Proteins in Vacuo: Relaxation of Unfolded Lysozyme Leads to Folding into Native and Non-Native Structures. A Molecular Dynamics Study

I. Velázquez,[†] C. T. Reimann,[‡,§] and O. Tapia*[,†]

Contribution from the Department of Physical Chemistry, Uppsala University, Box 532, S-751 21 Uppsala, Sweden, and Division of Ion Physics, Department of Materials Science, Uppsala University, Box 534, S-751 21 Uppsala, Sweden

Received March 22, 1999

Abstract: Unfolding and refolding processes for proteins in vacuo and gas phase are becoming the subject of experimental and theoretical attention. Recently, a molecular dynamics study of unfolding of disulfide-bond-intact lysozyme (DI-LYZ) in vacuo showed large-scale conformational changes (Reimann et al., *J. Phys. Chem. B* **1998**, *102*, 2277), thereby providing a configurational space for a denatured state. Here we study the opposite process: from the multimodal unfolding trajectory, a variety of conformations were selected for relaxation studies aimed at computationally mimicking "renaturing" conditions. For DI-LYZ, the relaxations gathered in two distinct classes of conformers as measured by their root-mean-square deviations (RMSD) from the X-ray structure. Structures originating from above the approximate midpoint of the main unfolding transition, with initial RMSD ranging from 8 to 17 Å, relaxed toward persistent compact structures having RMSD $\approx 7.5 \pm 0.5$ Å (Class I). They represent compact denatured albeit folded structures. Structures originating from conformers having initial RMSD < 8 Å yielded two subclasses of structures on relaxation: near-native (RMSD ≈ 3 Å) and "nativelike" (RMSD < 2 Å) (together comprising Class II). Both compact and elongated lysozyme species, reported in this work, are consistent with experimentally observed lysozyme conformers in vacuo. The relaxations under renaturation conditions do not elicit a random search of the conformational space. Rather, compact conformers with persistent tertiary and rich secondary structures rapidly form. Therefore, results of the work reported here and in a companion paper (Arteca et al., *Phys. Rev. E* **1999**, *59*, 5981) suggest that lysozyme undergoes a folding process in vacuo.

Introduction

There has been enormous progress, both experimental and theoretical,[1−4] in fields related to the protein folding problem. Understanding the molecular mechanisms involved requires inputs from research fields ranging from chemistry, physical chemistry, and biochemistry and most recently from ion physics. On the theoretical side, the mechanism that brings an amino acid sequence into a folded state still remains as an unsolved fundamental riddle.[4] In practice, molecular dynamics (MD) simulations of peptides in water[5,6] have shown that, for sufficiently long trajectories, the systems have a time evolution on the configurational space that populates folded as well as unfolded states. While long simulations of full proteins in water may eventually be feasible, there will remain the problem of disentangling solvent effects from the intrinsic properties of the biomolecule in the context of protein folding. Accessing such intrinsic properties implies knowledge of the protein in vacuo. When the protein folding problem started,[7] the idea of proteins in vacuo was not backed by any experimental experience. Today, a number of mass spectrometric based techniques are opening the field to study such objects.[8−10]

Experimentally, folding/unfolding/refolding processes in vitro are related to drastic changes in the medium surrounding the protein. Unfolding a system usually requires perturbations such as a high concentration of guanidium chloride or a shift in pH or temperature. To refold what has been denatured requires new conditions in which the aggressive medium is attenuated and changed into mild folding-promoting surroundings. However, even if these changes are carefully administered, the kinetics of unfolding/refolding often make it difficult to trap and study transient intermediates. In this situation, computer-assisted simulations are useful to model perturbations that drive unfolding in a similar fashion.[11,12] Protein unfolding intermediates are generated during an unfolding trajectory and herein resides the advantage of this approach: a variety of unfolded structures

* Corresponding author. Telephone: +46 18-471-36-59. FAX: +46 18-18-85-42. E-Mail: Orlando.Tapia@fki.uu.se

[†] Department of Physical Chemistry, Uppsala University.

[‡] Division of Ion Physics, Department of Materials Science, Uppsala University.

[§] Current Address: Department of Analytical Chemistry, Chemical Center, Lund University, Box 124, S-221 00 Lund, Sweden.

(1) Dobson, C. M.; Ptitsyn, O. B. *Curr. Opin. Struct. Biol.* **1997**, *7*, 1−2.
(2) Dill, K. A.; Chan, H. S. *Nature Struct. Biol.* **1997**, *4*, 10−19.
(3) Fersht, A. R. *Curr. Opin. Struct. Biol.* **1997**, *7*, 3−9.
(4) Shakhnovich, E.; Fersht, A. R. *Curr. Opin. Struct. Biol.* **1998**, *8*, 65−67.
(5) Daura, X.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. *J. Mol. Biol.* **1998**, *280*, 925−932.
(6) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740−744.
(7) Anfinsen, C. B. *Science* **1973**, *181*, 223−230.
(8) Shelimov, K. B.; Jarrold, M. F. *J. Am. Chem. Soc.* **1996**, *118*, 10313−10314.
(9) Valentine, S. J.; Anderson, J. G.; Ellington, A. D.; Clemmer, D. E. *J. Phys. Chem.* **1997**, *101*, 3891−3900.
(10) McLafferty, F. W.; Guan, Z. Q.; Haupts, U.; Wood, T. D.; Kelleher, N. L. *J. Am. Chem. Soc.* **1998**, *120*, 4732−4740.
(11) Alonso, D. O. V.; Daggett, V. *J. Mol. Biol.* **1995**, *247*, 501−520.
(12) Lazaridis, T.; Karplus, M. *Science* **1997**, *278*, 1928−1931.

10.1021/ja990911c CCC: $18.00
Published on Web 11/30/1999

become available for further study. To date, precise MD calculations have simulated unfolding of proteins in solution.[3] However, for proteins in vacuo, denaturing/renaturing conditions remain to be defined in current simulation schemes (see below).

A host of experimental techniques are producing considerable evidence that folding/unfolding processes occur in vacuo, i.e., without the involvement of solvent.[8−10] Unfortunately, these techniques do not provide atomic-resolved information concerning the accessible conformational space. MD simulations have been used in our group to supply such information.[13] An unfolding process was set up by using mild perturbing conditions: the protein, in absence of water molecules, was allowed to rotate, producing a kinematic centrifugal force.[13] This type of movement is what characterizes the absence of viscous forces that would otherwise appear in condensed phases. The use of a neutral protein model[14] with boundary conditions allowing for global rotation defines then an in vacuo denaturing milieu. The unfolding path found for disulfide-intact lysozyme (DI-LYZ) showed ordered sequences of subdomain unfolding that were compatible with experimental information published by Miranker et al.[15] The root-mean-square deviation (RMSD) profile displayed a set of almost native conformers as well as two upper plateaus (RMSD $\approx$ 10 and 15 Å, respectively) corresponding to largely unfolded structures. The sharp transition to the first plateau was found to be rich in mildly unfolded structures. To carry out MD simulations of protein refolding, as noticed by van Gunsteren et al., one of the problems resides in the construction of a configurational space representing the denatured state.[16] In this work, we use the configurations found under "denaturing" simulation conditions as a representative of such a state and subject them to renaturing simulations in vacuo.

"Renaturing" conditions are operationally defined via: (i) absence of explicit water (solvent) molecules; (ii) absence of frictive forces so that the system may rotate without restrictions; (iii) using a Maxwell−Boltzmann velocity distribution to initiate the MD run; and (iv) employing a weak coupling to a Berendsen's thermal bath.[17] The initial configurations derived from the "denaturing" run are non-equilibrium conformers with respect to the "renaturing" setup, implying that the systems will somehow relax. The crux of the problem now is to establish the nature of the relaxation processes: whether there is a simple random collapse or whether particular paths occur that may elicit a folding process.[3,18,19] Here we present results showing that disulfide-intact lysozyme relaxes, as any polymer would do under such circumstances, but it does so in such a way that its transit through conformational space can be seen as a protein folding process.

