Protein Science (1998), 7:860–874. Cambridge University Press. Printed in the USA.

Molecular dynamics simulations of hydrophobic collapse of ubiquitin

DARWIN O.V. ALONSO AND VALERIE DAGGETT

Department of Medicinal Chemistry, University of Washington, Seattle, Washington 98195-7610

(RECEIVED August 28, 1997; ACCEPTED December 19, 1997)

Abstract

Nine nonnative conformations of ubiquitin, generated during two different thermal denaturation trajectories, were simulated under nearly native conditions (62 °C). The simulations included all protein and solvent atoms explicitly, and simulation times ranged from 1–2.4 ns. The starting structures had α-carbon root-mean-square deviations (RMSDs) from the crystal structure of 4–12 Å and radii of gyration as high as 1.3 times that of the native state. In all but one case, the protein collapsed when the temperature was lowered and sampled conformations as compact as those reached in a control simulation beginning from the crystal structure. In contrast, the protein did not collapse when simulated in a 60% methanol: water mixture. The behavior of the protein depended on the starting structure: during simulation of the most native-like starting structures ($\leq$5 Å RMSD to the crystal structure) the RMSD decreased, the number of native hydrogen bonds increased, and the secondary and tertiary structure increased. Intermediate starting structures (5–10 Å RMSD) collapsed to the radius of gyration of the control simulation, hydrophobic residues were preferentially buried, and the protein acquired some native contacts. However, the protein did not refold. The least native starting structures (10–12 Å RMSD) did not collapse as completely as the more native-like structures; instead, they experienced large fluctuations in radius of gyration and went through cycles of expansion and collapse, with improved burial of hydrophobic residues in successive collapsed states.

Keywords: conformational sampling; folding pathways; hydrophobic collapse; molecular dynamics simulations

In general, native states of globular proteins are compact, show preferential burial of hydrophobic residues, have paired hydrogen bond donors and acceptors, and adopt one consensus three-dimensional structure. The relative importance of forces that stabilize the native state is still debated, but reduction of nonpolar surface area, electrostatics, hydrogen bonding, and the conformational preferences of the amino acids are widely accepted as contributors to protein structure and stability. The importance of the pathways from unfolded to native states stems from the combinatorics of the protein folding problem; many small globular proteins unfold and refold reversibly to the same native structure (Anfinsen, 1973; and for exceptions see Baker & Agard, 1994), but they cannot search all possible conformations for the most stable state in a reasonable period of time (Levinthal, 1969; and see review by Dill, 1990).

In light of these findings, there must be mechanisms to limit the conformational search to particular regions of conformational space or to narrow the pathways between the unfolded and native states. However, the mechanisms and pathways by which a protein proceeds from unfolded states to the native state are disputed. Two simple models are commonly considered, the framework model and the hydrophobic collapse model. According to the framework model, native-like secondary structure is first formed locally (in sequence) followed by a more complete collapse to the native state involving the formation of topologically distant (in sequence) tertiary contacts (Ptitsyn & Rashin, 1975; Kim & Baldwin, 1982, 1990). On the other hand, the hydrophobic collapse model predicts that the first step in protein folding involves collapse to bury nonpolar residues, and subsequently, the protein rearranges to form secondary structure and finally the native state (Ptitsyn, 1987; Baldwin, 1989). Not surprisingly, the fundamental driving forces for these two extremes differ; local conformational preferences and hydrogen bonding are invoked to explain the framework model and the hydrophobic effect is presumed to drive collapse. Of course, contributions from all of the above are both possible and probable.

To paraphrase Fersht and colleagues, the early events in protein folding currently make up the most obscure and controversial area of folding because these events have been inaccessible to experiment (Prat Gay et al., 1994; also see Evans & Radford, 1994; Engelhard & Evans, 1996), although experiments are beginning to access these time scales (Plaxco & Dobson, 1996; Wolynes et al., 1996; Eaton et al., 1997 and references therein). Computational techniques may also provide information about these experimentally-inaccessible events between the unfolded states and late folding

Reprint requests to: Valerie Daggett, Department of Medicinal Chemistry, University of Washington, H-174B, Box 357610, Seattle, Washington 98195-7610; e-mail: daggett@u.washington.edu.

intermediates. Most computational methods treating collapse and folding pathways have relied on simple models that usually only incorporate a limited number of energy parameters or simplified representations of the proteins and its environment (Levitt & Warshel, 1975; Covell & Jernigan, 1990; Skolnick & Kolinski, 1990; Moult & Unger, 1991; Guo et al., 1992; Hinds & Levitt, 1992, 1994; O'Toole & Panagiotopoulos, 1992; Camacho & Thirumalai, 1993; Karasawa et al., 1993; Covell, 1994; Hao et al., 1993; Sali et al., 1994; Bryngelson et al., 1995; Dill et al., 1995). These low-resolution techniques have the advantage that the energetics of many distinct conformations can be evaluated, and in some cases exhaustive enumeration of discrete approximate conformations is possible.

In contrast to such simplistic models, molecular dynamics simulations of proteins including all atoms in explicit solvent can, in principle, depict early stages of folding without assuming a particular dominant driving force and without the limitations of lattice representations. To objectively investigate events in folding, one must employ the same force field used for studies of the native state and unfolding process. That is, new terms to drive the process should not be included. Instead, the noncovalent interactions that represent the largest contributions to the forces that guide the motion of the protein in these other situations must also drive collapse. For example, the hydrophobic effect should result from the balance of forces in the standard pairwise interactions between different atoms, and in this work they are *not* introduced *ad hoc.* Similarly, a simple Coulombic term with standard parameters is used for representing hydrogen bonds and other electrostatic interactions.

To date, there have not been any molecular dynamics solution simulations of a protein addressing the possible early stages of folding. Instead, molecular dynamics simulations have focused on the unfolding process and characterization of predominantly early unfolding events (Daggett & Levitt, 1992, 1993; Mark & van Gunsteren, 1992; Daggett, 1993; Tirado-Rives & Jorgensen, 1993; Vishveshwara et al., 1993; Caflisch & Karplus, 1994, 1995; Li & Daggett, 1994, 1996, 1998; Alonso & Daggett, 1995; Hunenberger et al., 1995; Daggett et al., 1996; Laidig & Daggett, 1996; Storch & Daggett, 1996; Tirado-Rives et al., 1997). These early kinetic unfolding events are expected to occur late in the folding process. Recently, such molecular dynamics simulations have also been used to characterize the dynamical and structural properties of the denatured state (Bond et al., 1997; Kazmirski & Daggett, 1998). Nevertheless, simulations starting from unfolded structures may provide insight into the forces driving collapse and/or secondary structure formation. Recently, using a different, but complementary approach, Bozcko and Brooks (1995) and Guo et al. (1997) have investigated the energetics of this process and outlined the general folding in terms of thermodynamic coordinates.

Here, we describe unconstrained, all-atom molecular dynamics simulations with explicit water, in which expanded nonnative starting structures of ubiquitin collapse to a common state. The starting structures were generated by simulated thermal denaturation of the crystal structure (Fig. 1). The nonnative starting structures had α-carbon root-mean-square deviations (RMSDs) from the crystal structure ranging from 4–12 Å and radii of gyration 1.3 times the native value. During the subsequent simulations under quasi-native conditions (335 K and low pH), all but one of these nonnative starting structures collapsed to the same radius of gyration as that of the simulation starting with the crystal structure. Elevated temperature was used for the quenched simulations to facilitate conformational transitions relative to 298 K, and 335 K and low pH was chosen because the protein is quasi-stable under these conditions (Harding et al., 1991). The two starting structures that were very similar to the native structure (*native-like*, $4 \leq \text{RMSD} \leq 5$ Å) became more native-like by all criteria considered, and appear to represent the quasi-stable "native" state under these conditions of elevated temperature and low pH. Intermediate starting structures (*intermediate*, $5 < \text{RMSD} \leq 10$ Å) formed some native secondary and tertiary structure, but they did not refold. The most nonnative starting structures (*unfolded*, $10 < \text{RMSD} \leq 12$ Å) did not acquire native structure, and instead experienced nonspecific collapse and continued tosample different regions of conformational space by expanding and re-collapsing.

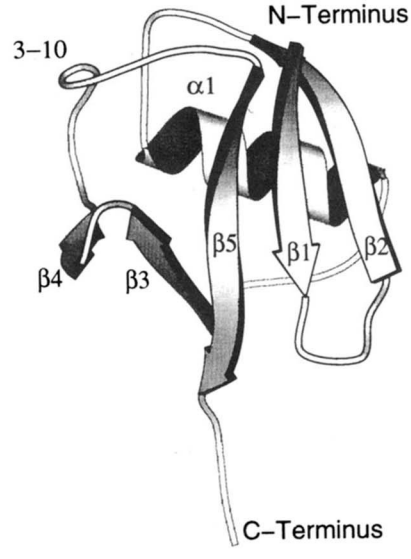

Fig. 1. Ribbon representation of the ubiquitin crystal structure (Vijay-Kumar et al., 1987) highlighting the secondary structure. The β-strands are numbered sequentially as they occur in the sequence: β1, residues 1–8; β2, residues 11–18; β3, residues 40–45; β4, residues 45–48; and β5, residues 68–72. The α-helix comprises residues 23–33. There is also a short 3–10 helix between residues 50–54.

Results

We report the results of simulations of nine different nonnative starting structures of ubiquitin under conditions that favor folding. The unfolded starting structures were generated from high temperature denaturation simulations of the crystal structure of native ubiquitin. A brief description of the unfolding pathways is given to illustrate the properties of the extracted (starting) structures for the quenched simulations. The simulations of the nonnative starting structures are then presented in three separate groups based on the degree of unfolding in the starting structures: (1) *native-like structures* (slightly expanded from the crystal state but with many of the native topological contacts); (2) *intermediate structures* (expanded relative to the crystal structure with few native topological contacts, designated intermediate because they fall roughly between groups 1 and 3); and (3) *nonnative, unfolded structures.*

The unfolding simulations and generation of starting structures

The starting structures for the collapse simulations were extracted from a broad sampling of two low pH denaturation simulations at

498 K, designated D1 and D2. The α-carbon RMSD from the crystal structure for these two denaturation runs and for 298 K native control simulations (N298 at low pH and N298n at neutral pH) is given in Figure 2. The RMSD for the N298n control simulation was 1.8 Å over the last 0.5 ns of the 2.4 ns trajectory (Table 1). Considering a longer time period, 1.0–2.4 ns, the RMSD ranged from a short-lived maximum of 2.5 Å at 1.3 ns to 1.2 Å at 1.9 ns, with an average of 1.8 Å. Low pH destabilized the protein, yielding a RMSD of 2.7 Å for the N298 control after 1.6 ns. The points along the high temperature denaturation trajectories from which the structures were extracted are indicated, and these structures are given as insets to Figure 2, with labels from I-A to I-I. The deviation from the starting structure provides an overview of the two denaturation simulations but no details on the specific structure lost.

In the D1 simulation, the protein lost helix docking and tertiary contacts within 0.05 ns (50 ps), which was then followed by a loss of interactions between β-strands 1 and 2 [β(1–2)] in the next 25 ps (see Fig. 1 for secondary structure designations). Contacts between β(1–5) were disrupted at ~0.14 ns. Though the β(1–2) main chain contacts were lost early, they fluctuated with time and periodically dropped below 5 Å. In contrast, the contacts between β(1–5) remained consistently above 15 Å after 0.25 ns.

In the second denaturation simulation (D2), the helix docking contacts were also lost very early, and by 0.05 ns the distances between the ends of the helix and the closest point on the β-sheet had increased from 5 to 7 Å at the N-terminus, and to 10 Å at the C-terminus. Denaturation of the β-sheet proceeded more slowly than in D1. The contacts between β(1–5) remained intact for the first 0.18 ns of the simulation. Then, all of these contacts broke within 0.02 ns of one another. Similarly, the contacts between β(3–5) broke at approximately 0.3 ns. In contrast, β(1–2) immediately began fraying from the top of the sheet, as depicted in Figure 1 (residues 2 and 16), but it was not until 0.45 ns into the simulation that all of the contacts between the strands were lost.

