JMB

Increasing Temperature Accelerates Protein Unfolding Without Changing the Pathway of Unfolding

Ryan Day, Brian J. Bennion, Sihyun Ham and Valerie Daggett*

Department of Medicinal Chemistry, University of Washington, P.O. Box 357610 Seattle, WA 98195-7610, USA

We have traditionally relied on extremely elevated temperatures (498 K, 225 °C) to investigate the unfolding process of proteins within the time-scale available to molecular dynamics simulations with explicit solvent. However, recent advances in computer hardware have allowed us to extend our thermal denaturation studies to much lower temperatures. Here we describe the results of simulations of chymotrypsin inhibitor 2 at seven temperatures, ranging from 298 K to 498 K. The simulation lengths vary from 94 ns to 20 ns, for a total simulation time of 344 ns, or 0.34 μs. At 298 K, the protein is very stable over the full 50 ns simulation. At 348 K, corresponding to the experimentally observed melting temperature of CI2, the protein unfolds over the first 25 ns, explores partially unfolded conformations for 20 ns, and then refolds over the last 35 ns. Above its melting temperature, complete thermal denaturation occurs in an activated process. Early unfolding is characterized by sliding or breathing motions in the protein core, leading to an unfolding transition state with a weakened core and some loss of secondary structure. After the unfolding transition, the core contacts are rapidly lost as the protein passes on to the fully denatured ensemble. While the overall character and order of events in the unfolding process are well conserved across temperatures, there are substantial differences in the timescales over which these events take place. We conclude that 498 K simulations are suitable for elucidating the details of protein unfolding at a minimum of computational expense.

© 2002 Elsevier Science Ltd. All rights reserved

Keywords: energy landscape; transition state of unfolding; activated process; molecular dynamics; protein unfolding

*Corresponding author

Introduction

The mechanisms of protein folding and unfolding have been the subject of intensive experimental and theoretical study for several decades. The details of this process at an atomic level, however, have proved elusive to traditional experimental and theoretical methods. The advent of molecular dynamics simulations over the last 15 years has shed some light on this process. In an effort to obtain atomic-level information about the folding/unfolding process, we have been employing computer simulation methods in close collaboration with experimentalists for over a decade. The overall results have been in very good agreement with experiment. In order to unfold a protein, we perform the simulations at very high temperatures,

typically 498 K, or 225 °C. Such drastic measures have been necessary because of the large difference in the experimental timescale for unfolding and that achievable with available computer power. However, as CPU power increases, we can extend the timescale accessible to computer simulations, allowing the denaturation of proteins to be simulated at much more reasonable temperatures.

It has been claimed that high temperature distorts the energy landscape and that high temperature simulations are thus irrelevant. This is usually given as a statement of fact. We do not, however, know that the energy landscape is changed dramatically. It may be that high temperature merely affects the rate of unfolding, as it would for a traditional activated process. Given our experience with many different protein systems in which very good, and even quantitative agreement, is obtained between experiment and high temperature simulations, it seems unlikely that the high-temperature energy landscape is
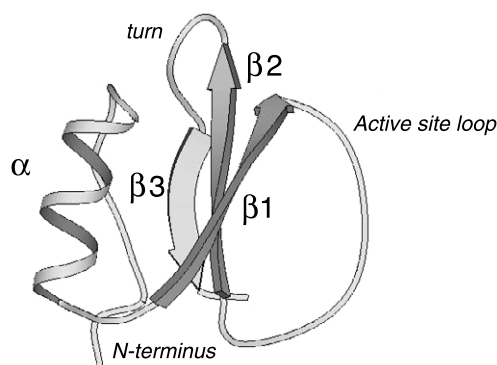

Figure 1. The crystal structure of CI2. The N terminus and helix make up region 1 and the active site loop (miniloop) and β-sheet make up region 2.

grossly different from that probed by experimentalists at lower temperatures. Nevertheless, we have chosen to test this idea more directly and determine to what extent temperature may be affecting the unfolding pathway. To this end, we have performed simulations of chymotrypsin inhibitor 2 (CI2) in water at seven different temperatures ranging from 298 K to 498 K (25–225 °C), with the 348 K simulation falling near the experimental T_m.[1]

CI2 was chosen for this study for a variety of reasons. It is a small, 64-residue single domain globular protein (Figure 1). It folds *via* a two-state mechanism with a single rate-determining transition state that is the same for folding and unfolding. It has been the subject of numerous experimental studies by Fersht and co-workers to completely map its folding/unfolding pathway.[2–10] Previous computational studies have predicted the structure of the transition state for unfolding, which was found to be in quantitative agreement with experimental data.[2,11,12] Further computational and experimental tests support the transition state models.[7,13,14] In addition, the high temperature unfolding pathway[11,15–17] and denatured state of CI2 have been described in detail.[18] Thus, it is a good model for the study of protein unfolding, and by extension folding pathways and the effect of temperature on these pathways.

Results and Discussion

## Behavior of the water was a function of temperature

CI2 was simulated at seven different temperatures: 298, 348, 373, 398, 448, 473, and 498 K. It is important to investigate the properties of pure water under these conditions as the solvent–protein and solvent–solvent interactions play critical roles in unfolding and folding and the simulations contain one protein molecule and approximately 2600 water molecules. Consequently, we also performed simu-

lations of pure water at these temperatures to ensure that the solvent properties are reasonable. We have already shown that our water model behaves as expected with changes in temperature and pressure,[19] but those earlier simulations were very short (200 ps) by today's standards. The protocols and conditions used here conform to those of the protein simulations presented and were performed for 4 ns.

Three experimentally observable aspects of the high temperature simulations were evaluated: water self-diffusion, water–oxygen radial distributions, and orientational relaxation. Water dynamics were analyzed by monitoring diffusion and orientation relaxation of the molecular dipole. In general, water diffusion in the simulations is in good agreement with experiment at the lower temperatures but lower than the experimental values at high temperature (Table 1). The discrepancy between experiment and simulation is less than we had reported previously,[19] however, due to improved sampling. Only slight changes to the density are required to correct the diffusion constant at high temperature. These changes lead to little or no change in the structural properties of the solution (unpublished results). Our measured τ_2 relaxation rate of 1.9 ps at 298 K is in good agreement with the experimental value of 2 ps.[20] With the exception of the 348 K simulation, the relaxation times decrease with increasing temperature. Water structure was examined by calculating the oxygen–oxygen radial distribution function at each temperature. The 298 K simulation is in good agreement with experiment (Table 1). As the temperature is increased, the first peak drops and shifts slightly, and the second peak's height is decreased and its radius is significantly increased. These trends are consistent with our earlier work[19] and findings from other groups.[21,22] Further details regarding the behavior of our water model and detailed solvation properties of peptides and proteins have been presented by Beck *et al.*[23] Overall our pure water simulations are in agreement with the available experimental results across a variety of temperatures, suggesting that this water model does not introduce artifacts when used in high-temperature protein denaturation studies.