Methods

Control MD Run. Atomic coordinates for DI-LYZ were taken from the crystal structure of wild-type hen egg white lysozyme at 1.7 Å resolution,[20] obtained from the Brookhaven protein data bank (E.C. 3.2.1.17; PDB #1hel). The coordinates of the polar hydrogens were generated from standard geometries. This structure was subjected to cycles of energy minimization.

The DI-LYZ structure was submitted to a control MD run of 1 ns, carried out in a standard fashion using the GROMOS-87 37D4 force field,[14,21] i.e., the NP framework,[13,14] under conditions of constant volume and temperature. The initial velocities were assigned from a Maxwellian distribution characteristic of a temperature of 293 K. Finally, a weak coupling ($\tau = 0.1$ ps) to the Berendsen thermal bath[17] was utilized. DI-LYZ presented a flat RMSD profile representing a small deviation from the crystal structure (RMSD < 2 Å), corresponding to a slight compaction of the conformers in the early stages of the trajectory. This results show the ability of GROMOS 37D4 force field to represent neutral proteins in vacuo in the neighborhood of the crystal structure.

Unfolding Run. In the unfolding run the protein model was strongly coupled ($\tau = 0.01$ ps) to the Berendsen thermal bath, inducing an exchange of kinetic energy between the internal and the external degrees of freedom.[22,23] The system was allowed to rotate. The model protein, submitted to these conditions, undergoes multiple unfolding transitions via a coupling-induced centrifugal force.[13] For the DI-LYZ trajectory, the first transition occurs at roughly 550−600 ps (Figure 1).[13]

Relaxation Runs. The starting conformations for weak-coupling relaxation MD runs were selected structures, taken to coincide with significant conformational changes occurring during the first unfolding transition and totally extended conformers from the DI-LYZ strong-coupled MD trajectory (Figure 1). These are summarized as follows with identifying symbols, where the subscript refers to the time at which each conformation occurred in the strong coupling denaturation run:[13] DI-LYZ$_{560}$, DI-LYZ$_{570}$, DI-LYZ$_{575}$, DI-LYZ$_{580}$, DI-LYZ$_{900}$, and DI-LYZ$_{990}$.

For each selected conformer, rotation and translation about the center of mass were halted, and the atomic coordinates were then used as the starting coordinates for the relaxation MD runs using a weak coupling to the Berendsen thermal bath. Initial atom velocities were randomly selected from a Maxwellian distribution characteristic of a temperature of 293 K. The global system was otherwise left free. The trajectories were integrated for a time period of 1.0 ns.

Analyses. The trajectories were analyzed for RMSD from the native structure, and radii of gyration (R_{gyr}). All comparisons are made with respect to the X-ray structure, unless the contrary is explicitely stated. DI-LYZ possesses two domains (α and β, ref 20). However, from our previous unfolding study, we find it convenient to think in terms of *two* α subdomains,[13] α_1 and α_2, and a β domain. α_1 comprises residues 1−39; α_2 88−129; and β 40−87. α_1 and α_2 have been separately synthesized and studied experimentally.[24] The theoretical analyses are accordingly resolved at the (sub)domain level.

The program MOLMOL[25] was used to monitor the changes in secondary structure during the relaxation runs, at time intervals of 10 ps. The parameters used to trace the existence of an hydrogen bond were a maximum hydrogen-donor−acceptor angle of 35° and a maximum hydrogen-acceptor distance of 2.4 Å.[26]

Relaxing Unfolded Lysozyme with MD Simulations

Selection of Unfolded Coordinate Sets. From the unfolding profile,[13] which shows sharp transitions as well as two plateaus

(13) Reimann, C. T.; Velázquez, I.; Tapia, O. *J. Phys. Chem. B* **1998**, *102*, 2277−2283.
(14) Åqvist, J.; van Gunsteren, W. F.; Leijonmarck, M.; Tapia, O. *J. Mol. Biol.* **1985**, *183*, 461−477.
(15) Miranker, A.; Robinson, C. V.; Radford, S. E.; Aplin, R. T.; Dobson, C. M. *Science* **1993**, *262*, 896−900.
(16) van Gunsteren, W. F.; Hünenberger, P. H.; Kovacs, H.; Mark, A. E.; Schiffer, C. A. *Philos. Trans. R. Soc. London, Ser. B* **1995**, *348*, 49−59.
(17) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(18) Levinthal, C. J. In *Mössbauer Spectroscopy in Biological Systems*; Debrunner, P., Tsibris, J. C. M., Eds.; University of Illinois Press: Urbana, IL, 1968; p 22.
(19) Levinthal, C. J. *J. Chim. Phys.* **1968**, *65*, 44−45.
(20) Wilson, K. P.; Malcolm, B. A.; Matthews, B. W. *J. Biol. Chem.* **1992**, *267*, 10842.
(21) van Gunsteren, W. F.; Berendsen, H. J. C. *Groningen Molecular Simulation (GROMOS) Library Manual*; Biomos: Nijenborgh 16, 9747 AG Groningen, The Netherlands, 1987.
(22) Tapia, O.; Nilsson, O. In *Molecular Aspects of Biotechnology: Computational Models and Theory;* Bertrán, J., Ed.; Kluwer: The Netherlands, 1992.
(23) Lemak, A. S.; Balabaev, N. K. *Mol. Simul.* **1994**, *13*, 177−187.
(24) Yang, J. J.; van den Berg, B.; Pitkeathly, M.; Smith, L. J.; Bolin, K. A.; Keiderling, T. A.; Redfield, C.; Dobson, C. M.; Radford, S. E. *Folding Des.* **1996**, *1*, 473−484.
(25) Koradi, R.; Billeter, M.; Wüthrich, K. *J. Mol. Graphics* **1996**, *14*, 51−55.
(26) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577−2637.

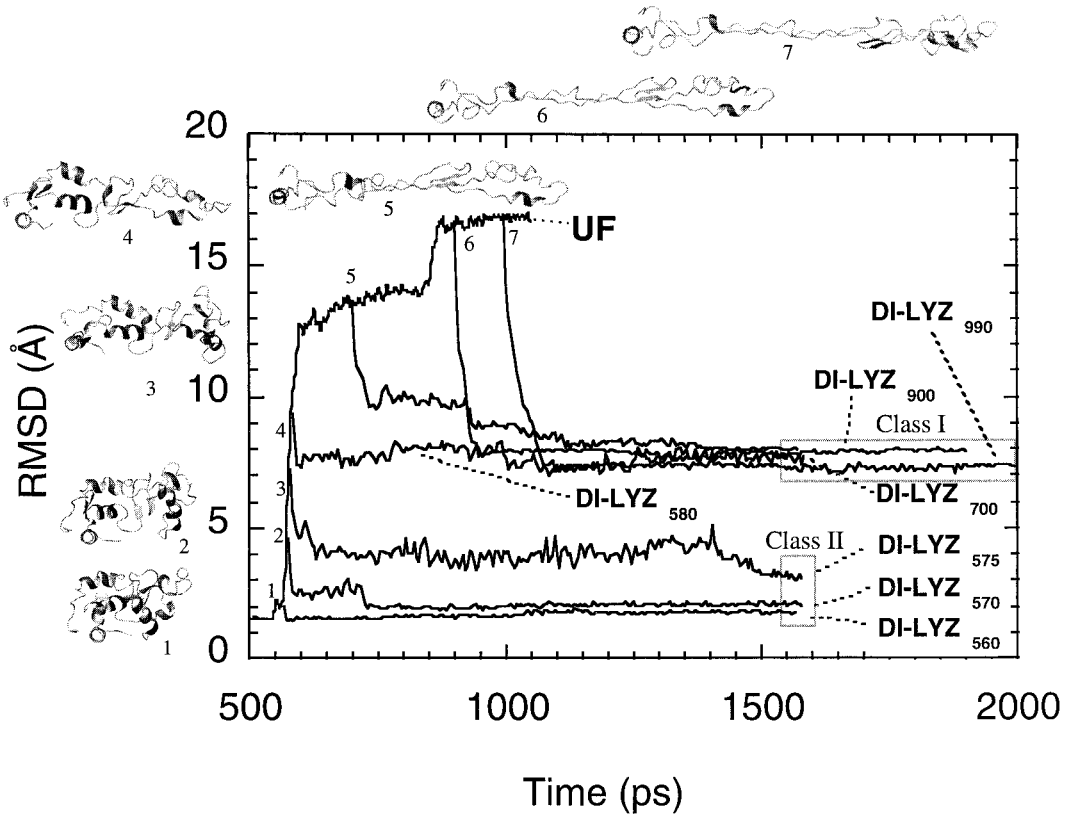

Figure 1. Root-mean-square deviations (RMSD) of the C_α atoms from the native structure over time. Unfolded seeding conformers for relaxation molecular dynamics (MD) were selected from an unfolding MD trajectory (UF, ref 13) and are shown; numbers (1−7) indicate from where along UF they were taken. Time profiles for the seven relaxation MD runs are shown, labeled DI-LYZ$_{560}$, DI-LYZ$_{570}$, DI-LYZ$_{575}$, DI-LYZ$_{700}$, DI-LYZ$_{580}$, DI-LYZ$_{900}$, and DI-LYZ$_{990}$, where the subscript indicates the time during UF where the seeding conformation occurred. As described in the text, relaxation toward two RMSD regions can be appreciated.