Nine structures were extracted from these denaturation simulations for investigation under near-native conditions (Fig. 2). The α-carbon RMSD from the crystal structure and the radius of gyration (R_g) of the extracted structures are given in Table 1. The starting structures spanned the range from being slightly expanded, but native-like, to almost completely unfolded. The percentage of repeating native β-structure in the plateau regions of D1 and D2

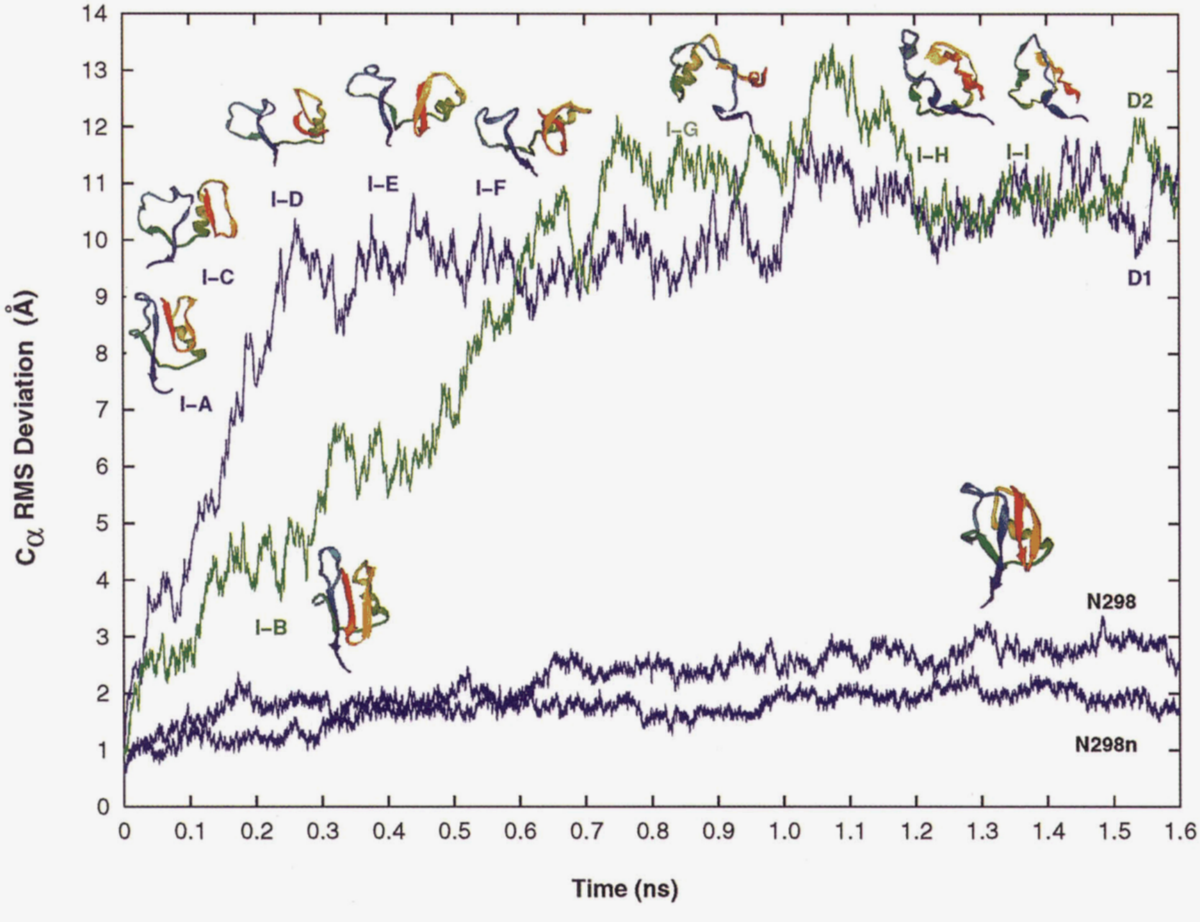

Fig. 2. The α-carbon RMSD from the crystal structure plotted as a function of simulation time for the native 298 K control simulation at low (N298) and neutral pH (N298n), and the two denaturation simulations at 498 K (D1 and D2). Optimal superposition of structures to remove rotational and translational motion was achieved using the method described by Kabsch (1976) and only residues 1–72 are included in the calculation. Main chain ribbon traces of each of the starting structures for the quenched simulations are illustrated and labeled. The protein structures are colored from red at the N-terminus through blue at the C-terminus. Positions of native secondary structure are depicted as a thickening of the ribbon (this does not mean that the structure was native, however).

Table 1. *Summary of simulations and overall properties observed*

	Properties of starting structures[a]				Simulations in water[b]				Simulations in 60% methanol[b]			
	Origin	Quench time (ns)	Initial R_g (Å)	Initial RMSD (Å)	Simulation time (ns)	Final R_g (Å)	Best & (final) RMSD (Å)	Collapse observed?	Simulation time (ns)	Final R_g (Å)	Best RMSD (Å)	Collapse observed?
Native starting structures												
N298n	Xtal	—	11.5	—	2.4	11.6	(1.8)	—	—	—	—	—
N298	Xtal	—	11.5	—	1.6	12.1	(2.7)	—	—	—	—	—
N335	Xtal	—	11.5	—	2.4	12.5	(4.0)	—	—	—	—	—
Native-like starting structures												
I-A	D1	0.06	12.9	4.0	1.0	12.4	3.1 (3.7)	Yes	1.0	13.3	4.0	No
I-B	D2	0.25	13.3	5.0	1.0	12.4	3.1 (4.2)	Yes	0.1	13.5	5.0	No
Intermediate starting structures												
I-C	D1	0.20	14.5	7.6	2.4	12.3	5.7 (7.5)	Yes	2.4	14.4	7.6	No
I-D	D1	0.25	14.6	9.5	2.4	12.6	6.1 (7.3)	Yes	0.4	16.1	9.5	No
I-E	D1	0.35	14.2	9.1	1.4	13.4	8.2 (8.7)	Fluctuating				
Unfolded starting structures												
I-F	D1	0.45	14.6	10.1	1.0	15.3	9.3 (10.2)	No				
I-G	D2	0.80	14.1	11.1	2.4	13.1	10.0 (10.9)	Fluctuating				
I-H	D2	1.50	13.3	10.5	2.4	13.5	10.1 (12.1)	Fluctuating				
I-I	D2	1.53	14.2	11.2	1.0	12.9	10.5 (11.1)	Fluctuating				

[a]The designation for each simulation is given in the first column: "N" indicates that the simulation began from the crystal structure (Xtal), "I" that it began from a nonnative structure sampled from a high temperature denaturation trajectory (D1 or D2). Quench time: the time during the denaturation simulations when a structure was extracted for simulation at lower temperature. Initial R_g: the radius of gyration of the starting structure, calculated as described in the legend to Figure 5. Initial RMSD; the α-carbon RMS deviation from the crystal structure, calculated as described in the legend to Figure 2.

[b]Simulation time: the duration of the native (either at 298 or 335 K) and quenched simulations under quasi-native conditions of 335 K and low pH or in 60% methanol. Final R_g: the average radius of gyration over the last 0.5 ns of each simulation. Best RMSD: the minimum α-carbon RMS deviation from the crystal structure observed during the simulation (not given for simulations starting from the crystal structure). Final RMSD: the average α-carbon RMSD from the crystal structure over the last 0.5 ns of the simulation, given in parentheses. The C_α RMSD values decrease by 1–2 Å when compared to the control at 335 K, N335.

was nearly zero (3%), but the α-helix remained ~70% intact through approximately the first nanosecond of both the D1 and D2 simulations, after which it dropped and oscillated between adopting a single helical turn and 0% helix. Discussion of the specific structure present in the starting structures is deferred until later, where we examine five of the collapse simulations in detail.

The quenched simulations

This section describes the results of simulations of the nonnative starting structures under native-like conditions. A slightly elevated temperature (335 K, 62 °C) was used to facilitate conformational transitions while remaining near-native conditions, as judged by experimental studies (Harding et al., 1991). Unless otherwise noted, all simulations were performed in water at 335 K. Other simulations are also presented for comparison but are not discussed in depth: N298 and N298n (introduced above), and I-X_M, simulations beginning with partially unfolded structures at 335 K in 60% methanol (Alonso & Daggett, 1995), where X denotes the starting structure and the subscript M is used to distinguish the methanol simulations from those in water.

Briefly, our results fall into three main categories: (1) the native-like and compact starting structures (I-A and I-B) collapsed and regained most native properties; (2) the intermediate starting structures (I-C, I-D, I-E) collapsed to a compact state with many native properties as well as substantially misfolded sections; and (3) the most nonnative structures (I-F, I-G, I-H, and I-I) collapsed but then fluctuated between collapsed and expanded states, with the exception of I-F, which did not collapse. To simplify the discussion of the various results, we discuss representative structures from these three groups rather than present the details of all simulations. This breakdown along the folding/unfolding reaction coordinate is similar to that described by Matthews (1993) in reviewing experimental work.

Native-like starting structures

During the course of the simulations of I-A and I-B, the protein gained native secondary structure and native hydrogen bonds, and the RMSD from the crystal structure decreased. However, the I-A and I-B starting structures were native-like based on RMSD from the crystal structure (4 and 5 Å, respectively) and radius of gyration (1.1 times native) (Table 1). Nonetheless, there were also significant differences between these starting structures and the native state. As mentioned in the previous section, the helix pulled away from the β-sheet early in the denaturation simulations. The distances between residues at either end of the helix and residues in β2 in I-A increased to about 11 Å, as compared to 5 Å in the crystal structure. The helix pulled away further but unevenly in I-B; the distance between the N-terminus of the helix and β2 increased to 11 Å, whereas the distance from the C-terminus of the helix increased to 20 Å.

The C_α RMSD of I-A decreased to 3.5 Å during the simulation (averaged over the last 100 ps) with a minimum value of 3.1 Å, illustrating that the most native-like structures obtained during the quenched simulations remained dynamic at 335 K, and were not locked into a particular conformation after collapsing. However, the minimum RMSD was maintained when the structure were quenched to 298 K, as was also found for the other simulations discussed below (data not presented). The final I-A structure was evaluated using the program Procheck (Laskowski et al., 1993) as an independent and objective evaluation of the structure, and the output from the "secondary structure and estimated accessibility" portion of this program (Fig. 3) illustrates that the protein acquired secondary structure in β1, β2, and the 3–10 helix. Similarly, comparison of the hydrogen bonding patterns in the initial and final 1 ns I-A structures and the crystal structure illustrates that native hydrogen bonds were acquired during the course of the simulation, although some hydrogen bonds were long ($3 \leq d \leq 4$ Å) due to the increased temperature and associated expansion of the protein at 335 K (Fig. 3). But, since hydrogen bonds are represented by a Coulombic potential in our simulations, there are still appreciable electrostatic interactions between charged groups at 4 Å. Nevertheless, eight hydrogen bonds were acquired between β(1–2), for example. The hydrogen bonds not associated with regular secondary structure, those between residues 22 and 52, did not reform during the I-A simulation, and these were also lost in the N335 control simulation.

During the I-B simulation, the contacts between the helix and β-sheet were re-established (see the side view in Fig. 4B). The β-sheet in the starting structure was more regular and planar than the curved surface more typical of a β-sheet; however, during the simulation, the sheet became more native-like and, hence, less uniform. The final structure differed from the crystal structure in two main aspects. The β-sheets were not precisely native, as characterized by (ϕ,ψ) angles; this can be seen in Figure 4 as the uneven shading in the ribbon depicting β1. The number of main chain–main chain hydrogen bonds increased from 33 to 51 during the simulation and was comparable to that of N298n (also at 51). Furthermore, the RMSD decreased by ~2 Å during the simulation comparable to that of the N335 control simulation (Fig. 4A) and many native contacts were established (Fig. 4C).

Intermediate starting structures

General properties

We dub three structures (I-C, I-D, I-E) extracted from the denaturation trajectory as "intermediate" nonnative structures. Their RMSD from the crystal structure ranged from 7.6–9.5 Å (Table 1, Fig. 2). Not only did these partially unfolded starting structures differ from the crystal structure, but they were structurally heterogeneous, differing greatly from one another. For example, the α-carbon RMSD between the I-C and I-D starting structures was 6 Å, and that between I-C and I-E was >7 Å.