Overview of the unfolding simulations

Simulations of CI2 were performed at the temperatures described above for the pure water simulations at neutral pH. The T_m for CI2 is approximately 75 °C, or 348 K.[1] At this temperature $\Delta G_{N–D} = 0$, and we would expect the protein to unfold and refold, provided it has enough energy to overcome the activation barrier for unfolding. Below this temperature (298 K), the protein should be stable and remain folded. Above this temperature, the protein should unfold. These expectations were indeed borne out (shown in part in Figure 2). Below, we describe in detail the behavior of the protein at each temperature. All simulations above the T_m were carried out for at

Table 1. Structural and dynamical properties of pure water as a function of temperature
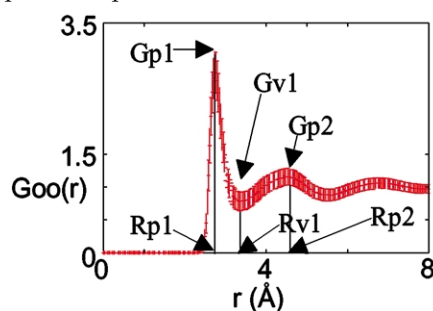
Temperature (K)	$\langle D \rangle$ (Å^2/ps)[b]	Radial distributions $G_{oo}(r)$						Orientational relaxation τ_2 (ps)[c]
		Heights[a]			Positions (Å)[a]			
		Gp1	Gv1	Gp2	Rp1	Rv1	Rp2	
298 MD	0.25	3.17	0.82	1.05	2.70	3.30	4.55	1.9
EXPT[d]	0.24	2.91	0.73	1.14	2.9	3.3	4.5	2.4
348 MD	0.56	2.82	0.90	1.05	2.70	3.55	5.25	0.8
EXPT[e]	0.60	~2.8	~0.9	~1.1	2.7	3.5	5.5	NA
373 MD	0.77	2.68	0.91	1.07	2.70	3.75	5.40	1.1
EXPT[e]	0.81	~2.7	~0.8	~1.0	2.7	3.9	5.3	NA
398 MD	0.99	2.57	0.91	1.07	2.70	3.70	5.40	0.4
EXPT[f]	1.21	2.85	0.84	1.09	2.82	3.69	4.47	NA
448 MD	1.48	2.36	0.91	1.08	2.75	4.10	5.45	0.3
EXPT	1.96							NA
473 MD	1.73	2.29	0.91	1.08	2.75	3.95	5.40	0.2
EXPT	2.38							NA
498 MD	1.98	2.22	0.90	1.08	2.75	3.95	5.40	0.2
EXPT[g]	2.75	1.83	0.93	1.03	2.93	5.13	5.8	NA

[a] Gp1 and Gp2 refer to peak values in the radial distribution, while Gv1 refers to the valley in the distribution. Rp1 and Rp2 refer to distances in angstroms for the peaks. Rv1 is the distance for the first valley (see inset).
[b] Experimental water self-diffusion coefficients taken from Krynicki *et al.*[34] At 298 K, $\langle D \rangle \sim 0.23-0.26$ Å^2 ps^{-1}.
[c] Experimental relaxation values for water at 298 K from Halle & Wennerstrom.[20] Relaxation values do not currently exist for other temperatures.
[d] Experimental water oxygen–oxygen radial distribution values from Soper & Phillips[35] and values from Soper *et al.*,[21] dataset "a" 298 K. http://www.isis.rl.ac.uk/disordered/Database/DB_Main.htm
[e] Radial distribution data from Narten & Levy.[36] Values for peak heights and valley are approximate.
[f] Experimental water oxygen–oxygen radial distribution values from Soper & Phillips[35] and values from Soper *et al.*,[21] dataset "a" 423 K. http://www.isis.rl.ac.uk/disordered/Database/DB_Main.htm
[g] Experimental water oxygen–oxygen radial distribution values from Soper & Phillips[35] and values from Soper *et al.*,[21] dataset "a" 573 K. http://www.isis.rl.ac.uk/disordered/Database/DB_Main.htm

least 10 ns after the protein unfolded in order to sample the denatured state. This led to simulations of the following lengths: 50 ns at 298 K, 80 ns at 348 K, 94 ns at 373 K, 40 ns at 398 K and 448 K, and 20 ns at 473 and 498 K. Thus, we are presenting a total of 344 ns, or 0.344 μs, of simulation time. We will first present each simulation for a given temperature in detail and then discuss the similarities and differences between the pathways.

298 K

The native state simulation was carried out for 50 ns. The protein remained very native-like with an α-carbon root mean square deviation (RMSD) of 1.7 ($\pm$0.2) Å from the crystal structure from 1 ns to 50 ns (Figure 2(a)). The deviation was localized to the mobile active site loop and N terminus. Omitting these regions from the comparison results in a C^α RMSD of 0.7 ($\pm$0.1) Å to the crystal structure (data not shown). The mobility of the loop is in agreement with NMR relaxation data[24] and a detailed comparison of a shorter native simulation of CI2 with the generalized NH order parameters has been presented.[25] The integrity of the packing interactions is evident in the plot of the number of tertiary contacts with simulation time (Figure 2(b)). These results demonstrate that our protocols and procedures can reproduce the native state over long simulation periods. This simulation serves as a control for the high temperature simulations, and will be discussed further below as necessary.