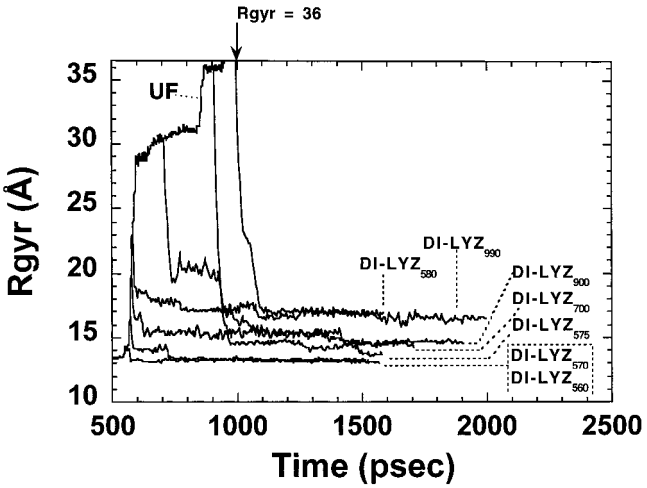

Figure 2. Radii of gyration (R_{gyr}) over time for DI-LYZ$_{560}$, DI-LYZ$_{570}$, DI-LYZ$_{575}$, DI-LYZ$_{580}$, DI-LYZ$_{700}$, DI-LYZ$_{900}$,and DI-LYZ$_{990}$. Note the sharp transition presented by the most extended conformers toward more compact structures.

in the RMSD and R_{gyr} time profiles (Figures 1, 2), a set of conformers were selected for relaxation studies. The structures, shown in Figure 1, characterize the main unfolding transition (560−595 ps) as well as the two upper plateaus. The cross-RMSD values (Table 1) show that successively chosen unfolded structures differ from each other by at least ∼2 Å, except for DI-LYZ$_{900}$ and DI-LYZ$_{990}$, which are very similar to each other.

Global Structure. We first consider starting conformations characterized by RMSD > 8 Å. These four conformations (labeled 4−7 in Figure 1) are considerably elongated. DI-LYZ$_{580}$, DI-LYZ$_{700}$, DI-LYZ$_{900}$, and DI-LYZ$_{990}$ relaxed rapidly to persistent (possibly metastable) states with RMSD values between 7 and 8 Å (7.5 ± 0.5 Å; Figure 1). The final structures reached in these trajectories were not identical. For example, they display significantly different R_{gyr} values (cf. Figure 2), showing different degrees of compaction. Also, the cross-RMSD values show that they are as different from each other as they are from the native structure (Table 1). This would mean that these unfolded structures migrate toward a sort of attractor defining a hyperspherical shell region characterized by RMSD ≈7 Å. We regard these conformations as members of a class of compact non-native structures hereafter referred to as Class I.

For starting conformations characterized by RMSD < 8 Å, the relaxation behavior is different: the three unfolded conformers apparently display a "memory" of their original folded state and easily reach quasi-native or native conformations on relaxation. At the onset of unfolding this may be seen as a natural result, and consequently, the relaxed structures are grouped together in Class II. As an example, DI-LYZ$_{560}$ had an initial denatured conformation presenting an RMSD ≈ 1.9 Å (structure 1 in Figure 1) and R_{gyr} only slightly greater than that of the native form (Figure 2). After relaxation for roughly 20 ps, the agreement with the X-ray structure improved, RMSD ≈1.7 Å, while R_{gyr} decreased to 13 Å (the same value as that for the native conformation), indicating the rapid reoccupation of the crystal native conformational space.

For starting structures characterized by somewhat larger RMSD, the relaxation presents some special characteristics. DI-LYZ$_{570}$ initially has RMSD ≈ 4.6 Å and thus was significantly denatured (structure 2 in Figure 1). The system rapidly relaxed

Table 1. Cross-RMSD Values for the Selected DI-LYZ Unfolded Conformers (light background), Remarking in Bold the RMSD with the Preceding Conformer; and Cross RMSD Values for the Final Conformations after 1 ns of Relaxation (shadowed background), Remarking in Bold RMSD Values that Correspond to Essentially the Same Structure[a]

	native	560	570	575	580	700	900	990	
native		**1.92**	3.41	7.18	8.47	13.53	16.65	16.95	native
560	**1.73**		**2.28**	6.11	7.46	12.72	15.84	16.17	560
570	**1.79**	**1.11**		**4.38**	5.88	11.19	14.38	14.69	570
575	3.04	2.68	2.82		**2.39**	7.65	10.88	11.19	575
580	7.82	7.64	7.93	6.32		**6.20**	9.54	9.88	580
700	8.17	7.64	7.92	7.38	6.76		**3.92**	4.18	700
900	7.86	7.62	7.63	8.18	10.07	7.89		**1.01**	900
990	7.10	6.81	6.80	7.58	9.62	7.70	4.25		990
	native	560	570	575	580	700	900	990	

[a] Units are in Å.

to a more compact structure (RMSD ≈2.5 Å). From 95 to 140 ps after initiation of relaxation, R_{gyr} reflected a more fluctuating pattern, prior to a swift conformational change occurring between 140 and 160 ps. The structure stabilized down to RMSD ≈ 1.8 Å, indicating that the conformer had folded back to a nativelike structure. The RMSD profile indicates that the relaxation is not a simple collapse. After a total of 1.0 ns of relaxation, RMSD and R_{gyr} values continued to reflect a good agreement between the relaxed conformations and the X-ray structure.

The initial DI-LYZ$_{575}$ (RMSD = 7.2 Å) can be considered as a distinctly unfolded structure (structure 3 in Figure 1). This one presented the most variable profile for both RMSD and R_{gyr}, characterized by a sudden initial collapse that took place within the first 50 ps of relaxation, leading to an oscillating plateau on which the system was trapped for about 750 ps. The final change consisted of a noticeable increase in compactness, accompanied by a drop in RMSD toward the X-ray structure. However, the last snapshot structure (RMSD ≈ 3 Å) cannot be considered a nativelike conformation, and we assign it to a near-native subclass included in Class II. The nearness will be more apparent when analyzing the (sub)domain structure evolution (see below).

Domain Structure. There are clear indications that the sets of residues assigned to the native (sub)domains, denatured in the initial structures, can also access more compact and sometimes even nativelike conformations during the relaxation dynamics.

The relaxation profiles in the (sub)domain RMSDs were multi-plateau, and the transitions observed were sigmoidal in nature. The larger changes in the α_2 subdomain and the β domain RMSDs tended to occur in unison (see arrows in Figure 3), even discounting the initial collapse in the first picoseconds of the relaxation runs. By contrast, the α_1 subdomain displayed a lack of synchrony. The α_2 subdomain, the most extensively affected under the assayed denaturation conditions,[13] presented the richest variety of relaxation patterns. The β domain also displayed a palpable variety of relaxation patterns. On the other hand, the α_1 subdomain presented a fairly simple behavior. In general, relaxing (sub)domains reached more compact metastable conformations which, however, were far from the native one.

