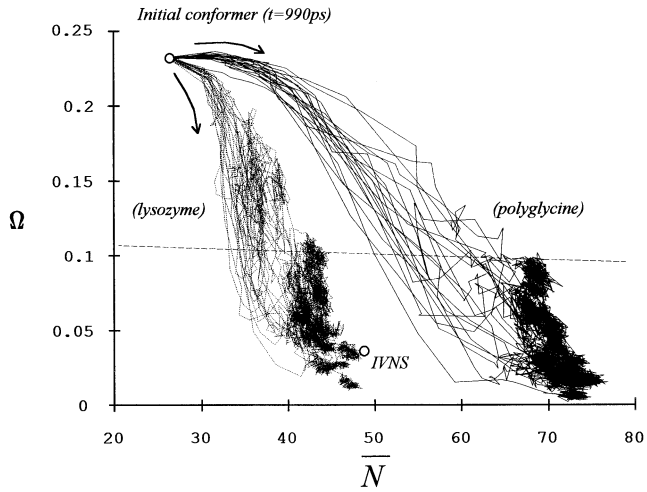

Figure 15. Role of primary sequence on the relaxation behavior. The figure shows the results of twenty MD trajectories for lysozyme and a polyglycine chain, both starting from the same R990 *unfolded transient of disulfide-intact lysozyme. The polyglycine chain leads to a continuum of compact structures with no intermediates. Its behavior is akin to the hypothetical path (1) in Figure 1, describing unfolded homopolymers undergoing a "pearling" transition. The relaxation of the* R990 *transient resembles that of the* R900 *transient (cf. Figure 14).*

sequences involves a limited amount of conformational stress. The latter can easily be accomodated with local motions, as also noted in the literature for different systems.[117] Also, polyglycine is known to be "adaptable" to a variety of native folds.[118] As shown in Figure 15, the MD trajectories for polyglycine lead very rapidly to compact stable structures with significant chain entanglement. Note that its final persistent conformers span the entire range of asphericity values below the SAW line, with small differences in entanglement. All trajectories follow the same pattern in ($\overline{N}$,Ω) space, consistent with a fast variation from unfolded to folded states with no short-lived intermediates in between. The polyglycine paths suggest a uniform mechanism leading to polymer collapse. The behavior resembles the hypothetical path (1) in Figure 1, dominated by an initial growth of local tertiary structure without global compactification. Path (1) was believed to mimic the collapse transition for simple homopolymers; our results show that polyglycine resembles these systems. Here, the lack of side chains allows an initial gain in folding complexity by forming denser local neighborhoods, i.e., a "pearling transition."[73] This is followed by a gradual compactization of the chain. This behavior is in marked contrast with that of unfolded lysozyme (cf. Figure 15), whose relaxation follows a pattern more similar to the hypothetical path (2) in Figure 1. In summary, the above simulations show, first of all, that polypeptides without well-defined native structure can indeed produce compact structures *in vacuo.* In polyglycine, the relaxation consists indeed of a single step: the formation of a compact chain. We find no evidence of configurational fluctuations involving partly folded intermediates. Although this situation can appear in polypeptides other than glycine,[115] the result illustrates that MD simulations of relaxation reflect correctly the dependence on sequence. A well-defined folding process is associated with a given primary structure. The occurrence of such a process is not a given for all conceivable sequences.

8. SUMMARY AND FINAL REMARKS

We have reviewed several recent developments on the behavior of *in vacuo* proteins. Although the experimental characterization of these systems is still limited, the evidence points clearly toward the occurrence of unfolding and refolding structural transitions when proteins and protein ions migrate in a low-pressure gas or in a vacuum. Understanding the nature of *in vacuo* transitions is an important step toward solving the problem of protein folding, in particular toward clarifying the specific role of the solvent. In this context, computer simulations provide a valuable tool to interpret and complement experimental observation. By using *1hel*-lysozyme as a working example, we have illustrated the type of insights provided by MD simulations, together with a characterization of molecular shape that conveys the backbone geometry and topology. In the particular case of disulfide-intact lysozyme, MD simulations are consistent with a number of experimental results:

1. Unfolded conformers are elongated structures, with characteristic radial spans of approximately 65Å.
2. Unfolding takes place as a multi-step transition, leading to the formation of a variety of elongated conformers.
3. There is a critical charge for unfolding lysozyme ions, which is $q^* \approx 9 \pm 1$ under the experimental conditions of heated electrospray.
4. Within the range of the critical charge q^*, compact and noncompact conformers coexist.
5. Refolding toward compact conformers can take place spontaneously *in vacuo,* for neutral or low-charge species initially unfolded.