348 K

The protein remained native-like for approximately 3 ns at 348 K and then began to deviate
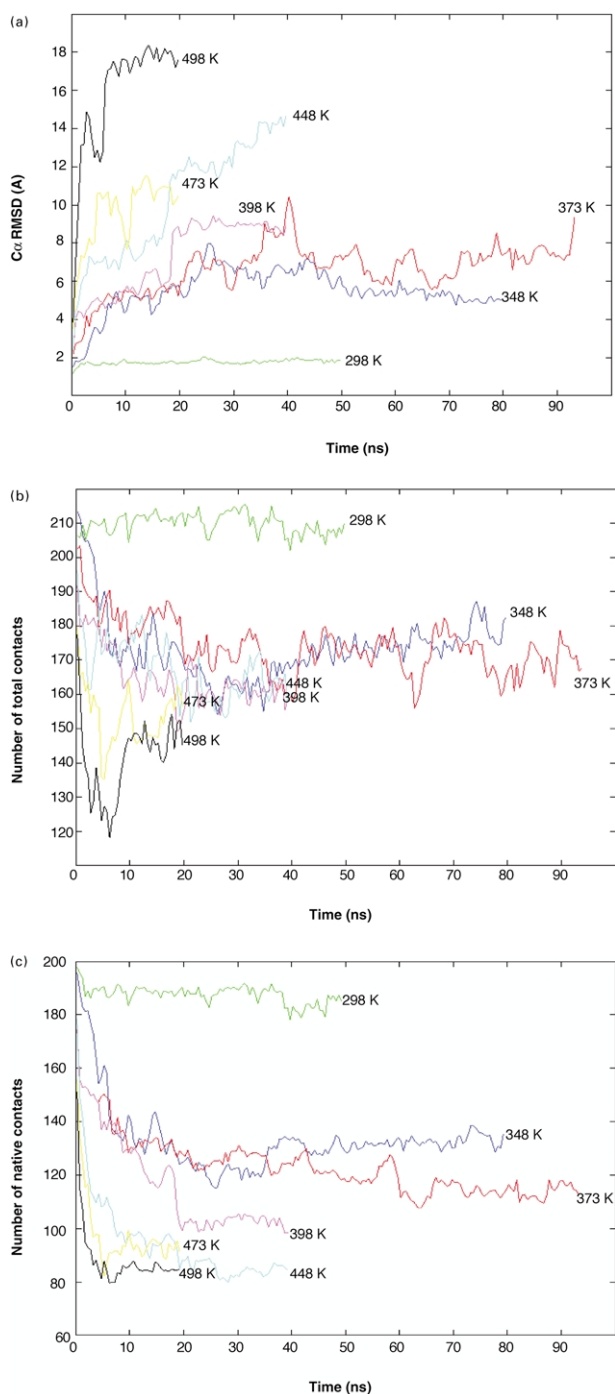
Figure 2. Trace of structural changes in term of. C^α RMSD from the crystal structure with (a) time and (b) temperature. Total tertiary contacts with time and temperature. A contact is defined as occurring when aliphatic carbon atoms of two side-chains come within 5.4 Å or any other atom of two side-chains comes within 4.6 Å. (c) Native tertiary contacts with time and temperature.

more dramatically from the crystal structure (Figure 2(a)). Around 100 ps, the active site loop and the turn between β1 and β2 became mobile. By 800 ps, the N terminus became more mobile and began to separate from the rest of the protein

(Figure 3). Around 3 ns, the turn/loop pulled away from the active-site loop as the β-sheet splayed. After another 0.5 ns, the bottom of the core began to open due to movement of the N terminus. Just before 4 ns, Trp5, the fluorescence probe used experimentally to monitor folding and unfolding, became exposed to the solvent. By 5.4 ns, Trp5 was again buried in the core due to the return and rotation of the N terminus. By 6 ns, the packing of the hydrophobic core was sufficiently weak that the top of the core opened up. At 6.3 ns, the active-site loop moved closer to the top of the β1 (Figure 3). After this time, the helix began to slide relative to the β-sheet. This sliding led to a 90° rotation of the helix in the plane of the sheet (Figures 3 and 10) and attenuation of associated native and non-native packing interactions (for example between Lys17 and Asp55). At approximately 7 ns, the N terminus moved away from the protein, exposing the bottom of the core to the solvent, and the N terminus and the loop were very dynamic. From 11 ns to 12 ns, the movement of the active site loop led to drastic disruption of the β(1-2) sheet, but the breakage was repaired over time.

By 15.5 ns the combined structural changes were enough to completely expose the hydrophobic core to solvent, but the core re-collapsed. This breathing motion was repeated. At 18.4 ns the orientation of the helix was ~80 degrees with respect to β1 and most of the native tertiary interactions in the core were broken (Figure 3). At 21.6 ns the active-site loop approached the helix and the Lys17-Asp55 salt bridge reformed. Some of the hydrophobic interactions in the core also reformed as a result of this motion. At 23.4 ns the N terminus moved away from the core again and the active-site loop became very dynamic. A nanosecond later, the loop collapsed down onto the protein and the protein lost all secondary (aside from one turn of helix) and tertiary structure with a large increase in the C^α RMSD. For roughly the next 25 ns the protein remained disrupted and highly dynamic with a C^α RMSD > 7 Å (Figures 2(a) and 3). After ~50 ns, the helix moved back towards the core and the C^α RMSD decreased. After this time, the loop and the N and C termini were very dynamic, but the radius of gyration, C^α RMSD, and the packing interactions in the core did not change much. Around 62 ns, the active-site loop moved back up, the fraying of the β(1-2) sheet was corrected, and the protein recovered many of its native packing interactions. After another 6 ns, the N terminus and active site loop again became disrupted and the C^α RMSD rose, but it recovered again. The final structure at 80 ns had a fairly nicely packed hydrophobic core, a high degree of secondary and tertiary structure, and a C^α RMSD < 5 Å.

373 K

Early in the simulation (~100 ps) at 373 K, the top of the helix pulled away from the protein and

Unfolding at 348 K

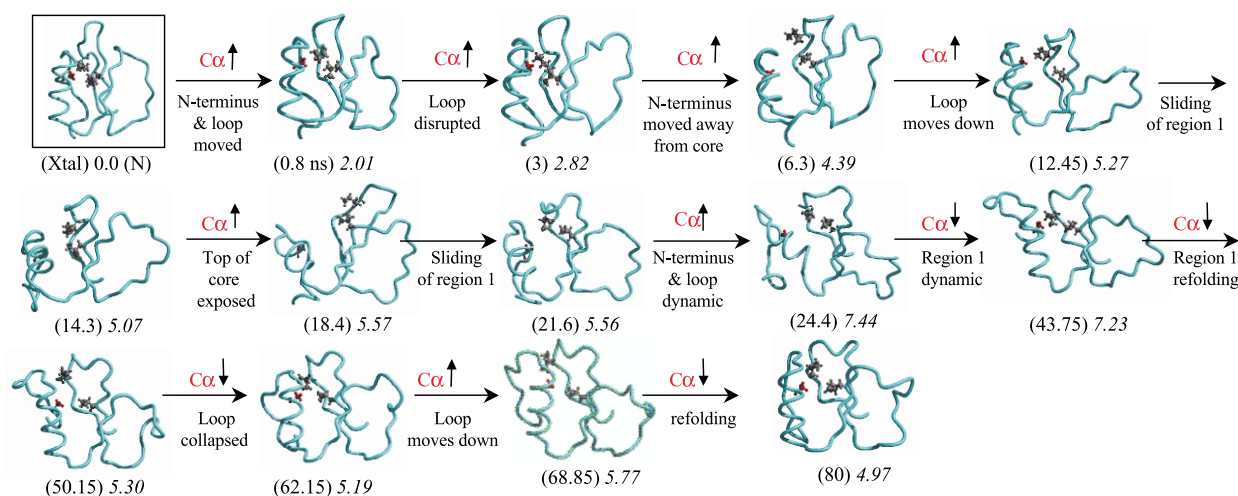**Figure 3**. Structures representative of the conformational changes observed in the unfolding trajectory at 348 K. In this figure and Figures 4–9, side-chains of residues involved in the nucleus for the transition state of unfolding, A16 (in red), L49 and I57 (in gray) are shown. The C$^\alpha$ symbol indicates the trend in RMS deviation of all the α carbons from the native state structure. The RMS deviation of all C$^\alpha$ atoms between each individual snapshot and the native structure is displayed in italics, with the time point of the snapshot given in parentheses. N, TS and D indicate the native, transition and denatured states.

the interaction between Ser12 and Asp55 was disrupted. At 200 ps, the N terminus moved away from the core (Figure 4). At 350 ps, the C terminus moved away from the core exposing the top and back of the core to the solvent. However, at 450 ps, the N and C termini came back to the core and the lost packing interactions were recovered. About 200 ps later the termini moved out again, exposing the core. After this point, the N terminus became very mobile. At 1.05 ns, the tertiary interactions at the top of the core were broken, and the active site loop became very mobile. Around 2.7 ns, the helix began to slide along the sheet, leading to an almost 90° rotation, as described above for the 348 K simu-

lation. For the next 1.2 ns the protein experienced alternating cycles of this sliding motion and breathing of the core.