In two of these cases (I-C and I-D), the protein collapsed to approximately the same radius of gyration as that of the N335 simulation (Table 1; Fig. 5A). The rate of collapse differed, how-

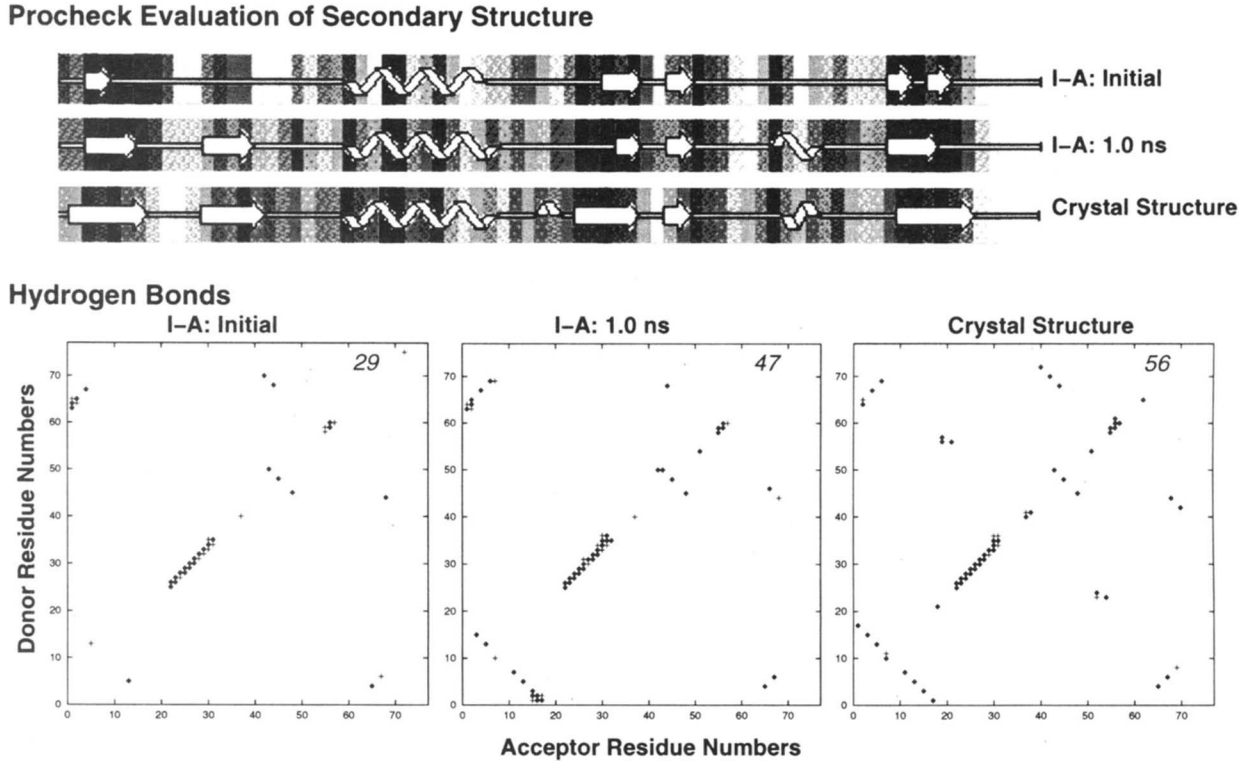

Fig. 3. Secondary structure and hydrogen bond content for the I-A simulation. Secondary structure was evaluated using the program Procheck (Laskowski et al., 1993). The hydrogen bond maps show interactions between main chain oxygens listed on the *x*-axis and backbone amide hydrogens on the *y*-axis. A 3 Å cutoff was used for the solid squares and 4 Å was used for the crosses. No angular constraints were used. The number of pairs under 3 Å is given explicitly in the right-hand corner. For comparison, the final structure from the N298n control simulation contained 51 hydrogen bonds given the 3 Å cut-off.

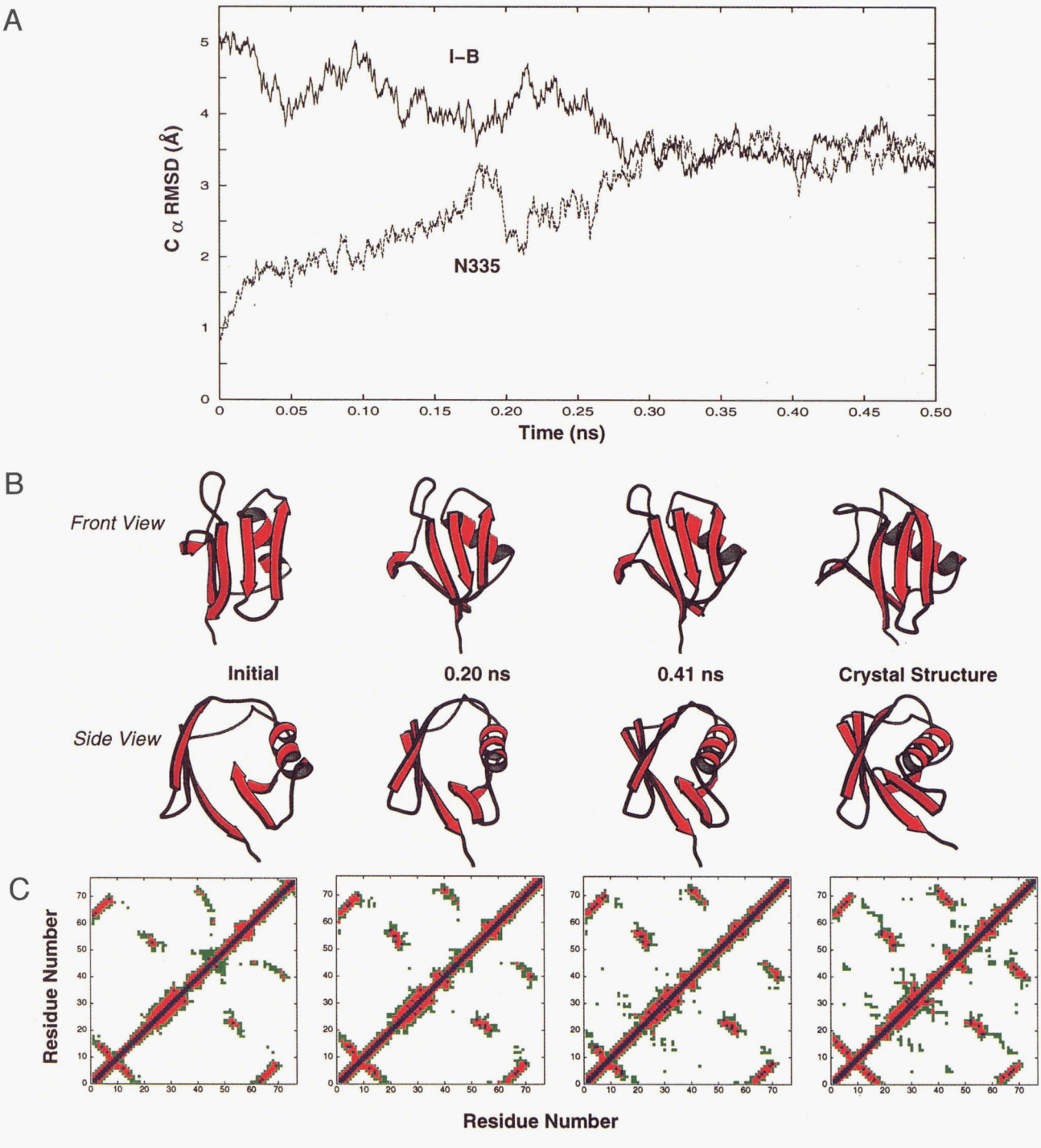

Fig. 4. Properties of the I-B simulation. **A:** The C_α RMSD from the crystal structure of the I-B and N335 simulations as a function of time. **B:** Main chain traces depicting some of the structural changes occurring during the simulation are displayed. The structures from all time points are shown in two orientations: the top is the standard orientation, as given in Figure 1, and the bottom is a side view with the $\beta 2$ strand in front on the left. The positions of the native secondary structural regions along the sequence are depicted as a red ribbon. **C:** α-Carbon contact maps for the structures shown in B. The contact maps are colored according to the following C_α-C_α contact distances: blue, $d \leq 5$ Å; red, $5 < d \leq 7.5$ Å; and green, $7.5 < d \leq 10$ Å.

ever; I-C collapsed to the radius of gyration of the N335 simulation in ~0.5 ns, while it took I-D ~1.5 ns (Fig. 5A). The α-carbon RMSD from the crystal structure for these simulations also decreased (Table 1). In the third simulation, I-E, the protein collapsed, but to a lesser degree and oscillated with time (Table 1).

The observed collapse was solvent dependent: structures collapsed in water and did NOT collapse in 60% methanol (Table 1 and compare the I-C and I-C_M curves in Fig. 5A). This effect is also reflected in the solvent accessible surface area (Fig. 5B). In water, the solvent exposure of both the hydrophobic and polar

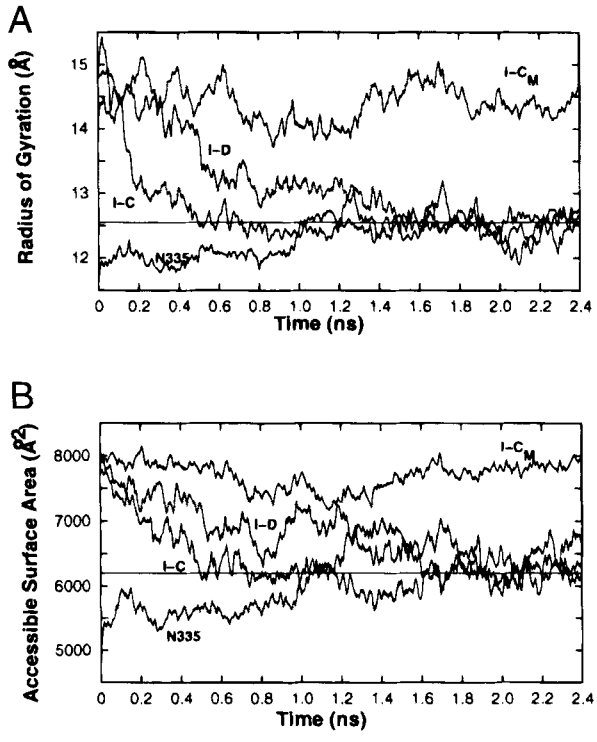

Fig. 5. The radius of gyration and the solvent accessible surface area as a function of time for partially unfolded structures (I-C and I-D) in pure water or 60% MeOH (subscript M) compared to starting from the crystal structure (N335). The values are plotted as running averages over 1 ps. **A:** The radius of gyration was calculated using the positions of the α-carbons: $R_g = 1/n\Sigma(\mathbf{r}_i - \langle\mathbf{r}\rangle)^2$, where $\mathbf{r}_i$ represents the position vector of each α-carbon and $\langle\mathbf{r}\rangle$ is the average position of those carbons, $1/n\Sigma(\mathbf{r}_i)$. **B:** The solvent accessible surface area was calculated using the method of Lee and Richards (1971) with a 1.4 Å probe.