This agreement between theory and experiment provides the framework that allows us to conjecture some of the microscopic features that characterize the unfolding and refolding paths. Using the computer simulations summarized in this work, the observations that emerge suggest a number of general principles that may be valid for folding-unfolding transitions in general:

1. Folding-unfolding transitions occur within a manifold of paths, exhibiting substantial fluctuations in accessible configurations and molecular shape features.
2. The unfolding of neutral lysozyme appears as a concerted process, characterized by a simultaneous change in asphericity and chain entanglement. This result suggests a coupling between the loss of compactness and the undoing of local folding features (e.g., secondary structural elements). Qualitatively, the unfolding transition is not changed by the strength of the coupling to a thermal bath. The latter affects only the persistence and relative population of unfolding intermediates.
3. The unfolding of lysozyme ions with $q > q^* + 1$ appears as a two-step (nonconcerted) process, characterized by a fast initial loss of chain entanglement, with little effect on the global anisometry. Despite the differences in the actual paths, the final unfolded conformers for lysozyme ions resemble those of neutral lysozyme in point 2 above.
4. In absence of unfolding conditions, *in vacuo* lysozyme remains in a slightly compactified (metastable) form of crystal native state (the so-called "*in vacuo* native structure," *IVNS*). When unfolding conditions are present, the transition does not take place directly from this compactified structure. Rather, the protein swells back toward the characteristic shape of the native structure in the crystal, and only then unfolding elicits. In other words, the crystal structure acts as an attractor for relaxations of the compact *IVNS.*
5. Relaxation from the unfolded state of neutral lysozyme occurs within a manifold of accessible paths, leading both to quasi-native and nonnative persistent conformers. Quasi-native refolding takes place when relaxing from partly unfolded conformers; in this case, the path retraces reversibly the unfolding trajectory. When relaxing from fully unfolded conformers, the paths deviate markedly from the unfolding trajectory. In this latter case, refolding is initiated by partial polymer collapse, leading to flexible structures that are not maximally compact. This process produces usually long-lived (denatured) compact conformers, but it can also lead to configurational trapping at noncompact structures.
6. Refolding is sequence dependent. The threading of a polypeptide sequence over an unfolded lysozyme conformer produces a radically different relaxation behavior.
7. The general pattern of the unfolding and refolding transitions is not affected by mild changes in the attractive interaction between amino acid residues. For instance, a lysozyme ion with subcritical charge and decreased attraction would unfold following the same pattern of a high-charge ion with "standard" attraction. These observations indicates

that capturing the essential characteristics of the folding transition do not require a very precise parametrization of the model force field.

Some of these findings are likely associated with particular features of disulfide-intact lysozyme, e.g., the presence of unbroken disulfide bridges and α/β domains. However, some of the above observations are in agreement with some general principles believed to govern the folding of all globular proteins. For instance, the current consensus is that protein-folding kinetics reflects the existence of multiple paths over a rugged free-energy landscape, biased toward the native fold.[24–34,107–109] Over this surface, some paths can produce native-like conformers, whereas others can be "derailed" (or "topologically frustrated") toward folded denatured structures.[26–32,112] An early approach to protein folding conjectured that full polymer collapse was a necessary step required to reduce the configurational space accessible to an unfolded protein.[23,38] Our results, however, are more consistent with an alternative view, whereby convergence toward the native structure can also follow paths involving only partial compactization.[34,37,61,90] This behavior fits also with results showing that the balance between the formation of a tertiary fold and the degree of compactness is affected by composition and the strength of the monomer interaction.[40,41] In conclusion, the folding of most globular proteins is believed to be determined by molecular forces, depending only on the primary sequence and the protein's environment. Modeling this phenomenon within the framework of *ab initio* methods[119–121] requires a representation of the protein geometry, a force field, and a configurational searching technique. This study contributes to the field by introducing a reliable protocol to expand the study of structural transitions to include proteins in-flight in a vacuum. In this context, our approach enables further exploration of the role of molecular forces in these transitions[122] by probing the effects of variable charges and monomer–monomer interactions. In terms of geometrical and topological shape descriptors, this study provides a framework to interpret the results of MD simulations in light of basic concepts such as the degree of polymer compactness and the organization of a tertiary fold.