For the Class I members, the three domains were initially highly unfolded and relaxed as shown in Figure 3. The DI-LYZ$_{900}$ and DI-LYZ$_{990}$ trajectories were special in that the cross-RMSD of the seeding structures was 1.0 Å, indicating close structural equivalence. However, the relaxations took separate paths. This is especially noticed with the α_2 subdomain, for which the RMSD dropped to 3−4 Å for DI-LYZ$_{900}$ while dropping to only 6 Å for DI-LYZ$_{990}$. (RMSDs for α_1 dropped to roughly the same level for these two relaxation trajectories; a corresponding effect was seen for β.)

A remarkable case was presented by the trajectory DI-LYZ$_{700}$. A very large change of the RMSD brought the β domain from a fully unfolded state (6.8 Å) to a totally folded nativelike conformation (Figure 3). The transition took place via two metastable states separated by fast and slow (diffusive-like) transitions; a trace of a change in the RMSD profile of α_2 could be sensed to coincide closely with the sharp second relaxation step of β (see arrows in Figure 3).

Some synchrony was observed in the relaxation of α_2 and β for the DI-LYZ$_{580}$ trajectories (see arrows in Figure 3). This relaxation was also unique in displaying a slow *unfolding* transition immediately after the fast initial collapse (Figure 3).

During relaxations leading to the Class II members, (sub)-domain RMSDs either stayed in the region characteristic of the native conformation, or else returned rapidly to this region. This is thought to be correlated with the relatively minor unfolding of the seeding structures for DI-LYZ$_{560}$ and DI-LYZ$_{570}$ (all (sub)domains), and for DI-LYZ$_{575}$ (α_1 subdomain). In the last trajectory, the α_2 and β (sub)domains exhibited synchronous adjustments (see arrows in Figure 3) in RMSD during relaxation, culminating in a final collapse to near-native or native conformations, respectively. This final collapse was signaled by oscillations/fluctuations which were large compared to those displayed by the other trajectories.

Secondary Structure. The X-ray structure[20] shows that 64 of the amino acids (50%) take part in the formation of the native secondary structure. We examined the persistence of these structures as well as that of alternative secondary structures that appeared during the relaxation runs. The relative persistence of some secondary structural features is used to classify them as: *transient secondary structures* when these last for a few ps, and *metastable secondary structures* when these are present for more than 100 ps.

Amount of Native Secondary Structure Retained by Unfolded Conformers. The analysis of the unfolded conformers selected for seeding MD relaxation runs reveals how much nativelike secondary structure DI-LYZ had retained during the unfolding. It can be seen from Figures 4a and 5 that the remaining secondary structure in the unfolded conformers corresponds mainly to that of the native conformation. The seeding structures for DI-LYZ560, DI-LYZ570, DI-LYZ575, and DI-LYZ990 presented 58, 49, 55, and 20 residues, respectively, involved in native secondary structure (Figure 5). The seeding conformations also displayed a few residues participating in non-native secondary structure formation.

Degree of Folding and Amount of Secondary Structure. The amount of secondary structure for each of the trajectories has been computed versus time to follow up its persistence (Figure 5). The total amount of secondary structure was roughly preserved (Figures 4b and 5). However, slight tendencies to lose or gain some secondary structure could be observed. For example, the DI-LYZ$_{990}$ relaxation trajectory displayed a net

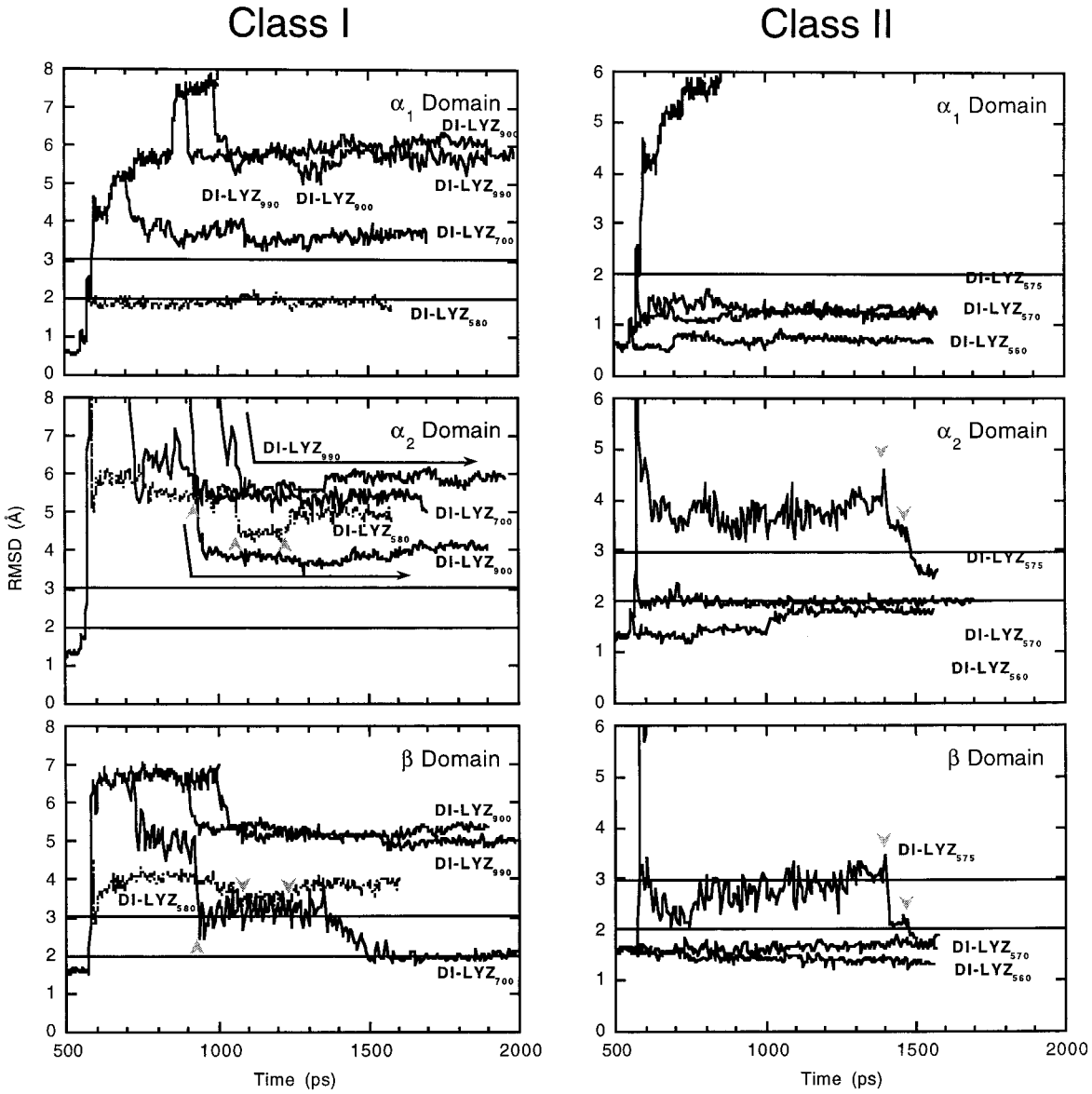

Figure 3. Time profile of the RMSD values resolved by (sub)domains α_1, α_2, and β for DI-LYZ$_{560}$, DI-LYZ$_{570}$, and DI-LYZ$_{575}$ (Class II), and DI-LYZ$_{580}$, DI-LYZ$_{700}$, DI-LYZ$_{900}$, and DI-LYZ$_{990}$ (Class I). The arrows give examples of synchronous fluctuations in (sub)domain-resolved RMSD. Comparison of these curves with those presented in Figures 1 and 2 allows additional correlated and independent changes to be observed.

increase in the amount of secondary structure by about 75% (from 20 to 35 residues), reaching peaks of 140% increase (48 residues). Notice the general pattern of ample fluctuations in the amount of secondary structure that was found to be greater than the overall loss or gain during relaxation (Figures 4,5). In general, the trajectories that relaxed to Class I structures presented an increased β-strand content and a reduced amount of helical features compared to the native structure; an exception was DI-LYZ$_{580}$, which displayed very little β-strand content.