At 3.9 ns, the active-site loop came closer to the top of β1, in contrast to the 348 K simulation in which this loop moved closer to the helix in a sliding motion. At this point, the helical structure was also more unfolded than in the previous simulations. At 4.2 ns, the active site loop pulled away from β1, and 100 ps later Trp5 became fully exposed to solvent. By 4.7 ns, the sliding motion described above came into play and the Lys17 and Asp55 interaction formed. Native helical structure was subsequently recovered. At 5.25 ns, the helix
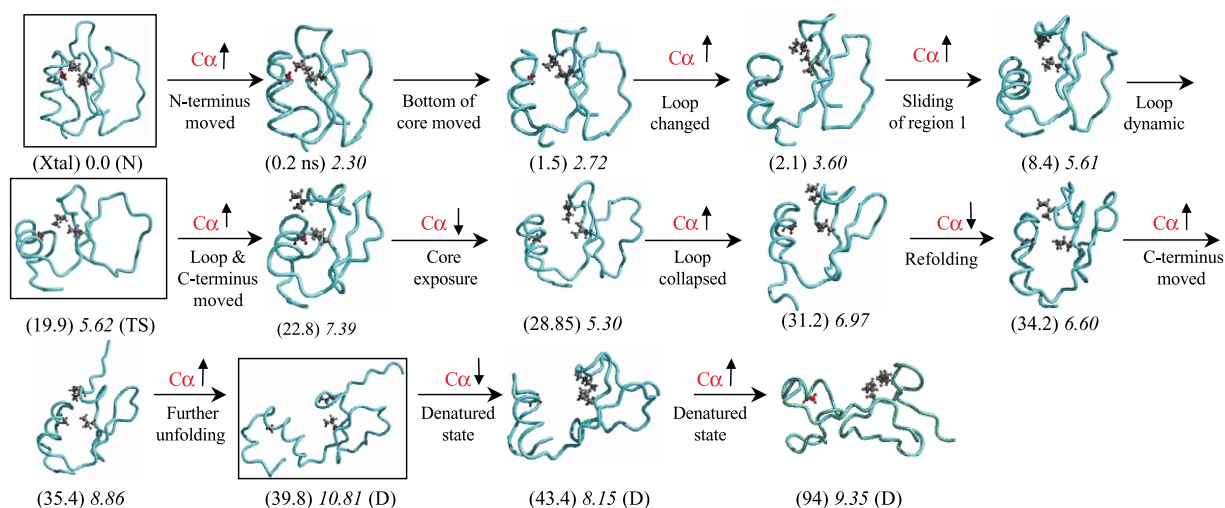
Unfolding at 373 K

Figure 4. Structures representative of the conformational changes observed in the unfolding trajectory at 373 K.

orientation was ~45° relative to the β1 axis and the active-site loop had formed tertiary interactions with the front side of the helix. For the next 3 ns, the protein cycled through sliding and recovery motions (Figure 4). The loop became very mobile and around 9.9 ns the top of the core opened up, but closed at 10 ns. After this time, the loop became very dynamic and the interaction between β1 and β2 weakened. At 13.1 ns, the conformation was similar to the 8.4-ns structure. The process went through another cycle until returning to the 8.4-ns-like structure at 17.75 ns. At 18.35 ns, the N terminus pulled away from the core, further exposing the bottom of the core to solvent. At 21.1 ns, the C terminus pulled away from the core and the C^α RMSD increased sharply. After this time, the loop collapsed back and came closer to the N terminus. At 35.4 ns, the C terminus moved up and was aligned with the helix axis. This movement was not observed in any other simulations, and it led to a dramatic increase in C^α RMSD (Figures 2(a) and 4). At 39.8 ns, the interaction between Lys18 and Asp55 was broken and the core became more exposed to solvent. For the next 4 ns the C^α RMSD fluctuated due to heightened mobility of the N and C termini. The packing interactions between β1 and β2 were partially retained, the helix unfolded and refolded various times, but the protein was severely disrupted with a C^α RMSD of 8–10 Å for the remaining 50 ns.

398 K

By 350 ps at 398 K, the N terminus moved away from the hydrophobic core, causing the bottom of the core to open (Figure 5). The N terminus continued to move away from the core and by 600 ps the active-site loop had become distorted and the turn between β2 and β3 pulled away from the protein. By 850 ps, however, the loop moved back into

place and hydrophobic interactions involving β2 were recovered. Up until approximately 4.5 ns, the main events were the conformational changes in the active-site loop and N terminus. At about this time, the helix began a sliding motion similar to that observed at 348 K and 373 K.

At 8.2 ns, the interaction between Ser12 and Asp55 was broken, the helix slid back towards the protein, the active-site loop moved closer to the front side of the helix, and the N terminus became totally exposed to solvent. At 8.85 ns, the Lys17-Asp55 interaction reformed due to the sliding of the helix against the sheet, but the sliding motion continued. At 11.8 ns, the active-site loop came forward and the top of the core recovered some packing interactions. At 11.95 ns, the interaction between Lys17 and Asp55 broke again. By 12.29 ns, the core had become fully exposed to solvent. Sliding continued to occur, which brought residues 17 and 55 back into contact around 12.95 ns. After this time point, the C^α RMSD increased sharply due to motion of the N and C termini. At 14.4 ns, the N terminus returned to the core, partially protecting it from solvent and causing a decrease in the C^α RMSD.

At 14.8 ns, the packing in the core region became weakened and the solvent accessible surface area of the core increased. The nucleation sites around Ala16 and Leu49 also came apart. At 14.85 ns, the N terminus moved away from the core, increasing the C^α RMSD. The core opened up further at 15.25 ns.