ACKNOWLEDGMENTS

We thank Prof. Curt Reimann (Lund) for many enjoyable and valuable discussions, Jason Sutherland and Denise Caughill (Sudbury) for their collaboration with part of the research, and Naomi Grant (Sudbury) for her comments on the manuscript. G.A.A. acknowledges financial support from the Natural Sciences and Engineering Research Council (NSERC) of Canada, and O.T. from the Swedish Natural Sciences Council (NFR).

REFERENCES

1 McCammon, J.A., and Harvey, S.C. *Dynamics of Proteins and Nucleic Acids* Cambridge University Press, Cambridge, 1987
2 Lazaridis, T., and Karplus, M. *Science* 1997, **278**, 1928
3 Daura, X., Jaun, B., Seebach, D., van Gunsteren, W.F., and Mark, A.E. *J. Mol. Biol.* 1998, **280**, 925
4 Duan, Y., and Kollman, P.A. *Science* **282**, 1998, 740
5 Doniach, S., and Eastman, P. *Curr. Op. Struct. Biol.* 1999, **9**, 157
6 Daura, X., van Gunsteren, W.F., and Mark, A.E. *Proteins* 1999, **34**, 269
7 Elber, R., Meller, J., and Olender, R. *J. Phys. Chem. B* 1999, **103**, 899
8 Shelimov, K., and Jarrold, M.F. *J. Am. Chem. Soc.* 1996, **118**, 10313
9 Gross, D.S., Schnier, P.D., Rodríguez-Cruz, S.E., Fagerquist, C.K., and Williams, E.K. *Proc. Natl. Acad. Sci. U.S.A.* 1996, **93**, 3143
10 Shelimov, K., and Jarrold, M.F. *J. Am. Chem. Soc.* 1997, **119**, 2987
11 Shelimov, K.B., Clemmer, D.E., Hudgins, R.R., and Jarrold, M.F. *J. Am. Chem. Soc.* 1997, **119**, 2240
12 Valentine, J.S., Anderson, J.G., Ellington, A.D., and Clemmer, D.E. *J. Phys. Chem. B.* 1997, **101**, 3891
13 McLafferty, F.W., Guan, Z., Haupts, U., Wood, T.D., and Kelleher, N.L. *J. Am. Chem. Soc. 1998*, **120**, 4732
14 Englander, S.W. *Annu. Rev. Biophys. Biomol. Struct.* 2000, **29**, 213
15 Konermann, L. *Sci. Prog.* 1998, **81**, 123
16 Kelleher, N.L., Lin, H.Y., Valaskovic, G.A., Aaserud, D.J., Fridriksson, E.K., and McLafferty, F.W. *J. Am. Chem. Soc.* 1999, **121**,806
17 Jarrold, M.F. *Acc. Chem. Res.* 1999, **32**, 360
18 Hoaglund-Hyzer, C.S., Counterman, A.E., and Clemmer, D.E. *Chem. Rev.* 1999, **99**, 3027
19 Sullivan, P.A., Reimann, C.T., Axelsson, J., Altmann, S., Quist, A.P., and Sundqvist, B.U.R. *J. Am. Soc. Mass. Spectrom.* 1996, **7**, 329
20 Reimann, C.T., Sullivan, P.A., Axelsson, J., Quist, A.P., Altman, S., Roepstorff, P., Velázquez, I., and Tapia, O. *J. Am. Chem. Soc.* 1998, **120**, 7608
21 Wolynes, P.G. *Proc. Natl. Acad. Sci.U.S.A.* 1995, **92**, 2426
22 Lifshits, I.M., Grosberg, A.Yu., and Khokhlov, A.R. *Rev. Mod. Phys.* 1978, **50**, 683
23 Chan, H.S., and Dill, K.A. *Annu. Rev. Biophys. Biophys. Chem.* 1991, **20**, 447
24 Gutin, A.M., Abkevich, V.I., and Shakhnovich, E.I. *Biochemistry* 1995, **34**, 3066
25 Du, R., Pande, V.S., Grosberg, A.Yu, Tanaka, T., and Shakhnovich, E.I. *J. Chem. Phys.* 1998, **108**, 334
26 Dobson, C.M., and Karplus, M. *Curr. Op. Struct. Biol.* 1999, **9**, 92
27 Garel, T., and Orland, H. *Europhys. Lett.* 1988, **6**, 307
28 Sali, A., Shakhnovich, E., and Karplus, M. *J. Mol. Biol.* 1994, **235**, 1614
29 Yee, D.P., Chan, H.S., Havel, T.F., and Dill, K.A. *J. Mol. Biol.* 1994, **241**, 557
30 Guo, Z., and Thirumalai, D. *Folding Design* 1997, **2**, 377
31 Dill, K.A., and Chan, H.S. *Nat. Struct. Biol.* 1997, **4**, 10
32 Socci, N.D., and Onuchic, J.N., and Wolynes, P.G. *Proteins* 1998, **32**, 136
33 Plaxco, K.W., and Baker, D.B. *Proc. Natl. Acad. Sci. U.S.A.* 1998, **95**, 13591
34 Thirumalai, D., and Klimov, D.K. *Curr. Op. Struct. Biol.* 1999, **9**, 197
35 Daura, X., Mark, X., and van Gunsteren, W.F. *Comp. Phys. Commun.* 1999, **123**, 97
36 Arteca, G.A. Molecular shape descriptors. In: K. Lip-