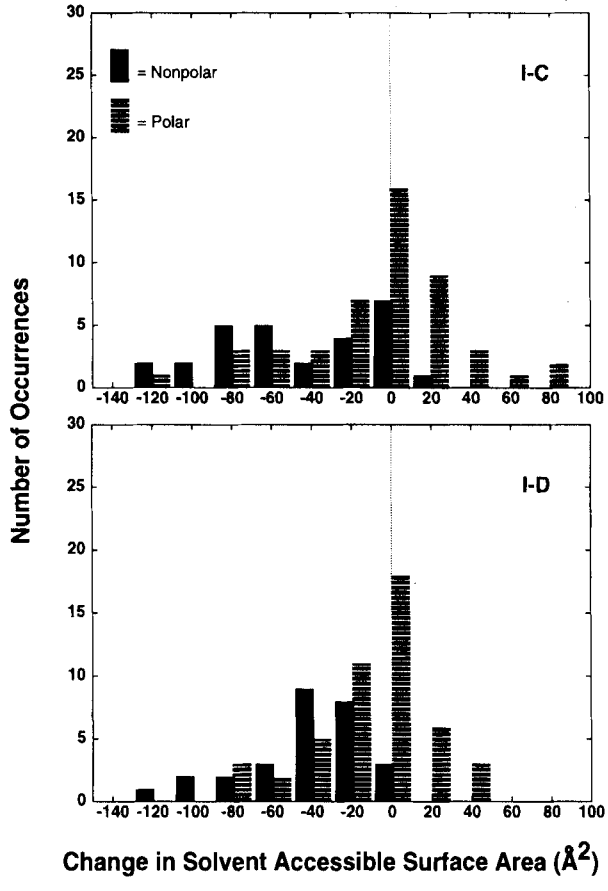

Fig. 6. Histograms of changes in solvent accessible surface area. Hydrophobic (dark bars) and polar residues (light bars) are plotted for the I-C and I-D simulations. The change was calculated between the values for the initial structure and the average over the last 0.1 ns of the simulations. The bin sizes and limits were the same for polar and hydrophobic residues and are only offset on the graph for the sake of clarity. The bins spanned ± 10 Å^2 around the following centers: -140, -120, -100, ..., 0, $+20$, $+40$, $+60$, $+80$ Å^2.

amino acids decreased over the course of the simulations, but the hydrophobic residues lost on average eight times as much surface area per residue as the polar groups (Fig. 6; Table 2). The change in accessibility ranged from -120 to $+90$ Å^2/residue for the I-C simulation. Of the 48 polar residues, approximately half gained in solvent accessible surface area while the other half lost. The largest single group of polar residues changed by less than 10 Å^2/residue with an average loss of $\sim$3 Å^2. In contrast, almost all of the hydrophobic residues lost significant solvent accessible surface area, and 60% lost more than 30 Å^2/residue, with an average loss of 44 Å^2/residue. The hydrophobic residues that changed the least or actually gained in exposure are on the surface of the protein in the crystal structure (Phe 4, Val 17, Phe 45, Ala 46, Val 70, and Leu 73).

Both the collapse and burial of the hydrophobic residues is made evident by comparing structures from corresponding simulations performed in methanol and water (Fig. 7). The simulations began with an expanded structure with many hydrophobic residues exposed to solvent (labeled as I-C: Initial in Fig. 7). The overall solvent exposure of I-C decreased slightly in methanol (-140 Å^2), but a much greater collapse was observed in water ($-1{,}500$ Å^2) and the hydrophobic residues became preferentially sequestered in the protein interior. Similarly, I-D collapsed with a distinct clustering of nonpolar residues (Fig. 7); however, the preferential burial of the nonpolar residues was slightly less effective than for I-C (Fig. 6; Table 2).

The energetics of the system are determined by the balance of protein-protein, protein-water and water-water interactions. The number of water-protein interactions decreased significantly over the course of the I-C simulation (Fig. 8A), and the decrease paralleled the rapid decrease in the radius of gyration (Fig. 5A). The number of main chain intra-protein hydrogen bonds increased in number by $\sim$10 (Fig. 8B), but this increase occurred slowly after the protein collapsed, that is between approximately 1.2 and 2 ns. As the protein collapsed, the water formerly in contact with the protein was free to interact with bulk water, such that the bulk water-water interactions increased (Fig. 8A). The number of protein-protein tertiary contacts increased steadily over the first 0.6 ns of the simulation (Fig. 8B) and did not change as abruptly as the radius of gyration or the interactions involving water.

Specific structure acquired during collapse

Various properties of the I-C and I-D collapse simulations averaged over the entire protein or classes of residues have been presented above, but we now consider the specific structure acquired during the simulations. The following questions are ad-

Table 2. *Changes in solvent exposed surface area*

Simulation[a]	Nonpolar/polar ratio[b]	Nonpolar residues		Polar residues	
		Initial[c]	Final[d]	Initial[c]	Final[d]
N335	—	981	1,586	3,797	4,627
I-C	7.7	2,700	1,476	5,007	4,733
I-D	3.7	2,774	1,601	5,149	4,605
$I\text{-}G_1$	1.6	2,530	2,025	5,190	4,637
$I\text{-}G_2$	2.1	2,530	1,828	5,190	4,605
$I\text{-}G_v$	0.3	2,530	2,092	5,074	3,020

[a]The simulations are described in Table 1. $I\text{-}G_1$: The first collapsed state of the I-G simulation, 0.85–0.95 ns. $I\text{-}G_2$: The second collapsed state of the I-G simulation, 1.9–2.0 ns. $I\text{-}G_v$ is the last 100 ps of an I-G simulation in vacuo, 1.1–1.2 ns.

[b]Ratio: per residue loss in solvent accessible surface area for nonpolar residues versus polar residues = $(\langle \Delta A_{non}/n_{non} \rangle)/(\langle \Delta A_{polar}/n_{polar} \rangle)$, where ΔA is the change in solvent accessible surface area and n is the number of residues. The change is between the starting structure of the simulation and the average over the last 0.01 ns, unless specified otherwise.

[c]Initial: Solvent exposed surface area of the named class of residues in the initial structure.

[d]Final: The average total solvent exposed surface area of the named class of residues over the last 0.01 ns of a given simulation.

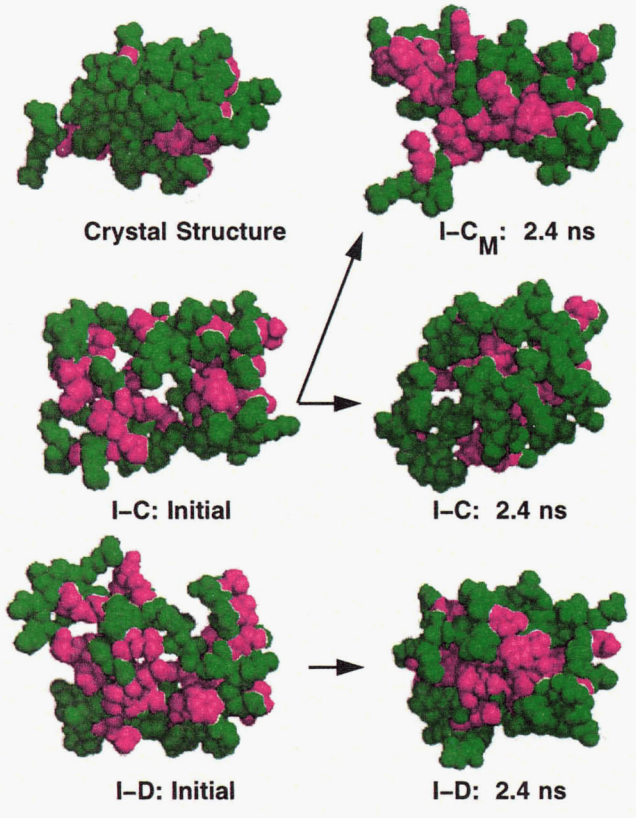

Fig. 7. Space filling models for the crystal structure and the starting and final structures from the I-C, $I\text{-}C_M$, and I-D simulations. Polar residues are shaded green and hydrophobic residues are magenta.

dressed below: What residual native structure remains in the starting structures for these simulations, what structure is formed during the collapse process, and what differences remain between the final conformations and the crystal structure?

Figure 9 shows main chain traces of the protein at various points in time during the I-C and I-D simulations, illustrating the large scale conformational changes that occurred. The main chain folding patterns of the starting structures were distinct from the crystal structure, and several distances between main chain atoms were 20 Å greater than in the crystal structure. However, both starting structures had a recognizable α-helix, and portions of β strands 1 and 2 in proximity. During the simulations, the nonlocal intrachain distances decreased dramatically (Fig. 9). For example, to provide perspective to Figure 9, the distance between $\beta 1$ and $\beta 5$ in I-D dropped by 10 Å during the first nanosecond.

The scarcity of native contacts in the starting structures and the increase in tertiary contacts over the course of the simulation are shown in the contact maps in Figure 10. The points are shaded according to the proximity of the α-carbons of the residues indicated on the axis (see caption). Figure 10 also relates the patterns appearing in the contact maps of the partially unfolded conformations with the corresponding contacts in the crystal structure.

The starting structure for the I-C simulation (Fig. 10) had few tertiary contacts compared to the crystal structure, and only $\beta(1\text{–}2)$ (region 1) and the α-helix (region 5) had near-native secondary structure contact patterns. The α-helix was intact as judged by the widening of the main diagonal for residues 23–33, but the distances between $i \rightarrow i + 4$ α-carbons were longer than in the crystal structure. The α-carbon distances between the adjacent strands in $\beta(1\text{–}2)$ were also longer, but they displayed a regular antiparallel pattern. Other β-structure was more disrupted, lacking most of the contacts between adjacent strands.

The density of contacts in the 1 ns structure of the I-C simulation was much greater than that of the starting structure (Fig. 10). Most of the contacts formed during the I-C simulation were tertiary contacts, both native and nonnative. The native contacts in region 6 are especially noticeable. Nonnative contacts in region 8 also formed as a result of a kink in the $\beta(1\text{–}2)$ hairpin and nonnative docking between $\beta(1\text{–}2)$ and the helix. Native contacts were restored between $\beta(1\text{–}5)$ during the simulation, such that the parallel β-sheet pattern involved more residues than the starting structure (region 2 of Fig. 10 and see Fig. 9). However, the strands were not as long, nor as close, as in the native state. Both main chain and side chain interactions between Met 1, Gln 2 and Gln 62, Lys 63 and Glu 64 position the two ends of the protein together. Although residues 62 and 63 fall outside of $\beta 5$, they were important in bringing $\beta 5$ into proximity of $\beta 1$. These polar interactions provided the specificity, or alignment, such that the proper hydrophobic contacts could then be formed between the side chains of Ile 3, Val 5, Leu 8 and Leu 67, Leu 69, and Leu 71. Contacts also formed between $\beta(3\text{–}5)$, but these were nonnative. The helix tightened up during the simulation, and two of the distances between $i \rightarrow i + 4$ α-carbon atoms became <5 Å.

Contacts between $\beta(1\text{–}2)$, $\beta(1\text{–}5)$, and $\beta(3\text{–}4)$ were much more disrupted in I-D than in I-C (compare the two initial structures in Figs. 9 and 10). The starting structure for I-D did not contain a regular antiparallel β-sheet pattern in region 1, representing $\beta(1\text{–}2)$. Even the relatively distant contact between residues 1 and 14 (6 Å) was nonnative. Similarly, the α-carbon pairs between strands 1 and 5 (region 2) were nonnative and farther than 7 Å apart. The prominent contacts between residue 7 (and flanking residues) and those near residue 26 were nonnative.

During the I-D simulation, the number of contacts in regions 2, 3, and, 6 increased appreciably compared to the starting structure

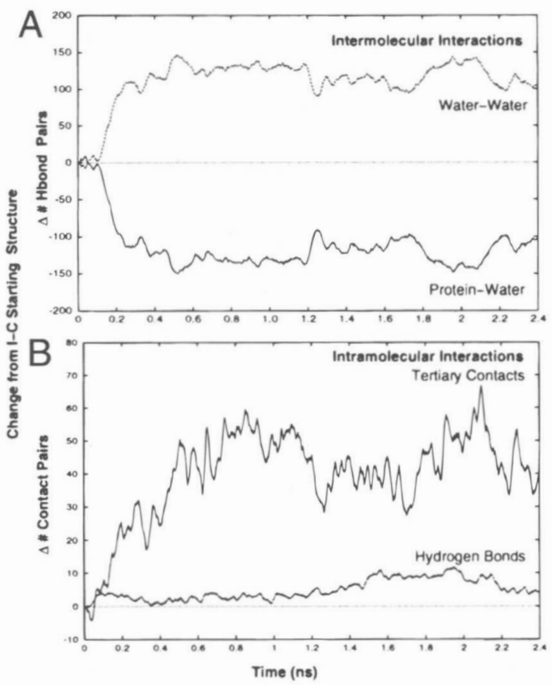

Fig. 8. Intermolecular and intramolecular interactions during the simulation of I-C in water. **A:** Intermolecular interactions were counted as the change in the number of contacts between water molecules or between the protein and water. **B:** The change in tertiary contacts (the number of protein atoms within 4.5 Å of other protein atoms separated by two or more residues) and protein hydrogen bonds (3 Å cut off).