At 18.55 ns, the loop collapsed down, and the C^α RMSD, radius of gyration and solvent accessible surface area increased dramatically. By 18.6 ns, the tertiary interactions between β1 and β2 were almost entirely broken. At 19.05 ns, the N terminus and the turn between β2 and β3 came together and formed new interactions involving Glu7, Arg43, and Asp45. Even as late as 21.85 ns the protein

Unfolding at 398 K

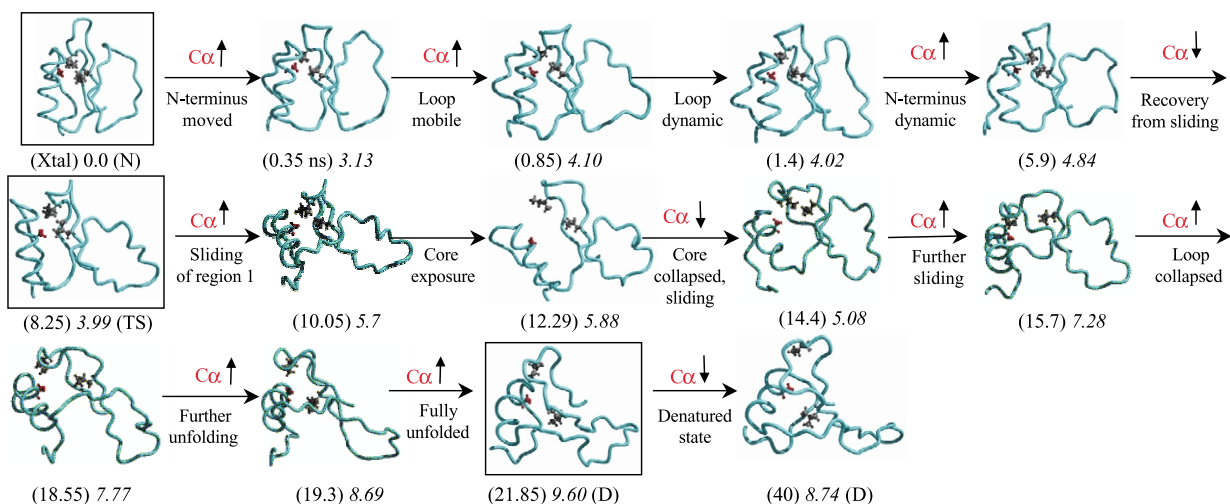**Figure 5.** Structures representative of the conformational changes observed in the unfolding trajectory at 398 K.
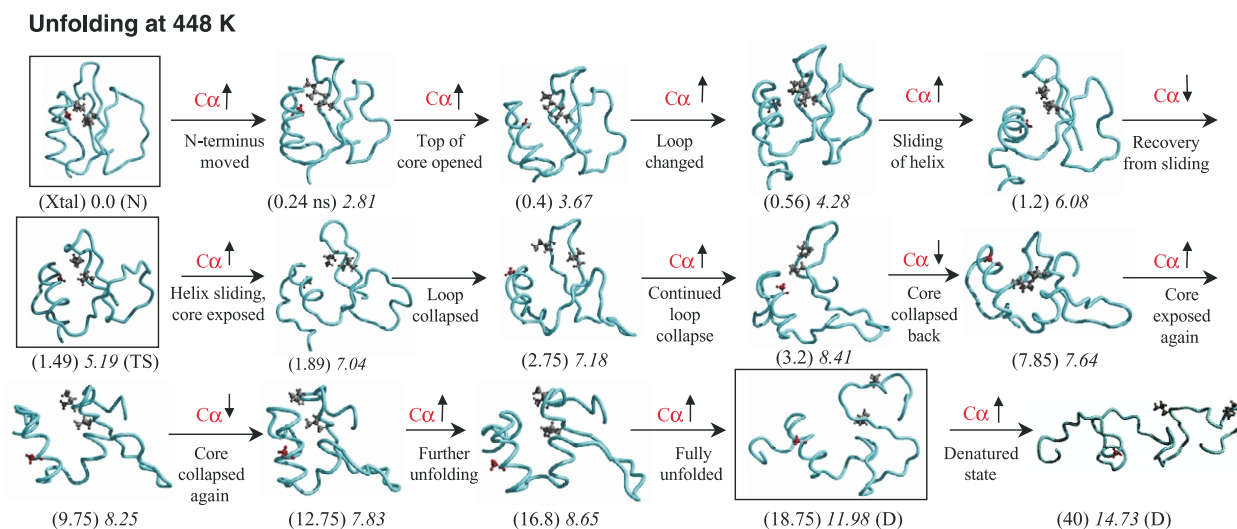
Unfolding at 448 K

Figure 6. Structures representative of the conformational changes observed in the unfolding trajectory at 448 K.

displayed some residual structure, but the protein was definitely unfolded and the bulk of its native interactions were lost. The last structure in the trajectory had a C^α RMSD of 8.74 Å, which is much lower than that obtained at higher temperatures. This difference is almost entirely due to the more expanded nature of the protein at the higher temperatures; the extent of local structure in the denatured state was in fact similar at the different temperatures.

448 K

During the first 200 ps of the 448 K simulation the N terminus and active-site loop became very mobile. At 230 ps, Lys2 and Glu7 came into close contact, causing the N terminus to move away from the core. Up through 500 ps, there was a weak breathing motion between the helix and sheet, but the core was not exposed to solvent during this process. At 560 ps, the helix moved down and the C^α RMSD increased due to the N terminus and turn displacements, but the solvent accessible surface area of the core region did not increase appreciably (Figures 2 and 6). At 820 ps, the sliding motion described above began to come into play. This led to the bottom of the core opening and the N terminus pulling away from the core at 910 ps. At 1.2 ns, the helix was almost perpendicular to β1.

After 1.45 ns, the active-site loop became very disrupted, which was followed by a loss of packing interactions between β1 and β2. At 1.58 ns, the loop collapsed down onto the protein and β2 rotated to fully expose Leu49 to solvent. β2 and β3 came apart with β2 moving towards the loop while β3 stayed with the helix. At 1.65 ns, the loop collapsed, which increased the C^α RMSD while decreasing the radius of gyration and β-content. Around 3.05 ns the loop collapsed further, disrupting the remaining interactions between β2 and β3.

At 3.9 ns, the N terminus pulled away and by 5 ns the C terminus moved away from the core. At 5.25 ns, the loop collapsed and the interactions between β1 and β2 were broken again. The nucleation site, consisting of residues 16, 49, and 57, was still intact. At 5.9 ns, the helix slid away from the core and not much later, 6.15 ns, Ala16 lost contact with Ile57. After 7.45 ns, the interactions between β2 and β3 were broken, but they were recovered immediately. This motion was repeated many times. At 9.25 ns, the helix and N terminus dropped relative to β2 and β3. At 12.75 ns, the core collapsed back and the C^α RMSD decreased. At 16.8 ns, the core became exposed again and the N terminus moved away. There was then a sharp increase in the C^α RMSD due to movement of the N terminus and the active site loop unfolding. The protein was fully unfolded at 18.75 ns with a C^α RMSD of 12.0 Å. The final C^α RMSD was 14.7 Å at 40 ns.