kowitz and D.B. Boyd (eds), *Reviews in Computational Chemistry*, vol. 9 (VCH, New York, 1996)
37 Alm, E., and Baker, D. *Curr. Op. Struct. Biol.* 1999, **9**, 189
38 Chan, H.S., and Dill, K.A. *Macromolecules* 1989, **22**, 4559
39 Lattman, E.E., Fiebig, K.M., and Dill, K.A. *Biochemistry* 1994, **33**, 6158
40 Abkevich, V.I., Gutin, A.M., and Shakhnovich, E.I. *J. Mol. Biol.* 1995, **252**, 460
41 He, S., and Scheraga, H.A. *J. Chem. Phys.* 1998, **108**, 287
42 Rudnick, J., and Gaspari, G. *J. Phys. A* 1986, **19**, L 191
43 Rudnick, J., and Gaspari, G. *Science* 1987, **237**, 384
44 Diehl, H.W., and Eisenriegler, E. *J. Phys. A* 1989, **22**, L 87
45 Baumgärtner, A. *J. Chem. Phys.* 1993, **98**, 7496
46 Arteca, G.A., and Mezey, P.G. *Biopolymers* 1992, **32**, 1609
47 Janse van Rensburg, E.J., Sumners, D.W., Wasserman, E., and Whittington, S.G. *J. Phys. A* 1992, **25**, 6557
48 Arteca, G.A. *Biopolymers* 1993, **33**, 1829
49 Arteca, G.A. *Phys. Rev. E* 1994, **49**, 2417
50 Arteca, G.A. *Phys.Rev. E* 1995, **51**, 2600
51 Arteca, G.A. *Macromolecules* 1996, **29**, 7594
52 Orlandini, E., Tesi, M.C., Whittington, S.G., Sumners, D.W., and Janse van Rensburg, E.J. *J. Phys. A* 1994, **27**, L333
53 Gutin, A., Sali, A., Abkevich, V., Karplus, M., and Shakhnovich, E.I. *J. Chem. Phys.* 1998, **108**, 6466
54 Baldwin, R.L., and Rose, G.D. *Trends Biochem. Sci.* 1999, **24**, 26
55 Baldwin, R.L., and Rose, G.D. *Trends Biochem. Sci.* 1999, **24**, 77
56 Bai, Y. *Proc. Natl. Acad. Sci. U.S.A.* 1999, **96**, 477
57 Arteca, G.A., Velázquez, I., Reimann, C.T., and Tapia, O. *Phys. Rev. E* 1999, **59**, 5981
58 Arteca, G.A., Velázquez, I., Reimann, C.T., and Tapia, O. *J. Chem. Phys.* 1999, **111**, 4774
59 Kholodenko, A.L., and Vilgis, T.A. *Phys. Rep.* 1998, **298**, 251
60 Stasiak, A., Katritch, V., Bednar, J., Michoud, D., and Dubochet, J. *Nature* 1996, **384**, 122
61 Katritch, V., Bednar, J., Michoud, D., Scharein, R.G., Dubochet, J., and Stasiak, A. *Nature* 1996, **384**, 142
62 Vologodskii, A.V., Crisona, N.J., Laurie, B., Pieranski, P., Katritch, V., Dubochet, J., and Stasiak, A. *J. Mol. Biol.* 1998, **278**, 1
63 Laurie, B., Katritch, V., Sogo, J., Koller, T., Dubochet, J., and Stasiak, A. *Biophys. J.* 1998, **74**, 2815
64 Cantarella, J., Kusher, R.B., and Sullivan, J.M. *Nature* 1998, **392**, 237
65 Buck, G. *Nature* 1998, **392**, 238
66 Rabin, Y., Grosberg, A.Yu., and Tanaka, T. *Europhys. Lett.* 1995, **32,** 505
67 Arteca, G.A., Nilsson, O., and Tapia, O. *J. Mol. Graph.* 1993, **11**, 193
68 Arteca, G.A., and Tapia, O. *J. Chem. Inf. Comp. Sci.* 1999, **39**, 642
69 Arteca, G.A., and Tapia, O. *Int. J. Quantum Chem.* 2000, **80**, 848
70 Arteca, G.A., and Caughill, D. I. *Can. J. Chem.* 1998, **76**, 1402