(Fig. 10). In regions 2 and 3, the distances between native contact pairs in β(1–5) and in β(3–5) became shorter, and the number of contacts increased. For example, β(3–5) (region 3) increased slightly in length (Fig. 10), and the side chain packing increased dramatically (Fig. 11), but while the distances between the relevant α-carbons in β(3–5) decreased, most of the side- and main-chain contacts were nonnative. The extent of motion giving rise to the increased interactions is exemplified by the prominent nonnative contact between residues 22 and 62 in region 6 of Figure 10, which resulted from a 10 Å movement of the main chain. Although the packing interactions improved between the β-strands (Figs. 9–11), the interactions were primarily through side chains. However, in region 2 [β(1–5)] one inter-strand hydrogen bond formed, and in region 3 [β(4–5)] two formed. Interestingly, many of the side chain interactions were nonnative, involving small, dynamic hydrophobic clusters (for example, see Fig. 11).

Unfolded starting structures

Four simulations beginning from very unfolded structures (I-F, I-G, I-H, and I-I with RMSD = 10–12 Å, Table 1) were investigated. In the discussion that follows, the I-G simulation is used as the representative of this group. Like the other simulations presented above, I-G collapsed to a compact state by ~0.9 ns, but unlike those cases, it did not remain collapsed (Fig. 12). After the initial collapse, the radius of gyration increased rapidly and then dropped again by 2 ns. Even in its collapsed state, the protein was not as compact as in the other simulations (Table 1; compare Figs. 5, 12).

The I-G starting conformation did not contain β-structure (Fig. 12B,C). Only the nonnative contacts near residues 2 and 59 appeared in any of the β-sheet regions of the contact map, and these residues are 16 Å apart in the crystal structure. In contrast to I-C and I-D, there were some tertiary contacts not associated with regular secondary structure in regions 6 and 7 (see Fig. 10). The contacts in region 6 were shifted in register compared to the native

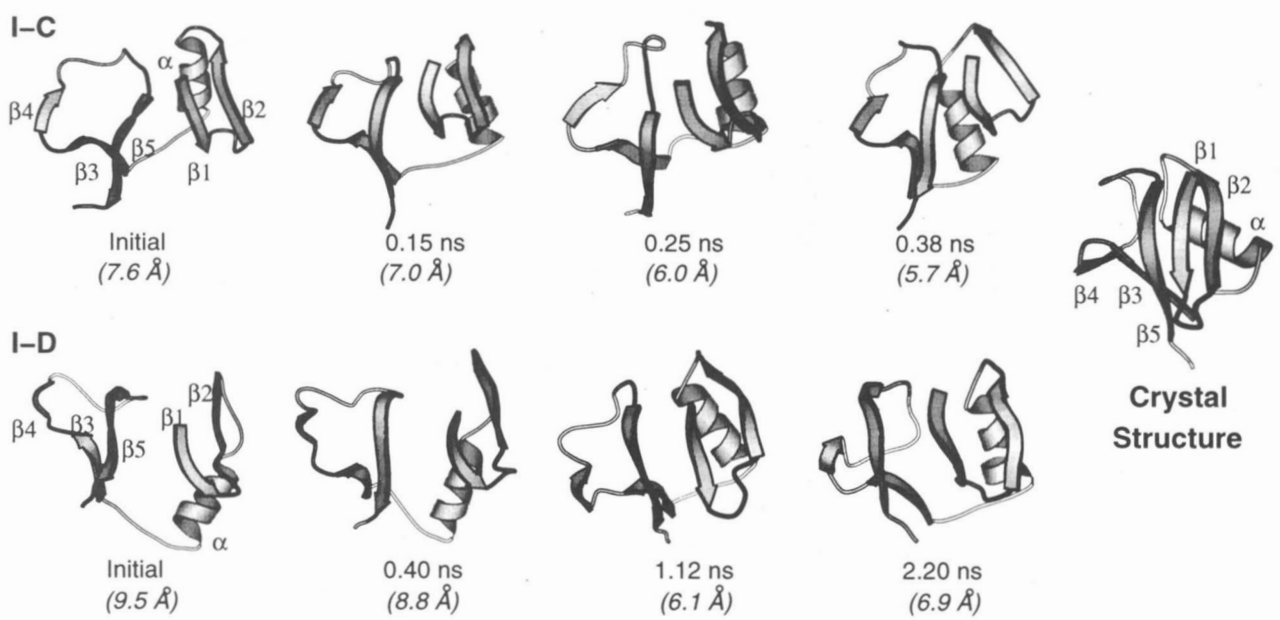

Fig. 9. Main chain traces of the structures from the I-C and I-D simulations at different time points. The crystal structure is shown for comparison. Native secondary structure regions are shaded for reference, but do not indicate that the structure adopted was necessarily native. The C_α RMSD to the crystal structure is given in parentheses for each structure.

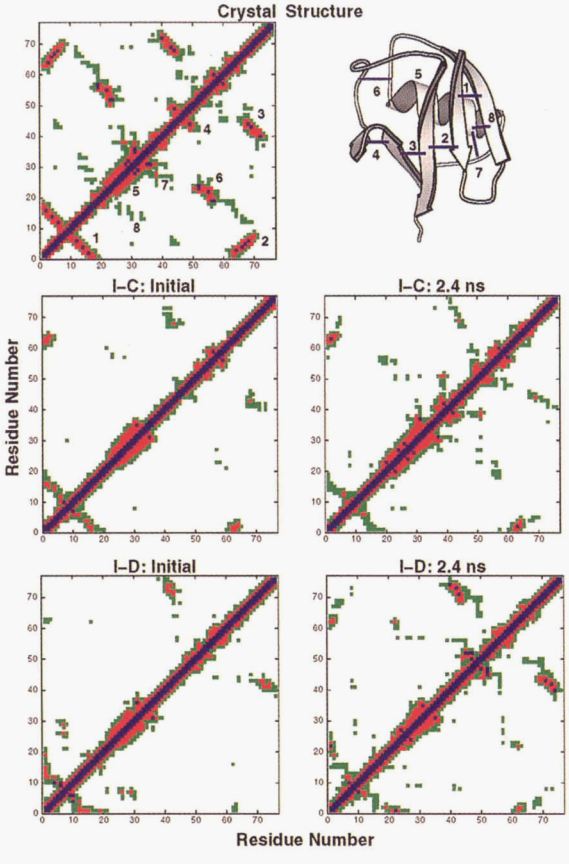

Fig. 10. α-Carbon contact maps for the initial and final structures from the I-C and I-D simulations. Secondary structure and tertiary contacts are numbered in the contact map and main chain trace for the crystal structure. The secondary structure is shaded, and residues 73–76 are not shown. The regions are numbered as follows: region 1, β(1–2); region 2, β(1–5); region 3, β(3–5); region 4, β(3–4) hairpin; region 5, α-helix; region 6, N-terminus of the helix (residues 19–24) to residues 52–56; region 7, the turn after the C-terminus of helix back onto the helix; and region 8, the docking between β-2 and the helix. The contact maps are colored as follows: blue, $d \leq 5$ Å; red, $5 < d \leq 7.5$ Å; and green, $7.5 < d \leq 10$ Å.

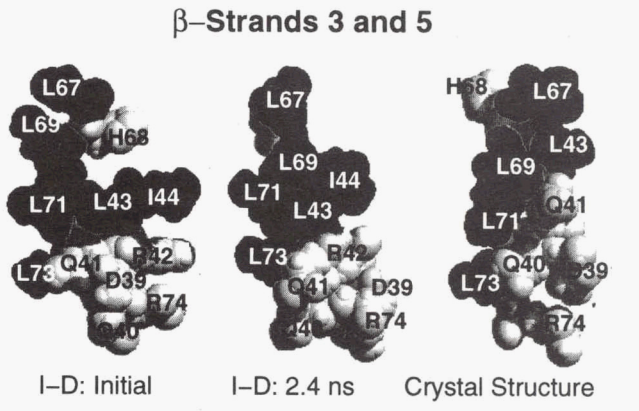

Fig. 11. Space filling models for portions of β-strands 3–5 for I-D and the crystal structure. These representations show all atoms of residues 40 through 44 in β3 and residues 67–73 in β5. The hydrophobic groups are black and polar residues are white/gray.

contacts. For example, residues 19 and 52 were in contact in the starting structure of I-G, instead of 19 and 57 as in the crystal structure.

The collapsed 0.9 ns structure from the I-G simulation contained many more contacts than the starting structure (Fig. 12); however, only in region 6 were these native contacts. Interestingly, most of the contacts formed between pairs in which at least one partner was in the latter half of the molecule (residue numbers >35, as illustrated in Fig. 12C). The N-terminal half made no close contacts with other N-terminal residues, but the C-terminal half folded back on itself and also made some contacts with residues near the N-terminus. In contrast, the C-terminus is mobile in the native state; that is, in the snapshots from I-G, the tail observed is at the N-terminus not at the C-terminus (note the different color of the tail at the bottom of the molecule for I-G and the crystal structure in Fig. 12, where the protein is colored from red at the N-terminus to blue at the C-terminus). Contacts formed between residues in region 2, β(1–5) and near region 8, β(3–5), but these were almost all nonnative. Thus, this early collapsed structure of I-G was very nonnative. The increase in the radius of gyration around 1 ns resulted from a general expansion and overall elongation of the molecule (see the 1.05 ns structure in Fig. 12B). Consequently, the numbers of contacts decreased (see region 6 in Fig. 12).

The protein re-collapsed between 1.1 and 2 ns, and different contacts formed during the second collapse (Fig. 12C). For instance, a nonnative contact between residues 19 and 46 formed, which was completely absent in the expanded 1.0 ns structure and shifted in the 0.9 ns structure (residues 19 and 55). Contacts began to form between residues in β(1–2), but they were nonnative. These contacts resulted from the formation of a small, and poorly packed, hydrophobic cluster of the side chains of Phe 4, Val 5, Leu 8, and Ile 13, shown as main chain contacts between residues 3 and 11 in Figure 12C.

The solvent exposed surface area of the hydrophobic residues decreased as the I-G simulation proceeded. Hydrophobic residues in both collapsed states were preferentially buried relative to the starting structure, and this preferential burial was more pronounced in the second collapse. In the first collapse, hydrophobic residues lost 1.6 times as much solvent-exposed surface area per residue as did the polar residues (relative to the initial structure) (Table 2). In the second collapse, that ratio increased to 2.1, as the solvent exposed surface area of the hydrophobic residues further decreased by another 180 Å^2. The extent of hydrophobic burial remained incomplete: the ratio of hydrophobic to polar burial was ~8 in the I-C simulation (Table 2). The I-H and I-I simulations showed the same general behavior, which is notable as these structures were completely devoid of native secondary structure (Fig. 2). A simulation of I-G *in vacuo* also collapsed (data not presented); however, *in vacuo* there was no preferential loss of hydrophobic surface area, rather there was a disproportionate loss of *polar* surface area (Table 2). Collapse is a common occurrence *in vacuo* where there is no solvent to compete with favorable intra-protein interactions. While such a collapse is not relevant to biological systems, it demonstrates that water was necessary for the preferential burial of hydrophobic residues.