473 K

At 30 ps, Ile57 and Ala58 moved away from Gly10, breaking the tertiary interactions at the top of the hydrophobic core (Figure 7). At 60 ps, the top of the α-helix also moved away from the core. Next, at 70 ps, β1 and the loop pulled away from the core. Then, at 90 ps, the N terminus moved out, exposing the bottom of the core to solvent. The core and loop became very dynamic such that at 260 ps, Trp5 became exposed to solvent. At 300 ps, the tertiary interactions in the core were weakened and the bottom of the core was exposed further to the solvent. After this time, there was a breathing type motion between the helix and sheet and the core was very dynamic. At 560 ps, the top of the helix slid back and a non-native interaction between Lys17-Asp55 formed.

At 600 ps, the helix slid back and twisted relative to the sheet. This motion continued over the next
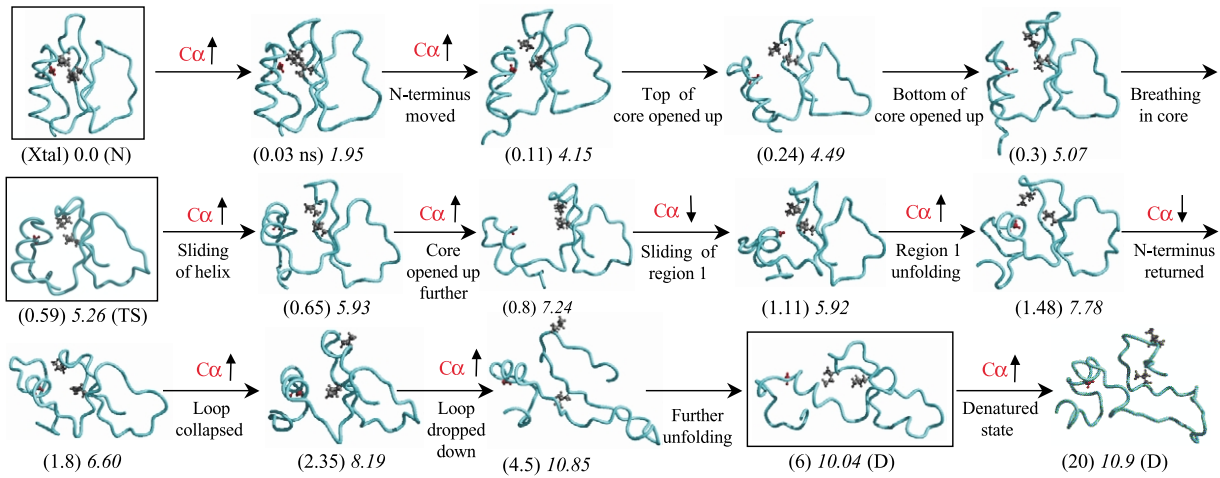
Unfolding at 473 K

Figure 7. Structures representative of the conformational changes observed in the unfolding trajectory at 473 K.

50 ps and led to an increase in the radius of gyration (R_g) and C^α RMSD (Figure 2(a)), and a decrease in the tertiary contacts (Figure 2(b) and (c)). At 760 ps, the helix pulled away from the protein, exposing the top of the core to the solvent, which then triggered the departure of the N terminus, exposing the back of the core to solvent (Figure 7). The hydrogen bond between Lys17 and Asp55 was still present. At 970 ps, the protein collapsed back and the core became less exposed to the solvent (Figure 7). At 10.8 ns, the structure resembled the 620 ps structure. At 11.1 ns, the helix slid almost 90°, but the nucleation site (residues 16, 49 and 57) remained intact.

At 1.48 ns, the N terminus became fully exposed to the solvent and highly mobile. At 1.52 ns, the bottom of the core was fully exposed and Trp 5 rotated away from the core. At 1.56 ns, the N terminus was pulled outside exposing the bottom of the core to solvent. The core then expanded and the C^α RMSD increased. After this time, the loop, N terminus and active site loop became very dynamic and the C^α RMSD increased. The N terminus then repacked against the core and the C^α RMSD and total solvent accessible surface area decreased. From 1.76 ns to 1.8 ns, the nucleation site residues were back in contact and the core was less exposed. At 1.83 ns, the top of the core was exposed again with the N terminus and loop becoming very dynamic by 1.91 ns. 40 ps later the core collapsed back. At 2.35 ns, the N terminus moved away from the core and the C^α RMSD increased. At 2.65 ns, the helix unfolded. At 2.8 ns, the interactions between β1 and β2 broke and β1 and the active site loop were pulled down away from the rest of the protein. After this time, the native contacts and secondary structure were almost completely destroyed. By 6 ns, the protein was fully unfolded and the C^α RMSD was 10.0 Å. For the final 14 ns, various denatured structures

were sampled. The final C^α RMSD was 10.9 Å (Figure 7).

498 K

Unfolding occurred much more rapidly at 498 K. By 10 ps the packing between the core residues Leu11 and Ile57 was compromised, which, in turn, weakened the tertiary interactions at the top of the core (Figure 8). By 55 ps, the top region of the core opened and the N terminus and the active site loop became mobile. At 80 ps, the bottom of the core started to open up by the N terminus, causing the top of the core to become more tightly packed. The protein then went through cycles of the breathing motion described above. At 100 ps, β3 and the C terminus moved away from the N terminus, weakening the hydrophobic core. At 140 ps, the N terminus pulled away from the core, and 15 ps later the top of the active site loop moved away from the core, causing fraying of the top of the β-sheet between β1 and β2. Also, the interaction between Phe50 and Val34 was lost by the rotation of the phenyl ring away from the β1. At 170 ps, the top of the N terminus moved farther away and the core opened further (Figure 8).

At 185 ps, the N terminus and Trp5 became exposed to solvent. Then, β1 and the loop moved away from the core. At 195 ps, the bottom of the N terminus moved towards the C terminus, closing the bottom of the core and decreasing the C^α RMSD. At 220 ps, the top of the α-helix pulled away and the loop became very dynamic. At 255 ps, the top of the core opened up further and it repeated the breathing motion described above. By 300 ps, the top of the core had opened up completely. In the next 30 ps, the helix and β1 pulled away from the core, the protein expanded dramatically, and the solvent accessible surface area increased.
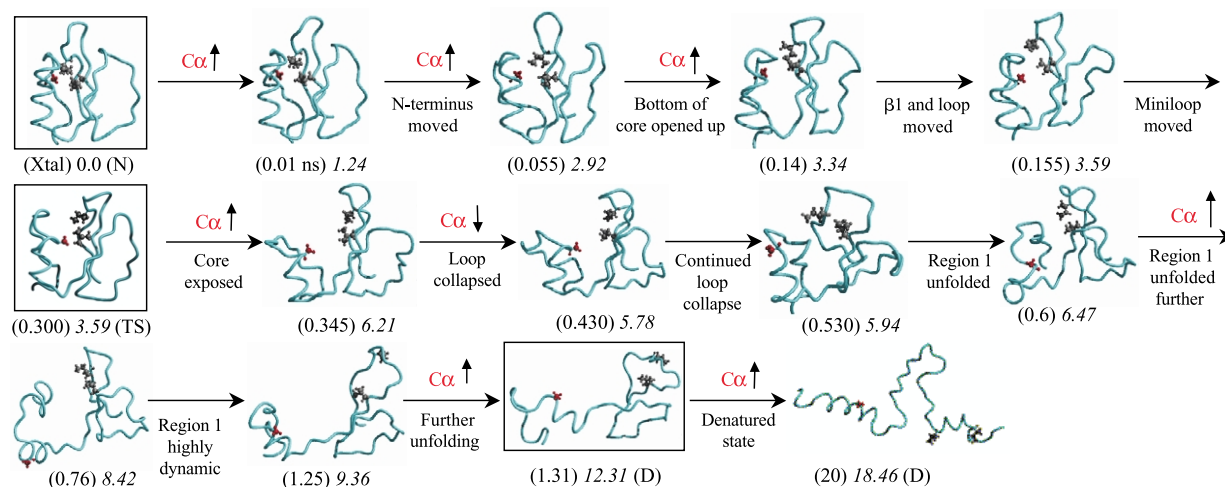
Unfolding at 498 K