71 Arteca, G.A. *J. Chem. Inf. Comp. Sci.* 1999, **39**, 550
72 Arteca, G.A. *Biopolymers* 1995, **35**, 393
73 Halperin, A., and Goldbart, P.M. *Phys. Rev. E* 2000, **61**, 565
74 Finkelstein, A.V. *Protein Eng.* 1997, **10**, 843
75 Privalov, P.L. *CRC Crit. Rev. Biochem. Mol. Biol.* 1990, **25**, 181
76 Alonso, D., and Daggett, V. *Protein Sci.* 1998, **7**, 860
77 Kazmirski, S.L., and Daggett, V. *J. Mol. Biol.* 1998, **277**, 487
78 Bruscolini, P., and Casetti, L. *Phys. Rev. E* 2000, **61**, R2208
79 Marchi, M., and Ballone, P. *J. Chem. Phys.* 1999, **110**, 3697
80 Ferrara, P., Apostolakis, J., and Caflisch, A. *J. Phys. Chem. B* 2000, **104**, 4511
81 Williams, M.A., Thornton, J.M., and Goodfellow, J.M. *Protein Eng.* 1997, **10**, 895
82 Reimann, C.T., Velázquez, I., and Tapia, O. *J. Phys. Chem. B* 1998, **102**, 9344
83 Brooks III, C.L. *Curr. Op. Struct. Biol.* 1998, **8**, 222
84 Reimann, C.T., Velázquez, I., and Tapia, O. *J. Phys. Chem. B* 1998, **102**, 2277
85 Miranker, A., Robinson, C.V., Radford, S.E., Aplin, R.T., and Dobson, C.M. *Science* 1993, **262**, 896
86 van Gunsteren, W.F., and Berendsen, H.J.C. *Groningen Molecular Simulation (GROMOS) Library Manual.* Biomos, Groningen, 1987
87 Åqvist, J., van Gunsteren, W.F., Leijonmarck, M., and Tapia, O. *J. Mol. Biol.* 1985, **183**, 461
88 Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak, J.R. *J. Chem. Phys.* 1984, **81**, 3684
89 Koradi, R., Billeter, M., and Wüthrich, K. *J. Mol. Graph.* 1996, **14**, 51
90 Klimov, D.K., and Thirumalai, D. *J. Chem. Phys.* 1998, **109**, 4119
91 Kantor, Y., and Kardar, M. *Europhys. Lett.* 1994, **27**, 643
92 Kantor, Y., and Kardar, M. *Europhys. Lett.* 1994, **28**, 169
93 Kantor, Y., Kardar, M., and Ertas, D. *Physica A* 1998, **249**, 301
94 Wolfling, S., and Kantor, Y. *Phys. Rev. E* 1998, **57**, 5719
95 Dobrynin, A.V., Rubinstein, M., and Obukhov, S.P. *Macromolecules* 1996, **29**, 2974
96 Przybylski, M., and Glocker, M.O. *Angew. Chem. Int. Ed. Eng.* 1996, **35**, 806
97 Suizdak, G., Bothner, B., Yeager, M., Brugidou, C., Fauquet, C.M., Hoey, K., and Chang, C.-M. *Chem. Biol.* 1996, **3**, 45
98 Reimann, C.T., Velázquez, I., Bittner, M., and Tapia, O. *Phys. Rev. E* 1999, **60**, 7277
99 Arteca, G.A., Velázquez, I., Reimann, C.T., and Tapia, O. *Chem. Phys. Lett.* 2000, **327**, 245
100 Miteva, M., Demirev, P.A., and Karshikoff, A.D. *J. Phys. Chem. B* 1997, **101**, 9645
101 Jackson, S.E. *Folding Design* 1998, **3**, R81
102 Hennig, M., Bermel, W., Spencer, A., Dobson, C.M., Smith, L.J., and Schwalbe, H. *J. Mol. Biol. 2* 1999, **88**, 705
103 Arteca, G.A., Reimann, C.T., and Tapia, O. *J. Phys. Chem. B* 2000, **104**, 11360
104 Baker, D. *Nature* 2000, **405**, 39