Plots of the number of amino acids forming secondary structural features versus RMSD are presented in Figure 4. Considering the seeding conformers, which are evidently created under strongly nonequilibrium conditions, it can be noticed that more unfolded conformers present a somewhat lower amount of secondary structure compared to conformers closer to the native structure. However, *after* relaxation, such a trend was no longer seen. For example, for each trajectory, relaxation toward more nativelike conformations did not correlate with a net gain in secondary structure. Moreover, after the fast collapse stage, clusters of conformers characterized by a very narrow range of RMSD displayed secondary structure contents fluctuating by ±20%, a large value. Finally, when comparing conformational clusters arising out of relaxation of different seeding structures, similar ranges of amount of secondary structure were experienced even though RMSD values varied from <2 to >6 Å. The perspective provided by these relaxation runs is that a ≈35% secondary structure content is compatible with both nativelike and non-native structures.

Nativelike Secondary Structure. The results of a secondary structure analysis for the different relaxation trajectories displayed patterns in common with that of the native structure (Figure 5). Residues 1−4 seldom participated in the formation of secondary structure, and residues 16, 23, 37, 47−49, 55−56, 67−68, 87, 103, 119, and 124−129 (at the carboxy terminal end) never participated in the formation of secondary structure. The region from residues 42−59 contains the three antiparallel β-strands characteristic of native DI-LYZ;[20] in the relaxation runs, these displayed a fluctuating character, but no alternate secondary structures appeared. The region covered from residue 63−79 did not present much secondary structure, while residues

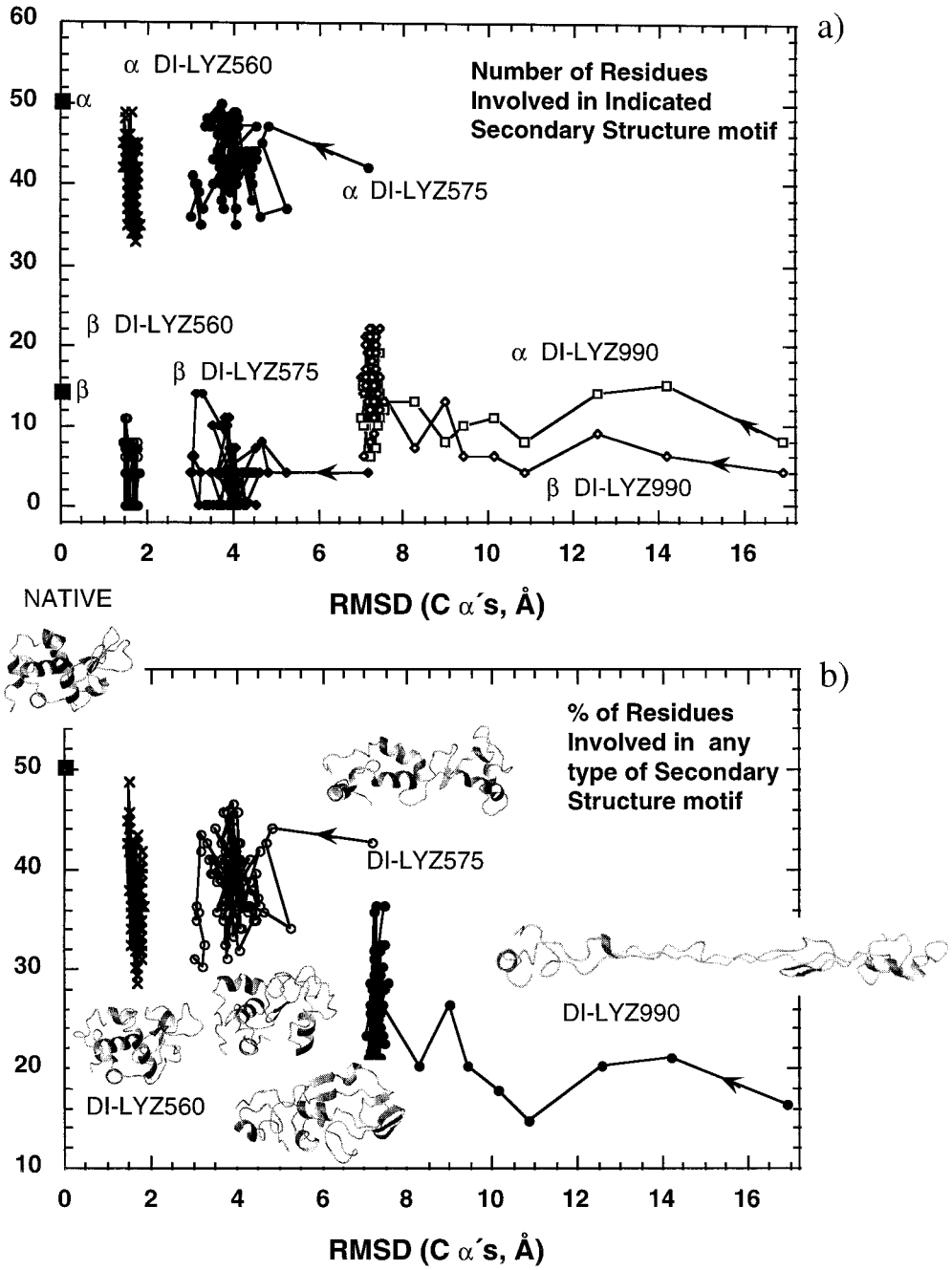

Figure 4. Top: number of residues involved in the formation of α-helix or β-strand versus RMSD over time for DI-LYZ$_{560}$, DI-LYZ$_{575}$, and DI-LYZ$_{990}$. Arrows indicate the time direction. Bottom: percentage of residues involved in helical or β-strand motifs versus RMSD over time for DI-LYZ$_{560}$, DI-LYZ$_{575}$, and DI-LYZ$_{990}$. A vertical clustering of points indicates a bad correlation between RMSD and the amount of secondary structure.

80−85 were always helical, differing from the native conformation in being more α rather than 3_{10} helix. From residues 87 through 100 a highly preserved piece of the long helix 4 was found, although it displayed a slight tendency to become 3_{10} helix in its first segment, residues 87−91, under conditions of severe unfolding for DI-LYZ$_{990}$.

Metastable Non-Native Secondary Structures. In Figure 5, several examples of the formation of non-native secondary structure lasting for times >100 ps can be seen. For example, metastable 3_{10} helix often forms between residues 17−20, between α-helices H1 and H2 (defined in ref 20). This is seen in *all* trajectories except DI-LYZ$_{575}$, in which a fleeting β-sheet is observed (residues 18−24). Immediately at the C-terminus side of the β sheet, helical motifs often form (residues 59−63), except for DI-LYZ$_{900}$ and DI-LYZ$_{990}$. For trajectories initiated from highly elongated conformers, DI-LYZ$_{700}$, DI-LYZ$_{900}$, and DI-LYZ$_{990}$, formation of pervasive non-native β sheets was seen in lieu of H2 (residues 31−35;[20] DI-LYZ$_{900}$, and DI-LYZ$_{990}$) and in the region of H5−H6 (residues 100−114;[20] DI-LYZ$_{700}$, DI-LYZ$_{900}$, and DI-LYZ$_{990}$). As a particularly dramatic example, DI-LYZ$_{990}$ presented an antiparallel β-sheet that was stable for almost the whole of the 1 ns relaxation trajectory (residues 100−114).