Discussion

Molecular dynamics simulations of nine nonnative conformations (labeled I-A through I-I) of ubiquitin generated from two thermal

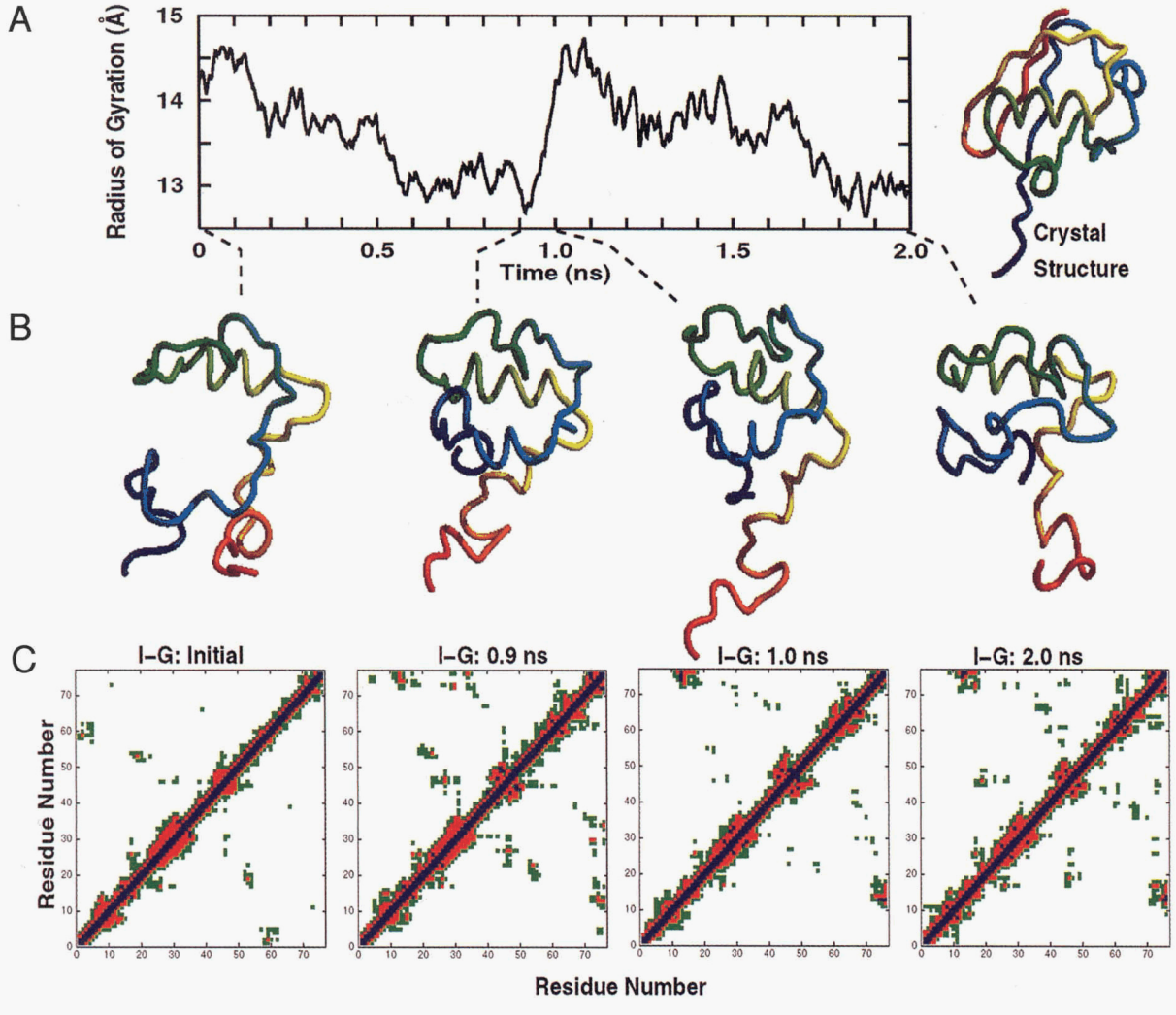

Fig. 12. Properties of the I-G simulation. **A:** The radius of gyration as a function of time, calculated as described in the legend to Figure 5. **B:** Main chain traces sampled from the I-G simulation at key points colored from red at the N-terminus through blue at the C-terminus. **C:** α-Carbon contact maps of the structures shown in B, using the same shading as described for Figure 10.

denaturation simulations have been performed under quasi-native conditions. These thermally-quenched simulations fall into three general classes, depending on the similarity of the starting structures to the crystal structure. (1) *Native-like* starting structures (RMSD $\leq$ 5 Å) approached the crystal structure: they gained native hydrogen bonds, they made native main chain contacts, they acquired native secondary and tertiary structure, their RMSD to the crystal structure decreased, and they became more native-like as evaluated by other independent criteria [for example, Procheck (Laskowski et al., 1993)]. The differences between the structures obtained in these simulations and the crystal structure were minor and appear to be due to the elevated temperature and low pH, such that I-A, I-B, and N335 are representative of the native state under these conditions. We also note that the control simulation at 298 K and neutral pH (N298n) exhibits a low average RMSD (1.8 Å after 2.4 ns). These results demonstrate that the simulations show the proper temperature-dependent effects and that the force field can realistically model the native state. (2) Starting structures *intermediate* between native and completely unfolded states (RMSD 5–10 Å) collapsed and became more native-like by a variety of criteria, but they did not refold. The radius of gyration and solvent accessible surface area in these simulations converged with the values obtained in simulations starting from the crystal structure at the same temperature. Furthermore, the decrease in accessible surface area was due to the preferential burial of hydrophobic residues. Some native tertiary contacts formed, primarily through side chains, but they did not, by and large, form the regular main chain contact patterns characteristic of native secondary structure. In contrast, collapse did not occur in control simulations in 60% methanol starting from the same structures. (3) The least native-like, *unfolded* starting structure (RMSD >10 Å) sampled collapsed states but experienced large fluctuations in radius of gyration throughout the simulation. Even in the collapsed states, the protein did not become as compact as in simulations starting from more native-like structures (the radii of gyration were 2–7% greater). Interestingly, more hydrophobic surface area was buried with successive cycles of collapse, but the protein did not refold nor even become more native-like in terms of specific secondary or tertiary structure. Unfortunately, while the unfolding reactions are continuous from native through fully unfolded states, the quenched sim-

ulations merely provide glimpses into possible early (nonspecific hydrophobic collapse and isolated units of unstable secondary structure), middle (more specific hydrophobic collapse and formation of secondary structure contacts), and late (consolidation of packing interactions and formation of secondary structural hydrogen bonds) folding events for ubiquitin. However, we caution that we cannot establish that the structures are on a folding pathway, which is similar to the situation encountered by experimentalists in establishing whether intermediates are on- or off-pathway (see Baldwin, 1996; Roder & Colón, 1997).

It has been difficult, in general, to decipher the exact order of events in folding because both collapse and secondary structure formation tend to occur in the experimental deadtime (typically the millisecond timescale). There is mounting evidence that collapse precedes or occurs concomitant with secondary structure formation in many small single domain proteins (Chaffotte et al., 1992; Khorasanizadeh et al., 1993, 1996; Mann & Matthews, 1993; Fersht et al., 1994; Feng & Widom, 1994; Itzhaki et al., 1994, 1995; Jones et al., 1994, 1995; Otzen et al., 1994; Sosnick et al., 1994; Viguera et al., 1994; Agashe et al., 1995; Eliezer et al., 1995; Huang & Oas, 1995; Jones & Matthews, 1995; Nölting et al., 1995, 1997; Schindler et al., 1995; Ballew et al., 1996; Houry et al., 1996; Lopez-Hernandez & Serrano, 1996; Pascher et al., 1996; Schindler & Schmid, 1996; Yi & Baker, 1996; and reviewed by Evans & Radford, 1994; Miranker & Dobson, 1996; Fersht, 1997). In the case of barstar, Agashe et al. (1995) conclude that "the earliest event of the major folding pathway of barstar is a nonspecific hydrophobic collapse that does not involve concomitant secondary structure formation." Similar studies of the 101 residue C-terminal fragment (F2-V8) of tryptophan synthase led Chaffotte et al. (1992) to conclude that: ". . . the condensed nonnative state of F2-V8 results from a very rapid nonspecific hydrophobic collapse." Furthermore, Jones et al. (1995) have reported that dihydrofolate reductase appears to collapse and then re-expand during folding. A preliminary account of the fast phase (μs) of folding of barstar by Nölting et al. (1995) suggests that the earliest steps in folding involve a transition between two "relatively solvent exposed states with little consolidation of structure." More recent studies by Nölting et al. (1997) confirm the earlier studies; however, they find, contrary to Agashe et al., that the collapsed state contains some secondary structure. The detection/formation of secondary structure appears to be dependent on the solvent: secondary structure is seen in the absence of denaturant but is not observed in 2 M urea (Nölting et al., 1997) or 1 M GndHCl (Agashe et al., 1995). In any case, the experimental evidence suggests that different proteins may undergo collapse and secondary structure formation at different relative points during folding (and see Fersht, 1997), and for ubiquitin the experimental results support the existence of a rapidly formed, collapsed intermediate.

The folding of ubiquitin has been studied extensively by Roder and co-workers (Briggs & Roder, 1992; Khorasanizadeh et al., 1993, 1996) and reviewed by Baldwin (1996) and Roder and Colón (1997). They conclude that a three state mechanism best describes the kinetics for the folding of ubiquitin. Khorasanizadeh et al. (1996) report the formation of an on-pathway collapsed intermediate within the 2–3 ms dead time of the experiments, which they describe as a "loosely folded state held together mainly by hydrophobic contacts" (also see Roder & Colón, 1997). This state affords little protection from exchange of labile amide hydrogens, indicating that it "still lacks stable, persistent hydrogen-bonded structure," but it appears to contain secondary structure (Khorasanizadeh et al., 1993, 1996). These experimental observations and interpretations are consistent with our "intermediate" (I-C and I-D) states, which have a dynamic and loose hydrophobic core and some secondary structure, although their dynamic nature leads to fluctuations in solvent exposure of the amide hydrogens and only transient formation of hydrogen bonds. Khorasanizadeh et al. (1996) also note that residue 45 appears to be buried in the intermediate (residue 45 is Tyr in wildtype and Trp in their mutant), giving 90% of the fluorescence of the native state. Tyr 45 was partially buried in all of our collapsed states, but the packing and sequestration from solvent was pronounced in the more native-like, intermediate structures (data not presented).

Since hydrophobic collapse of ubiquitin occurs during the experimental deadtime (ms), only an upper limit to the actual time for collapse can be determined. Collapse or fluctuations in the radius of gyration took place in the simulation on the 0.5 to 2 ns time scale, although we do not know whether the rates observed in simulations will map exactly to experimental times. The rapidity of the collapse was a result of several factors. First, our simulations were conducted at elevated temperature (62 °C), with a corresponding decrease in the solvent density. Furthermore, simple collapse to any one of many collapsed states is an easier search problem than a protein folding to one native and precisely-packed structure. Also, as shorter time scales become experimentally accessible through fast spectroscopic techniques (Jones et al., 1993; Ballew et al., 1996; Williams et al., 1996; and reviewed by Plaxco & Dobson, 1996; Wolynes et al., 1996; Eaton et al., 1997), it is becoming clear that both secondary and tertiary structure can form on the nanosecond-microsecond time scale even at room temperature, or lower. Thirdly, ubiquitin is small and contains a well-defined hydrophobic core. It is very stable against denaturation by heat, pH, and chemical denaturants (Lenkinsiki et al., 1977; Jensen et al., 1980; Harding et al., 1991), thus the hydrophobic residues may be situated along the sequence so as to make collapse facile. Fourth, the unfolded starting structures in these simulations were not generated randomly; in many cases the α-helix was partially intact and in some of the structures, incipient native β-structure was present. Finally, the radius of gyration of the unfolded structures fluctuated significantly, and most experimental probes will only measure the population average and cannot detect fluctuations of a single molecule on a nanosecond time scale. Also, it is worth noting that fast hydrophobic collapse is not necessarily a favorable event in the absence of a somewhat native chain topology or other specific interactions to guide the collapse and instead it may inhibit or compete with acquisition of native structure.