Figure 8. Structures representative of the conformational changes observed in the unfolding trajectory at 498 K.

At 340 ps, the loop pulled down further and over the next 5 ps most of the tertiary interactions in the core were broken and the core became fully exposed to solvent. At 355 ps, the helix and sheet moved further apart and the helix began to fray. At 420 ps, the core re-collapsed and the solvent accessible surface area decreased. Similar cycles were observed until 760 ps, when the protein opened up dramatically (Figure 8). The N terminus came back towards the core and the C^α RMSD again decreased at 840 ps (Figure 2). By 1.31 ns the protein was fully unfolded with a C^α RMSD of 12.3 Å. After that time, the protein sampled denatured conformations.

Comparison of the simulations at different temperatures

Temperature changes the timescale but not the pathway of unfolding

All of the simulations followed a similar order of unfolding events regardless of temperature. The temperature affected the timescale of the transitions but not the overall nature of those transitions (Figure 9). Similar, although less extensive, results have been obtained for the unfolding of the engrailed homeodomain[26] and the WW domain.[27]

The first step in the unfolding of CI2 was the expansion of the protein and disruption of the packing of the hydrophobic core. The loss of specific hydrogen bonds has been used to illustrate this expansion at 498 K.[15] Disruption of these hydrogen bonds occurred concomitantly with the expansion of the protein at all temperatures. As the hydrophobic core expanded, the packing of the N terminus against the α-helix became disrupted, the total number of tertiary contacts in the hydrophobic core decreased, and the solvent accessible surface area and radius of gyration increased.

While these linked events occurred in every simulation, they did so at different times, with the rate increasing with increasing temperature. This event occurred at approximately 50 ps at 498 K, 90 ps at 473 K, 240 ps at 448 K, 300 ps at 398 K, 600 ps at 373 K and 6.3 ns at 348 K (Figure 9).

Once the hydrophobic interactions in the core weakened, the protein began to experience breathing or sliding motions between the helix and N terminus and the β-sheet and the active site loop. In the highest temperature simulations ($\geq$448 K), the high thermal energy of the system accelerated the unfolding events with a vigorous breathing motion, whereas in the lower temperature simulations ($\leq$398 K), the protein experienced a less destructive sliding motion with a much slower rate. In addition to the hydrophobic core weakening, the active-site loop became very dynamic. At about 160 ps in the 498 K simulation, the active site loop pulled away from the rest of the protein, causing even greater exposure of the hydrophobic core to the solvent. This event occurred in all of the simulations but at later times at lower temperatures.

After the initial disruption of the core, the protein entered the transition state ensemble. The details for the identification and characterization of the transition state ensembles from the simulations followed published procedures.[11,15] The transition state ensembles will be described in detail elsewhere (R.D. & V.D., unpublished results). The transition state conformations identified at different temperatures are similar and all of them display partial disruption of the secondary structure and a weakened hydrophobic core (Figure 9). In addition, the average number of tertiary contacts in the unfolding transition state was essentially the same at all temperatures in which thermal denaturation occurred. This number ranged from 171 contacts at 498 K to 174 contacts at 373 K, with an average value of 172
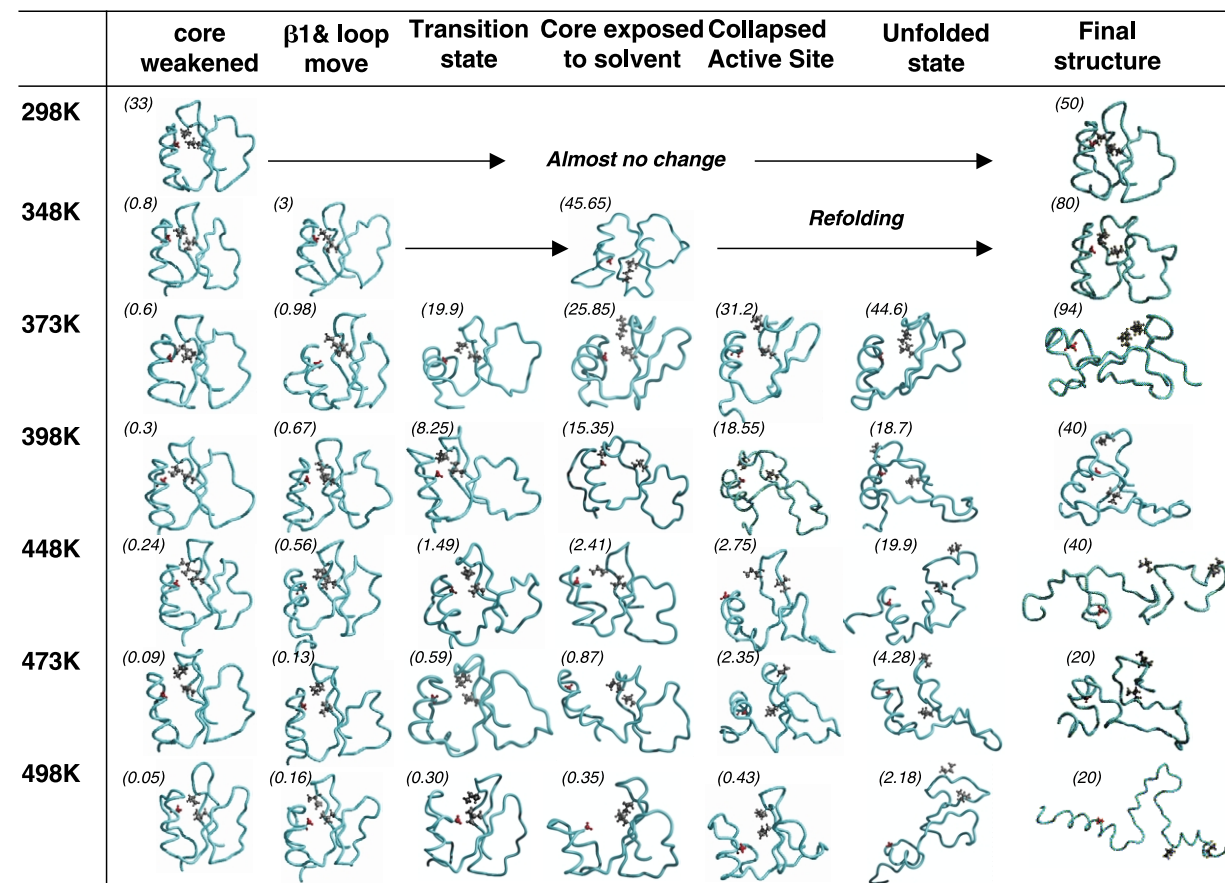
Figure 9. Similarity of unfolding events at seven different temperatures. No observable unfolding events occurred in the 298 K trajectory. At 348 K (which is near the CI2 melting temperature), only early events in the unfolding pathway were observed before the protein began to refold. All the other high temperature simulations show the same sequence of unfolding events. The transition state and unfolded time points were determined by conformational clustering.[11] The core exposed to solvent time point was based on the solvent accessible surface area of the side-chains of residues 5, 8, 16, 19, 20, 29, 31, 47, 49, 51, 57, and 61. The other time points were defined by qualitative observation of the trajectories. These times (in ns) are given in parentheses.