Transient Secondary Structures. Many of the secondary structures undergo several transitions during the relaxation trajectories. They may become either another secondary structure or a coiled coil. *All* of the trajectories presented transients in which secondary fold, native or non-native, was lost and later

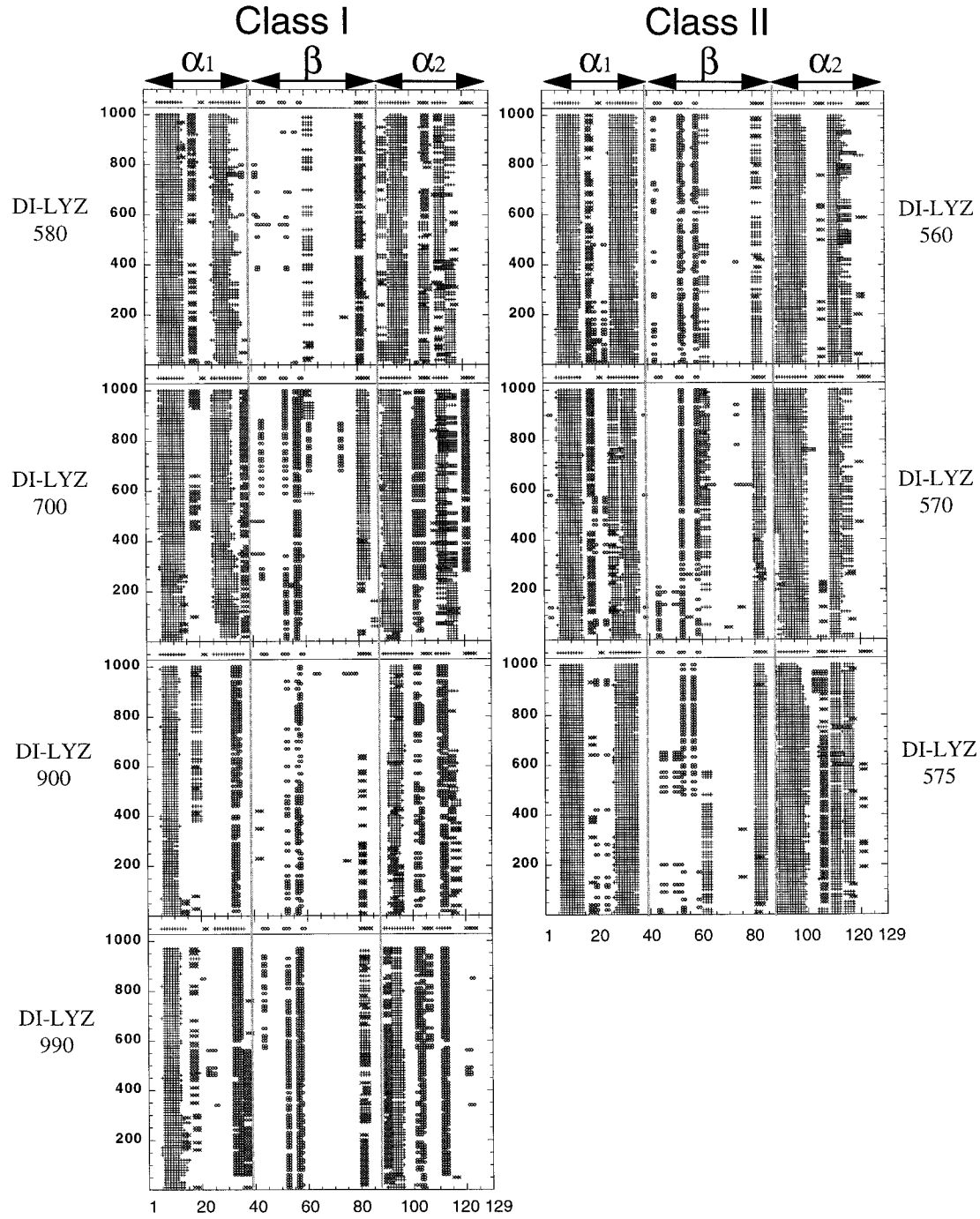

Figure 5. Time evolution of α-helices (+), 3_{10}helices (×), π-helices (×ff), and β-strands (◇) for the trajectories DI-LYZ$_{560}$, DI-LYZ$_{570}$, and DI-LYZ$_{575}$ (Class II), and DI-LYZ$_{580}$, DI-LYZ$_{700}$, DI-LYZ$_{900}$, and DI-LYZ$_{990}$ (Class I). Each of the 129 amino acids is represented for each trajectory. The top of each plot includes as a reference information on the secondary structure present in the native structure.

on retrieved. All of the amino acids in the sequence are involved in these transitions for at least one of the trajectories, the only exceptions being the above-mentioned amino acids never forming secondary structure. As an example, the relaxation of the DI-LYZ species often presented the formation of two transient β-strands involving residues 19−20 and 23−24 (Figure 5) forming an antiparallel β-sheet.

Discussion

The computational perturbation applied on lysozyme in vacuo by Reimann et al.[13] has yielded conformers differing greatly in their degree of unfolding (Figure 6, left-hand column). Conformers characterized by RMSD ≤ 4−5 Å and retaining some native secondary structure are analogous to conformers with similar RMSD values identified as "folding intermediates" in thermal denaturation MD studies.[11,27−29] Conformers with RMSD ≥ 7 Å or so are unfolded beyond what is typical of conformations commonly explored in other protein unfolding/refolding studies,[11,12,30] including those carried out on lysozyme in the condensed phase.[31,32] However, these extremely denatured conformers are a reasonable model for in vacuo conformers

(27) Ladurner, A. G.; Itzhaki, L. S.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 8473−8478.
(28) Li, A.; Daggett, V. *J. Mol. Biol.* **1996**, *257*, 412−429.
(29) Vijayakumar, S.; Vishveshwara, S.; Ravishanker, G.; Beveridge, D. L. *Biophys. J.* **1993**, *65*, 2304−2312.
(30) Marchi, M.; Ballone, P. *J. Chem. Phys.* **1999**, *110*, 3697−3702.

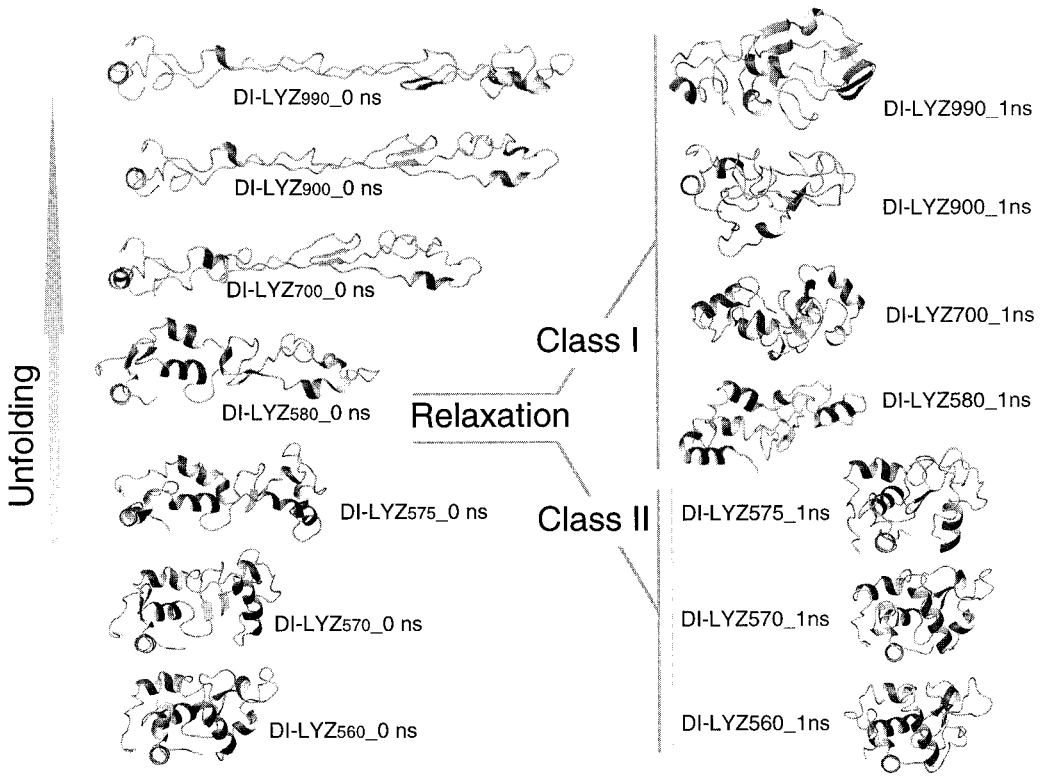

Figure 6. 6. The structures of a set of representative conformers are depicted. The structural effects of the unfolding forces can be appreciated by examining the left column (vertical arrow). Labels indicate trajectory and relaxation time at which the snapshot was taken. The relaxation of the structures on the left-hand side have yielded the ones on the right-hand side. Although the relaxed conformers of Class I present a high degree of compactness compared to their unfolded counterparts, structural differences among them can be appreciated.

studied experimentally.[9,33,34] For example, the lengths of the denatured conformers, ~100 Å, are comparable to lengths deduced in energetic surface imprinting experiments,[34] 135 ± 30 Å. Also, the conformers have orientationally averaged collision cross sections of ~1800 Å^2, compatible with the results of ion-drift mobility experiments.[9]

All of the structures shown in Figure 6 (left-hand column) are thus interesting subjects for MD relaxation studies in vacuo. Such studies probe the ability of protein intramolecular forces—the forces intrinsic to the protein—to relax an unfolded structure and lead to the formation of significant and persistent tertiary, subdomain, and secondary structures. These forces are represented in a well-known way by the GROMOS-87 potential.[14,21] The ensemble of relaxation studies presented in this work leads to a potentially counterintuitive result: in the absence of solvent, lysozyme can relax toward native and nativelike structures, supporting the suggestion made by Wolynes[35] that biomolecules may fold in vacuo.