105 Levinthal, C.J. In: *Mössbauer Spectroscopy of Biological Systems*, Debrunner, P., Tsibris, J.C.M., Munck, E., Eds. (University of Illinois Press, Urbana, 1968)
106 Levinthal, C.J. *J. Chim. Phys.* 1968, **65**, 44
107 Karplus, M., Shakhnovich, E.I. *Protein Folding*, T. Creighton, ed. (Freeman, New York, 1992,
108 Baldwin, R. *Nature* 1994, **369**, 183
109 Bryngelson, J., Onuchic, J.N., Socci, N.D., and Wolynes, P.G. *Proteins Struct. Funct. Gen.* 1995, **21**, 167
110 Kiefhaber, T. *Proc. Natl. Acad. Sci. U.S.A.* 1995, **92**, 9029
111 Mantagne, A., Radford, S.E., and Dobson, C.M. *J. Mol. Biol.* 1997, **267**, 1068
112 Shea, J.-E., Onuchic, J.N., and Brooks III, C.L. *Proc. Natl. Acad. Sci. U.S.A.* 1999, **96**, 12512
113 Li, A.J., and Daggett, V. *J. Mol. Biol.* 1998, **275**, 677
114 Velázquez, I., Reimann, C.T., and Tapia, O. *J. Am. Chem. Soc.* 1999, **121**, 11468
115 Arteca, G.A., Paulino, M., Reimann, C.T., and Tapia, O. *Phys. Chem. Chem. Phys.* 2000, **2**, 5259
116 Nilsson, O. *J. Mol. Graph.* 1990, **8**, 192
117 Lio, R., Walter, A.B., and Mathews, B.W.*J. Mol. Biol.* 2000, **295**, 129
118 O. Nilsson, O., and Tapia, O. *J. Mol. Struct. (Theochem)* 1992, **256**, 295
119 Goodfellow, J.M., and Levy, R.M. *Curr. Op. Struct. Biol.* 2000, **10**, 137
120 Lazaridis, T., and Karplus, M. *Curr. Op. Struct. Biol.* 2000,**10**, 139
121 Osguthorpe, D.J. *Curr. Op. Struct. Biol.* 2000, **10**, 146
122 Leckband, D. *Annu. Rev. Biophys. Biomol. Struct.* 2000, **29**, 1