Because the very early stages of protein folding are so difficult to characterize using experimental means, one would hope that simulations could shed some light on this area. We find that although there was a general tendency for the very nonnative expanded structures to collapse temporarily to "misfolded" compact states on the nanosecond timescale, they appear to be only marginally stable and are able to re-expand and fluctuate rapidly with respect to compactness. The average behavior of a population of such structures would show a much slower decrease in radius of gyration than the time scale of the fluctuation of an individual structure. The general behavior displayed in the simulations regarding early, nonspecific hydrophobic collapse is consistent with experimental observations. However, due to the lack of definitive experimental data during the very early stages of folding and the discontinuous nature of our quenched simulations, we cannot distinguish between the structures being early on the folding pathway,

a slow folding phase, or being folding incompetent and off pathway. Experiments of kinetic partitioning between folding and misfolding suggest that the folding efficiency is ~50% in the cell (Jaenicke, 1995), and, as such, insights into off-pathway events are also crucial to understanding and controlling protein folding. In addition, we note that while experimental studies demonstrate that ubiquitin folds rapidly, a minor slow-folding population is observed (Khorasanizadeh et al., 1996).

For the simulations that experienced more specific hydrophobic collapse but did not refold (the intermediate structures, I-C and I-D), many of the side chain interactions that formed upon collapse were nonnative and involved small, dynamic hydrophobic clusters (for example, see the 2.4 ns collapsed structure in Fig. 11). Some of these clusters were in the inter-strand hydrogen bonding plane and prevented the "normal" arrangement of side chains above and below the plane of the main chain hydrogen bonds in β-sheets. Interestingly, this behavior was observed in water, but the side chains moved out of the plane and hydrogen bonds formed in 60% methanol (see Alonso & Daggett, 1995). Interactions of β-sheet residues through side chain packing in the absence of stable inter-strand hydrogen bonds has been observed in a small designed peptide (Sieber & Moe, 1996) and in the denatured state of staphylococcal nuclease (Wang & Shortle, 1996). In addition, a fragment comprised of the β(1–2) hairpin from ubiquitin shows unambiguous $C_{\alpha H}$-$C_{\alpha H}$ NOEs across the sheet, while the NH groups readily exchange with solvent (Searle et al., 1995). Such an effect has also been observed during the refolding of hen egg white lysozyme, where Itzhaki et al. (1994) found that collapse is substantial in the experimental deadtime, but they state that the side chain packing is loose and only unstable hydrophobic clusters appear to be present. Also, during this time some secondary structure forms, but the secondary structure is not stable enough to confer protection from exchange to the amide hydrogens. Itzhaki et al. (1994) suggest that later protection from exchange is primarily the result of the development of stabilizing side chain interactions, which are predominately hydrophobic in nature.

The nonnative hydrophobic clusters observed in the simulations were energetically favorable and contributed to the stability of I-C and I-D, such that they did not proceed further toward the native state. Overall, this state is slightly expanded compared to the native state with a partially formed but compromised hydrophobic core, disrupted secondary structure, moderate hydrogen bond content, and a mixture of native and nonnative packing contacts. In fact, these properties are similar to those of the transition state of unfolding/folding of chymotrypsin inhibitor 2, another small and fast folding protein (Otzen et al., 1994; Li & Daggett, 1994, 1996; Itzhaki et al., 1995; Daggett et al., 1996). That the simulated structures have properties consistent with the intermediate observed experimentally and "get stuck," such that they appear to be before but near the transition state of folding, is consistent with the suggestion of Khorasanizadeh et al. (1996) that the transition state for the I $\rightarrow$ N process for ubiquitin is structurally similar to I. The more native like structures, I-A and I-B, however, did not experience a significant barrier to acquiring native structure and would then appear to be after the major transition state. The final structures from these simulations are slightly deformed relative to the crystal structure, but no more so than the control simulation beginning with the crystal structure, and the packing is significantly improved compared to I-C and I-D. In addition, we note that while ubiquitin is unusually stable to acidic pH, there is an increase in the exposed hydrophobic surface area in the folded conformation at low pH at 25 °C (Khorasanizadeh et al., 1993). Thus, I-A, I-B, and N335 all appear to be representative of the "native" state at 62 °C and low pH.

Methods

Three general types of molecular dynamics simulations were performed for this work: (1) high temperature denaturation simulations (D1 and D2) starting with the crystal structure to generate starting structures for the quenched simulations; quenched simulations of nonnative structures (I-A through I-I) under quasi-native conditions; and control simulations of the crystal structure under native and nearly-native conditions (N298, N298n, N335). The simulation protocols have been described by Alonso and Daggett (1995), but the pertinent details of the quenched simulations follow.

The program ENCAD was used for all simulations (Levitt, 1990), using the potential function described by Levitt et al. (1995, 1997) in all cases. In all quenched simulations the charge states of ionizable residues were set to approximately correspond to a pH of 2.0 to match the experimental work of Harding et al. (1991). We protonated the C-terminal carboxyl group and the side chains on Asp, Glu, His, Lys, and Arg. The structures extracted from the high temperature denaturation simulations were minimized 2,000 steps in vacuo to minimize any possible artifacts caused by the high temperature generation of the structures. After minimization, water molecules were added around the protein to fill a rectangular box, with walls at least 8 Å away from any protein atom, resulting in a total of over 3,000 water molecules in all cases. The density of the solvent was set to the experimental value at 335 K (0.982 g/mL; Kell, 1967) by adjusting the volume of the box, after which the box volume was held constant. Periodic boundary conditions were used to minimize boundary effects. The resulting systems were then prepared for molecular dynamics. The entire system, including protein and solvent, was minimized for 3,000 steps, followed by 1,000 steps of dynamics, and finally minimized for an additional 2,000 steps. This procedure was followed in order to maintain consistency between the 60% MeOH and the pure water simulations. For one starting structure (I-C), the effects of the preparation strategy were investigated by constraining the protein and equilibrating the water for 10,000 steps of molecular dynamics. The results were the same as those involving less equilibration of the water ($\Delta R_g < 0.01$ Å and $\Delta\alpha$-carbon RMSD < 0.5 Å), and in both cases the protein was well-solvated (data not presented). In addition, the same procedure was used for simulations with a quench temperature of 298 K except that the density was adjusted to 0.997 g/mL. Lowering the temperature led to comparable results but the rate of collapse was slower (data not presented), so due to limited computer resources, simulations were performed at 335 K.

Full molecular dynamics simulation of the entire system was then started after the above preparations. Atoms in the system were assigned velocities according to a Maxwellian distribution and allowed to move according to Newton's equations of motion. The velocities of the atoms were adjusted intermittently until the system reached the desired temperature, which usually takes 2–3 ps. A 2 fs (1 fs = 10^{-15} s) time step was used for both the preparation and the full simulation. An 8 Å nonbonded cutoff was employed and the nonbonded list was updated every 5 time steps. Structures were saved every 0.2 ps for analysis.

Acknowledgments

We thank Drs. Sheena Radford, David Baker, Phil Evans, Peter Kollman, and Ken Dill for stimulating discussions and/or helpful comments on the manuscript. We are also grateful to Drs. Alan R. Fersht and Bryan Jones for providing preprints of their work. We appreciate financial support from the National Science Foundation (MCB 9407903), and the Donors of The Petroleum Research Fund, administered by the American Chemical Society (ACS-PRF 28022-64), and primarily the Young Investigator Program of the Office of Naval Research (N00014-95-1-0484). Molecular graphics images were made using Molscript (Kraulis, 1991; Figs. 1, 9, 10) and UCSF MidasPlus (Ferrin et al., 1988; Figs. 2, 4, 7, 11, 12).

References

Agashe V, Shastry M, Udgaonkar J. 1995. Initial hydrophobic collapse in the folding of barstar. *Nature 377*:754–757.

Alonso DOV, Daggett V. 1995. Molecular dynamics simulations of protein unfolding and limited refolding: Characterization of partially unfolded states of ubiquitin in 60% methanol and in water. *J Mol Biol 247*:501–520.

Anfinsen CB. 1973. Principles that govern the folding of protein chains. *Science 181*:223–230.

Baker D, Agard DA. 1994. Kinetics versus thermodynamics in protein folding. *Biochemistry 33*:7505–7509.

Baldwin RL. 1989. How does protein folding get started? *Trend Biochem Sci 14*:291–294.

Baldwin RL. 1996. On-pathway versus off-pathway folding intermediates. *Folding & Design 1*:R1–R8.

Ballew RM, Sabelko J, Gruebele M. 1996. Direct observation of fast protein folding—The initial collapse of apomyoglobin. *Proc Natl Acad Sci USA 93*:5759–5764.

Bozcko EM, Brooks CL. 1995. First-principles calculation of the folding free energy of a three-helix bundle protein. *Science 269*:393–396.

Bond CJ, Wong K, Clarke J, Fersht AR, Daggett V. 1997. Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: Description of the folding pathway. *Proc Natl Acad Sci USA 94*:13409–13413.

Briggs MS, Roder H. 1992. Early hydrogen-bonding events in the folding reaction of ubiquitin. *Proc Natl Acad Sci USA 89*:2017–2021.

Bryngelson JD, Onuchic JN, Socci ND, Wolynes G. 1995. Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins Struct Funct Genet 21*:167–195.

Caflisch A, Karplus M. 1994. Molecular dynamics simulation of protein denaturation: Solvation of the hydrophobic cores and secondary structure of barnase. *Proc Natl Acad Sci USA 91*:1746–1750.

Caflisch A, Karplus M. 1995. Acid and thermal denaturation of barnase investigated by molecular dynamics simulations. *J Mol Biol 252*:672–708.

Camacho CJ, Thirumalai D. 1993. Kinetics and thermodynamics of folding in model proteins. *Proc Natl Acad Sci USA 90*:6369–6372.

Chaffotte A, Cadieux C, Guillou Y, Goldberg M. 1992. A possible initial folding intermediate: The C-terminal proteolytic domain of Trp synthase B chains folds less than 4 ms into a condensed state with nonnative-like secondary structure. *Biochemistry 31*:4303–4308.

Covell DG. 1994. Lattice model simulations of polypeptide chain folding. *J Mol Biol 235*:1032–1043.

Covell DG, Jernigan RL. 1990. Conformations of folded proteins in restricted spaces. *Biochemistry 29*:3287–3294.

Daggett V. 1993. A model of the molten globule state of CTF generated using molecular dynamics. In: Angeletti RH, ed. *Techniques in protein chemistry, IV.* New York: Academic Press. pp 525–532.

Daggett V, Levitt M. 1992. A model of the molten globule state from molecular dynamics simulations. *Proc Natl Acad Sci USA 89*:5142–5146.

Daggett V, Levitt M. 1993. Protein unfolding pathways explored through molecular dynamics simulations. *J Mol Biol 232*:600–619.

Daggett V, Li A, Laura S, Itzhaki LS, Otzen DE, Fersht AR. 1996. Structure of the transition state for folding of a protein derived from experiment and simulation. *J Mol Biol 257*:430–440.

Dill KA. 1990. Dominant forces in protein folding. *Biochemistry 29*:7133–7185.

Dill KA, Bromberg S, Yue KZ, Feibig K, Yee D, Thomas P, Chan HS. 1995. Principals of protein folding: A perspective from simple exact models. *Protein Sci 4*:561–602.

Eaton WA, Muñoz V, Thompson PA, Chan C, Hofrichter J. 1997. Submillisecond kinetics of protein folding. *Curr Opin Struct Biol 7*:10–14.

Eliezer D, Jennings P, Wright P, Doniach S, Hodgson K, Tsuruta H. 1995. The radius of gyration of an apomyoglobin folding intermediate. *Science 270*:447–448.

Engelhard M, Evans PA. 1996. Experimental investigation of sidechain interactions in early folding intermediates. *Folding & Design 1*:R31–R37.

Evans PA, Radford SE. 1994. Probing the structure of folding intermediates. *Curr Opin Struct Biol 4*:100–106.

Feng H-P, Widom J. 1994. Kinetics of compaction during lysozyme refolding studied by continuous-flow quasielastic light scattering. *Biochemistry 33*: 13382–13390.

Ferrin TE, Huang CC, Jarvis LE, Langridge R. 1988. The Midas display system. *J Mol Graph 6*:13–27.

Fersht AR. 1997. Nucleation mechanisms in protein folding. *Curr Opin Struct Biol 7*:3–9.

Fersht AR, Itzhaki LS, elMasry NF, Matthews JM, Otzen, DE. 1994. Single versus parallel pathways of protein folding and fractional formation of structure in the transition state. *Proc Natl Acad Sci USA 91*:10426–10429.