1. Folding-like Behavior of Lysozyme in Vacuo. Two classes of resultant conformers were found in the relaxation processes: one comprising non-native structures (Class I) clustered around a common origin with similar hyperspherical RMSD "radius", and another one comprising native and near-native structures (Class II). Perusal of a set of unfolded and relaxed conformers in Figure 6 visually suggests that the relaxations comprise a folding process. For example, Class I relaxed conformers were notably compact compared with their unfolded precursors (see Figure 2), while at the same time displaying persistent secondary and tertiary structures (Figures 4−6). Class II relaxed conformers, originating from unfolded states similar in terms of RMSD to ones identified as transition states in solution,[11,27,28] were not only compact, but they also even bore a close resemblance to the native one (Figures 1 and 6).

DI-LYZ$_{570}$ shows a rapid relaxation to a final structure which, on the basis of RMSD, can be considered to be completely refolded (though there is some incorrect secondary structure). By scrutinizing the behavior of DI-LYZ$_{570}$ in Figures 1 and 3, it can be seen that a certain degree of overall unfolding was manifested in the seeding structure without the (sub)domains being appreciably unfolded. On relaxation, the seeding structure evolved to RMSD = 2.6 Å in about 15 ps, while the (sub)-domains presented RMSD ≤ 2 Å. In this case the unfolding reflected mainly a relative reorientation of the (sub)domains, as in a hinge-like motion,[13] and the relaxation reflected a readjustment of these orientations.

Even though only seven relaxation trajectories were carried out in this study, in one case (DI-LYZ$_{700}$) the β-domain, originally unfolded to RMSD 6.6 Å, recaptured its native structure (Figure 3). Thus, a relaxation behavior uncovering a folding process is also observed at the (sub)domain level.

According to the portrayal of all the relaxed conformers in Figure 6, considerable structural differences exist among them. Yet, during the relaxation trajectories, the total energy has decreased drastically (Figure 7) to values which, considering fluctuations, do not allow any strong distinction to be made

(31) Mark, A. E.; van Gunsteren, W. F. *Biochemistry* **1992**, *31*, 7745−7748.
(32) Hünenberger, P. H.; Mark, A. E.; van Gunsteren, W. F. *Proteins: Struct., Funct., Genet.* **1995**, *21*, 196−213.
(33) Gross, D. S.; Schnier, P. D.; Rodriguez-Cruz, S. E.; Fagerquist, C. K.; Williams, E. R. *Proc. Natl. Acad. of Sci. U.S.A.* **1996**, *93*, 3143−3148.
(34) Reimann, C. T.; Sullivan, P. A.; Axelsson, J.; Quist, A. P.; Altmann, S.; Roepstorff, P.; Velázquez, I.; Tapia, O. *J. Am. Chem. Soc.* **1998**, *120*, 7608−7616.
(35) Wolynes, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 2426−2427.

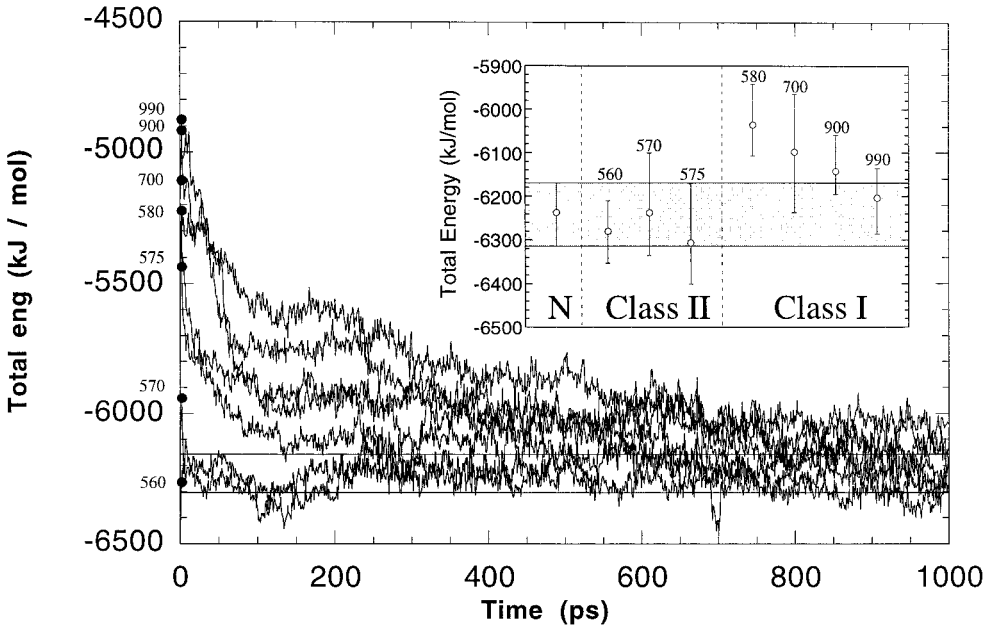

Figure 7. 7. Time profile of the total energy for relaxation trajectories DI-LYZ$_{560}$, DI-LYZ$_{570}$, DI-LYZ$_{575}$, DI-LYZ$_{580}$; DI-LYZ$_{700}$, DI-LYZ$_{900}$, and DI-LYZ$_{990}$ relaxation runs. Fluctuations in energy for the last 100 ps of each trajectory are shown in an inset, the gray bar covers the energy-fluctuation region of the native structure in vacuo. The overlapping with this region is apparent for all the trajectories besides DI-LYZ$_{580}$.

between Class II (nativelike) and Class I (non-native) conformers (Figure 7, inset). Not even the relaxation of the native state in vacuo yielded conformers distinctly lower in energy than all the others ("N" in Figure 7). Random collapse would also lead generally to reduced total energy. However, the close similarities between the final energies of Class II and Class I conformers also support the hypothesis that all the relaxed conformations shown in Figure 6 constitute viable folded conformers in vacuo.

Synchronous events in relaxation imply nonrandom, folding-like behavior, but asynchronous events do not necessarily imply a random collapse. Some features that point to folding include cooperative sigmoidal transitions observed at the intra-(sub)-domain level (Figure 3). Moreover, examples of synchrony are observed when comparing the behavior of β and α_2 (Figure 3). Some asynchrony is also observed, especially in the behavior of α_1 with respect to the other (sub)domains (Figure 3), but asynchrony does *not* imply a random collapse. For example, experimental studies on the folding of solution-phase hen lysozyme by Dobson and co-workers[15,36,37] show that lysozyme does not become organized in a single cooperative event but rather that the different domains of the protein become stabilized with very different kinetics. Furthermore, different populations of lysozyme molecules refold along different pathways.

2. Behavior of Secondary Structure on Relaxation and Connection to a Folding-Like Process. It should be noticed that (sub)domain transitions were not strongly correlated with changes in the fluctuation patterns of secondary structural features (compare Figures 3 and 5). Moreover, secondary structures could fluctuate to a certain extent, even after the final relaxed (sub)domain and/or tertiary fold was achieved. As an example, for DI-LYZ$_{700}$ the β-domain relaxed to a structure which was nativelike as indicated by RMSD = 2 Å, but it still displayed notable fluctuations in its secondary structure.

The transient fluctuations of secondary structure in vacuo are not a consequence specifically of the environment: this behavior has also been observed in solution-phase lysozyme.[32] Early in a thermal denaturation process of lysozyme before significant unfolding had initiated,[32] picosecond-scale transitions in secondary structure were observed. This suggests that fluctuations in secondary structure may be a general phenomenon occurring not only during relaxation/folding, but also even under steady-state conditions.

A key question pertaining to the protein folding problem concerns the extent to which transient secondary structures resemble the *native* ones during protein folding. Finkelstein and Ptitsyn[38] suggested that nativelike secondary structure can appear transiently in the denatured state, significantly reducing the number of degrees of freedom that may need to be searched in order to achieve a folded state. The relaxation processes in vacuo show some characteristics that can be related to this idea. However, our results indicate that *both* nativelike and non-native transient secondary structures occur during the relaxation of the unfolded DI-LYZ. As an example, Class I conformers display fluctuating character of the nativelike helices, residues 5−9, 25−36, and 87−100 (H1, H2, and H4 respectively[20]), while also displaying non-native fluctuating β-strand and β-sheet structures (see Figure 5; residues 100−120; DI-LYZ$_{700}$, DI-LYZ$_{900}$, and DI-LYZ$_{990}$). Moreover, most of the residues are found to participate in secondary structures that are different from the native ones at some time or other during relaxation. The apparent discrepancy with the picture of Finkelstein and Ptitsyn[38] is solved by the hypothesis that the relaxed Class I conformers are plausible folded structures in vacuo. These structures are different from the native structure in solution, so that it is entirely reasonable that they contain many examples of metastable, fluctuating secondary structures of all types.

The formation of non-native, metastable secondary structures illustrates how a peptide with the *same* amino acid sequence can result in the formation of a β-sheet *or* an α-helix. A similar structural adaptability was observed experimentally by Kabsh and Sander:[39] in a study of sequence homologies among different

(36) Miranker, A.; Radford, S. E.; Karplus, M.; Dobson, C. M. *Nature* **1991**, *349*, 633−636.
(37) Radford, S. E.; Dobson, C. M.; Evans, P. A. *Nature* **1992**, *358*, 302−307.
(38) Finkelstein, A. V.; Ptitsyn, O. B. *J. Mol. Biol.* **1976**, *103*, 15.
(39) Kabsch, W.; Sander, C. *Proc. Natl. Acad. Sci. U.S.A.* **1984**, *81*, 1075−1078.

proteins, they reported on a pentapeptide that was found to be a part of an α-helix in one protein while forming a part of a β-strand in another protein. The difference with the present work is that we observed different secondary folds for the *same* amino acid sequence in the *same* protein. Therefore, conformational changes in secondary structure may be attributed to changes in the micro-environmental conditions, i.e., different folds.

In experimental studies on protein folding, circular dicrohism (CD), and hydrogen−deuterium exchange techniques are typically used to monitor the amount of secondary structure retained by the protein probed.[40−42] The amount of retained (or gained) secondary structure is in turn thought to be related to the degree of folding achieved by the unfolded substates.[43,44] Interestingly, Figure 4 shows that in vacuo such is not the case: the amount of secondary structure does *not* correlate with the degree of folding, and metastable but non-native secondary structures can form. If many of the general results obtained in the present study in vacuo would also be representative of lysozyme in solution, then a strong CD signature may not generally be assumed to reflect the attainment of a nativelike structure without further evidence.

Secondary structures that are non-native but stable occur profusely in the DI-LYZ_{700}, DI-LYZ_{900}, and DI-LYZ_{990} trajectories; these display an increased amount of β-strand with respect to the native structure. The presence of β-strands in folded intermediates of human lysozyme, in regions that are α-helix in the native fold, has been associated with the formation of amyloid fibrils that underlay a range of fatal diseases.[45]

Final Remarks

A major question arises with the type of simulations reported here: to what extent is one observing some generic relaxation dynamics that any polymer would undergo, as opposed to an actual folding process. In the present work, we have attempted to answer this question with the aid of standard structural descriptors. A *nonrandom* relaxation process is evidenced by: ability to form compact, quasi-native structures; ability of a (sub)domain to fold to the native structure even if the other (sub)domains do not; cooperative (sub)domain relaxation; synchronous and asynchronous events in relaxation; formation of persistent secondary structural motifs; and relaxation to states of similar total energy, only slightly higher than that of the native state.

Further key insights can be obtained with the aid of complementary shape and topological descriptors, which we have employed with promising results to characterize the fold diversity in proteins related to the one examined here.[46] On the basis of the topological descriptors, additional clues of a folding process are uncovered.[47,48]

Another way to test whether the relaxations we monitor are random is to consider the behavior of proteins with the *same* backbone conformation but with a *different* amino acid sequence. If an identical relaxation behavior would be observed, i.e., if this behavior is not sequence-dependent, then the relaxation dynamics is generic rather than being indicative of a folding process. The simulations (manuscript in preparation[49]) using the most unfolded conformers but with a totally different sequence of the same length (from a protein having a well-defined crystal structure) show the same initial collapse, but subsequently, the difference in sequence shows up in a distinctive manner. Therefore, it is concluded that this computer-simulated relaxation is sequence dependent. A detailed analysis of these results will be communicated elsewhere.[49]

In summary, taking all of our results together, in particular the dynamical, structural, and energetic information reported here, one can be reasonably sure that this type of simulation yields nontrivial results which, at least for lysozyme, strongly suggest that folding is a property of the protein in absence of solvent effects. This idea is being experimentally confirmed for protein ions in vacuo.[50]

In vacuo, both compact and elongated lysozyme species are experimentally observed to exist,[9,33,34] giving consistency to our simulation results. However, in solution, lysozyme has also been demonstrated to adopt a denatured form rich in β-sheet,[45] as also seen in our simulation results. Therefore, one may conjecture that conformers similar to the in vacuo denatured but folded members of Class I may also exist in solution. Note that relaxed structure DI-LYZ_{580} (highest total energy, inset of Figure 7) may equally be a representative of the transition state conformations connecting Class II to Class I structures.

Acknowledgment. The authors thank the Swedish Research Council for Engineering Sciences (TFR), the Swedish National Board for Industrial and Technical Development (NUTEK), the Swedish Natural Sciences Council (NFR), and Dr. Gustavo Arteca for helpful discussions.

JA990911C

(40) Roder, H.; Elöve, G. A.; Englander, S. W. *Nature* **1988**, *335*, 700−704.
(41) Katta, V.; Chait, B. T. *J. Am. Chem. Soc.* **1993**, *115*, 6317−6321.
(42) Miranker, A.; Robinson, C. V.; Radford, S. E.; Dobson, C. M. *FASEB J.* **1996**, *10*, 93−101.
(43) Baldwin, R. L. *J. Biomol. NMR* **1995**, *5*, 103−109.
(44) Fink, A. L.; Oberg, K. O.; Seshardi, S. *Folding Des.* **1997**, *3*, 19−25.
(45) Booth, D. R.; Sunde, M.; Bellotti, V.; Robinson, C. V.; Hutchinson, W. L.; Fraser, P. E.; Hawkins, P. N.; Dobson, C. M.; Radford, S. E.; Blake, C. C. F.; Pepys, M. B. *Nature* **1997**, *385*, 787−793.
(46) Arteca, G. A.; Tapia, O. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 642−649.
(47) Arteca, G. A.; Velázquez, I.; Reimann, C. T.; Tapia, O. *Phys. Rev. E* **1999**, *59*, 5981−5986.
(48) Arteca, G. A.; Velázquez, I.; Reimann, C. T.; Tapia, O. *J. Chem. Phys.* **1999**, *111*, 4774−4779.
(49) Arteca, G. A.; Paulino, M.; Reimann, C. T.; Tapia, O., manuscript in preparation.
(50) Jarrold, M. F. *Acc. Chem. Res.* **1999**, *32*, 360−367.