Guo Z, Brooks CL, Boczko EM. 1997. Exploring the folding free energy surface of a three-helix bundle protein. *Proc Natl Acad Sci USA 94*:10161–10166.

Guo ZY, Thirumalai D, Honeycutt JD. 1992. Folding kinetics of proteins: A model study. *J Chem Phys 97*:525–535.

Hao MH, Pincus MR, Rackovsky S, Scheraga HA. 1993. Unfolding and refolding of the native structure of bovine pancreatic trypsin inhibitor studied by computer simulations. *Biochemistry 32*:9614–9631.

Harding MH, Williams DH, Woolfson DN. 1991. Characterization of a partially denatured state of a protein by two-dimensional NMR: Reduction of the hydrophobic interactions in ubiquitin. *Biochemistry 31*:3120–3128.

Hinds DA, Levitt M. 1992. A lattice model for protein structure prediction at low resolution. *Proc Natl Acad Sci USA 87*:2536–2540.

Hinds DA, Levitt M. 1994. Exploring conformational space with a simple lattice model for protein structure. *J Mol Biol 243*:668–682.

Houry W, Rothwarf D, Scheraga H. 1996. Circular dichroism evidence for the presence of burst-phase intermediates on the conformational folding pathway of ribonuclease A. *Biochemistry 35*:10125–10133.

Huang GS, Oas TG. 1995. Submillisecond folding of monomeric repressor. *Proc Natl Acad Sci USA 92*:6878–6882.

Hunenberger PH, Mark AE, Van Gunsteren WF. 1995. Computational approaches to study protein unfolding—Hen egg white lysozyme as a case study. *Proteins Struct Funct Genet 21*:196–213.

Itzhaki LS, Evans PA, Dobson CM, Radford SE. 1994. Tertiary interactions in the folding pathway of hen lysozyme: Kinetic studies using fluorescent probes. *Biochemistry 33*:5212–5220.

Itzhaki LS, Otzen DE, Fersht AR. 1995. Detailed structure of the transition state for folding of chymotrypsin inhibitor 2 analyzed by protein engineering. *J Mol Biol 254*:260–288.

Jaenicke R. 1995. Folding and association versus misfolding and aggregation of proteins. *Phil Trans Roy Soc Lond B Biol Sci 348*:97–105.

Jensen J, Goldstein G, Breslow E. 1980. Physical-chemical properties of ubiquitin. *Biochim Biophys Acta 624*:378–385.

Jones BE, Beechem JM, Matthews CR. 1995. Local and global dynamics during the folding of *Escherichia coli* dihydrofolate reductase by time-resolved fluorescence spectroscopy. *Biochemistry 34*:1867–1877.

Jones BE, Jennings PA, Pierre RA, Matthews CR. 1994. Development of nonpolar surfaces in the folding of *Escherichia coli* dihydrofolate reductase detected by 1-anilinonaphthalene-8-sulfonate binding. *Biochemistry 33*: 15250–15258.

Jones BE, Matthews CR. 1995. Early intermediates in the folding of dihydrofolate reductase from *Escherichia coli* detected by hydrogen exchange NMR. *Protein Sci 4*:167–177.

Jones CM, Henry ER, Hu Y, Chan CK, Luck SD, Bhuyan A, Roder H, Hofrichter J, Eaton WA. 1993. Fast events in protein folding initiated by nanosecond laser photolysis. *Proc Natl Acad Sci USA 90*:11860–11864.

Kabsch W. 1976. A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr Sect A 32*:922–933.

Karasawa T, Tabuchi K, Fumoto M, Yasukawa T. 1993. Development of simulation models for protein folding in a thermal annealing process-I: A simulation of BPTI folding by the pearl necklace model. *Comput Appl Biosci 9*:243–251.

Kazmirski S, Daggett V. 1998. Simulations of the structural and dynamic properties of denatured proteins: The "molten coil" state of bovine pancreatic trypsin inhibitor. *J Mol Biol 277(5)*, in press.

Kell GS. 1967. Precise representation of volume properties of water at one atmosphere. *J Chem Eng Data 12*:66–68.

Khorasanizadeh S, Peters ID, Butt TR, Roder H. 1993. Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry 32*:7054–7063.

Khorasanizadeh S, Peters ID, Roder H. 1996. Evidence for a three-state model of protein folding from kinetic analysis of ubiquitin variants with altered core residues. *Nat Struct Biol 3*:193–205

Kim PS, Baldwin RL. 1982. Specific intermediates in the folding reactions of small proteins. *Annu Rev Biochem 51*:459–489.

Kim PS, Baldwin RL. 1990. Intermediates in the folding reactions of small proteins. *Annu Rev Biochem 59*:631–660.

Kraulis PJ. 1991. MolScript: A program to produce both detailed and schematic plots of protein structure. *J Appl Cryst 24*:946–950.

Laidig KE, Daggett V. 1996. Molecular dynamics simulations of apocytochrome b_{562}: A "highly ordered" molten globule. *Folding & Design 1*:353–364.

Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Cryst 26*:283–291.

Lee B, Richards FM. 1971. The interpretation of protein structures: Estimation of static accessibility. *J Mol Biol 55*:379–400.

Lenkinsiki ER, Chen DM, Glickson JD, Goldstein G. 1977. Nuclear magnetic resonance studies of the denaturation of ubiquitin. *Biochim Biophys Acta 494*:126–130.

Levinthal C. 1969. How to fold graciously. Notes by Rawitch A. In: Debrunner P, Tsibris JCM, Munch E, eds. *Mossbauer spectroscopy in biological systems*. Montecello, IL: Allerton House. pp 22–24.

Levitt M. 1990. *ENCAD—Energy calculations and dynamics. Molecular Applications Group*. Stanford, CA: Stanford University and Yeda.

Levitt M, Hirshberg M, Sharon R, Daggett V. 1995. Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution. *Comp Phys Commun 91*:215–231.

Levitt M, Hirshberg M, Sharon R, Laidig KE, Daggett V. 1997. Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution. *J Phys Chem 101*:5051–5061.

Levitt M, Warshel A. 1975. Computer simulation of protein folding. *Nature 253*:694–698.

Li A, Daggett V. 1994. Characterization of the transition state of protein unfolding: Chymotrypsin inhibitor 2. *Proc Natl Acad Sci USA 91*:10430–10434.

Li A, Daggett V. 1996. Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 using molecular dynamics simulations. *J Mol Biol 257*:412–429.

Li A, Daggett V. 1998. Molecular dynamics simulation of the unfolding of barnase: Characterization of the major intermediate. *J Mol Biol 275*: 677–694.

Lopez-Hernandez E, Serrano L. 1996. Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2. *Folding & Design 1*:43–55.

Mann CJ, Matthews CR. 1993. Structure and stability of an early folding intermediate of *Escherichia coli* Trp aporepressor measured by far-UV stopped-flow circular dichroism and 8-anilino-1-naphthalene sulfonate binding. *Biochemistry 32*:5282–5290.

Mark AE, van Gunsteren WF. 1992. Simulation of the thermal denaturation of hen egg white lysozyme: Trapping the molten globule state. *Biochemistry 31*:7745–7748.

Matthews CR. 1993. Pathways of protein folding. *Annu Rev Biochem 62*:653–683.

Miranker AD, Dobson CM. 1996. Collapse and cooperativity in protein folding. *Curr Opin Struct Biol 6*:31–42.

Moult J, Unger R. 1991. An analysis of protein folding pathways. *Biochemistry 30*:3816–3824.

Nölting B, Golbik R, Fersht AR. 1995. Submillisecond events in protein folding. *Proc Natl Acad Sci USA 92*:10668–10672.

Nölting B, Golbik R, Neira JL, Soler-Gonzalez A, Schreiber G, Fersht AR. 1997. The folding pathway of a protein at high resolution from microseconds to seconds. *Proc Natl Acad Sci USA 94*:826–830.

O'Toole EM, Panagiotopoulos AZ. 1992. Monte Carlo simulation of folding transitions of simple model protein using a chain growth algorithm. *J Chem Phys 97*:8644–8651.

Otzen DE, Itzhaki LS, elMasry NF, Jackson SE, Fersht AR. 1994. Structure of the transition state for the folding/unfolding of the barley chymotrypsin inhibitor 2 and its implications for mechanisms of protein folding. *Proc Natl Acad Sci USA 91*:10422–10425.

Pascher T, Chesick JP, Winkler JR, Gray HB. 1996. Protein folding triggered by electron transfer. *Science 271*:1558–1560.

Plaxco K, Dobson CM. 1996. Time-resolved biophysical methods in the study of protein folding. *Curr Opin Struct Biol 6*:630–636.

Prat Gay GD, Ruiz-Sanz J, Davis B, Fersht AR. 1994. The structure of the transition state for the association of two fragments of the barley chymotrypsin inhibitor 2 to generate native-like protein: implications for mechanisms of protein folding. *Proc Natl Acad Sci USA 91*:10943–10946.

Ptitsyn OB. 1987. Protein folding: Hypotheses and experiments. *J Protein Chem 6*:273–293.

Ptitsyn OB, Rashin AA. 1975. A model of myoglobin self-organization. *Biophys Chem 3*:1–20.

Roder H, Colón W. 1997. Kinetic role of early intermediates in protein folding. *Curr Opin Struct Biol 7*:15–28.

Sali A, Shaknovitch E, Karplus M. 1994. How do proteins fold? *Nature 369*:183–184.

Schindler T, Herrler M, Marahiel M, Schmid FX. 1995. Extremely rapid protein folding in the absence of intermediates. *Nat Struct Biol 2*:663–673.

Schindler T, Schmid FX. 1996. Thermodynamic properties of an extremely rapid protein folding reaction. *Biochemistry 35*:16833–16842.

Searle MS, Williams DH, Packman LC. 1995. A short linear peptide derived from the N-terminal sequence of ubiquitin folds into a water-stable nonnative-hairpin. *Nat Struct Biol 2*:999–1006.

Sieber V, Moe GR. 1996. Interactions contributing to the formation of a β-hairpin-like structure in a small peptide. *Biochemistry 35*:181–188.

Skolnick J, Kolinski A. 1990. Simulations of the folding of a globular protein. *Science 250*:1121–1125.

Sosnick TR, Mayne L, Hiller R, Englander SW. 1994. The barriers in protein folding. *Nat Struct Biol 1*:149–156.

Storch E, Daggett V. 1996. Structural consequences of heme removal: Molecular dynamics simulations of rat and bovine apocytochrome b5. *Biochemistry 35*:11596–11604.

Tirado-Rives J, Jorgensen WL. 1993. Molecular dynamics simulations of the unfolding of apomyoglobin in water. *Biochemistry 32*:4175–4184.

Tirado-Rives J, Orozco M, Jorgensen WL. 1997. Molecular dynamics simulations of the unfolding of barnase in water and 8M aqueous urea. *Biochemistry 36*:7313–7329.

Viguera AR, Martinez JC, Filimonov VV, Mateo PL, Serrano L. 1994. Thermodynamic and kinetic analysis of the SH3 domain of spectrin shows a two-state folding transition. *Biochemistry 33*:2142–2150.

Vijay-Kumar S, Bugg CE, Cook WJ. 1987. Structure of ubiquitin refined at 1.8 Å resolution. *J Mol Biol 194*:531–544.

Vishveshwara S, Ravishanker C, Beveridge DL. 1993. Differential stability of β-sheets and α-helices in β-lactamase: A high temperature molecular dynamics study of unfolding intermediates. *Biophys J 65*:2304–2312.

Wang Y, Shortle D. 1996. A dynamic bundle of four adjacent hydrophobic segments in the denatured state of staphylococcal nuclease. *Protein Sci 5*:1898–1906.

Williams S, Causgrove RG, Fang KS, Callender RH, Woodruff WH, Dyer RB. 1996. Fast events in protein folding: Helix melting and formation in a small peptide. *Biochemistry 35*:691–697.

Wolynes PG, Luthey-Schulten Z, Onuchic JN. 1996. Fast folding experiments and the topography of protein folding energy landscapes. *Chem Biol 3*:425–432.

Yi Q, Baker D. 1996. Direct evidence for a two-state protein unfolding transition from hydrogen-deuterium exchange, mass spectrometry, and NMR. *Protein Sci 5*:1060–1066.