contacts. While there were fairly large fluctuations in this value, the average value did not drop below 172 contacts until near the transition state (Figure 2(b)). The time the system takes to reach the transition state ensemble, however, is very sensitive to the simulation temperature. The transition state ensembles are located at approximately 300 ps at 498 K, 590 ps at 473 K, 1.47 ns at 448 K, 8.25 ns at 398 K, and 19.9 ns at 373 K. Thus, the thermal denaturation of CI2 appears to be simply a matter of breaking enough interactions to allow complete denaturation. The rate at which these interactions are broken is dependent on the available thermal energy. We have not found any clear trends or hierarchies in the types of contacts contributing to the stability of the pre-transition state ensemble. A more thorough analysis of transition state structures and the forces contributing to their stability will be presented in a future paper (R.D. & V.D., unpublished results).

After the transition state, the protein experienced a large structural change and a sharp increase in C^α RMSD at all temperatures. The hydrophobic core region was fully disrupted and exposed to the sol-

vent and much of the native tertiary structure was lost, with the active site loop becoming totally deformed. Water molecules were often near the core region, but the mobility and dynamics of the water were high enough that no significant solvent-induced unfolding events were observed. At this stage of unfolding, the C^α RMSD increased from ~ 4 Å to ~ 8 (higher T) or ~ 6 (lower T) Å.

The last major structural change occurred as the protein entered the fully denatured state at 1.31 ns at 498 K, 6 ns at 473 K, 9.75 ns at 448 K, 21.85 ns at 398 K, and 39.95 ns at 373 K (Figure 9). Generally, at this time, most of the native secondary and tertiary structure was absent although some residual structure was observed. The packing of the hydrophobic core was destroyed, although new, small hydrophobic clusters formed. The protein was highly dynamic at all temperatures and explored many diverse conformations. The detailed characteristics of the denatured state of CI2 for multiple simulations at 498 K are reported elsewhere.[18] The C^α RMSD increased from ~ 8 (higher T) or ~ 6 (lower T) Å to ~ 12 (higher T) or ~ 8 (lower T) Å going into the denatured state. The more unfolded

and expanded denatured states in the higher temperature simulations are due to two factors. One is the diverse nature of the denatured state conformation and the simulation time limit for the lower temperature trajectories. Even 94 ns of simulation time at 373 K may not be sufficient to fully sample the conformational space available to the denatured state. Additionally, at the lower temperature, the solvent density is higher, which helps to maintain the protein in a relatively compact denatured state, limiting its C^α RMSD. The heterogeneity within the 498 K denatured states ($\langle C^\alpha$ RMSD$\rangle$ between structures is ~ 12 Å) is similar to the heterogeneity between the denatured states obtained at different temperatures ($\langle C^\alpha$ RMSD$\rangle$ between structures is ~ 12 Å).

The order of loss of the hydrogen bonding networks defining the protein's secondary structure (Figure 10(a)) was also observed to be essentially conserved across temperatures (Figure 10(b)). First lost are the very high contact order bonds between the N terminus and the third strand of the beta sheet. The hydrogen bonds of the beta sheet are lost next over a fairly long time span. Finally, the hydrogen bonds of the helix break.

Temperature-dependent differences in unfolding

Although the overall unfolding pathway is independent of temperature, differences are observed from trajectory to trajectory. The order of loss of specific side-chain contacts (Figure 10(a)) is not the same at the different temperatures studied (Figure 10(c)). This is also seen in the order of loss of specific main-chain hydrogen bonds making up the β-sheet in Figure 10(b). Many hydrophobic contacts and hydrogen bonds, both native and non-native, may contribute to the overall stability of a given region of protein as it unfolds. These contributions are similar in magnitude and highly degenerate. Thus, the order of events in the unfolding is not dependent on the order of loss of specific contacts.

In previously studied 498 K simulations of CI2 and the 498 K simulation herein, the main type of motion observed during the unfolding process was a breathing motion between the N-terminal strand and helix with respect to the β-sheet and active site loop. After the local unfolding of the N terminus and the increase in mobility of the loop, the thermal energy in the system is delivered to the hydrophobic core. This is seen in the disruption and then reformation of the tertiary interactions in the core. Eventually, the tertiary interactions in the core are weakened to the point that the system has enough energy to overcome the major energy barrier of unfolding and the hydrophobic core becomes fully exposed to the solvent. Further unfolding events occur rapidly as the protein progresses to the denatured state.

In the 473 K and other lower temperature simulations, a new motion is introduced during unfolding of the core region (Figure 11). At 650 ps of the 473 K simulation, the helix, which is parallel to the β strands in the native state, is rotated along an axis perpendicular to the plane of the sheet. In order to maintain tertiary interactions, the loop connecting strands 2 and 3 pulls over the helix. In contrast to the breathing motion described above, the total solvent accessible surface area of the core region and the total volume of the protein are not increased significantly by this sliding motion. This may be why the sliding motion is accessible at lower thermal energies than the breathing motion. At 473 K, loss of secondary and tertiary interactions occurred concurrently with the helix sliding to lead to the core opening fully to the solvent.

At 448 K, the sliding of the helical region began at about 1.2 ns, before the transition state of unfolding. Interestingly, the sliding motion was not a precursor to the unfolding event. Instead, the helix moved back to almost its native orientation. Another new observation at this temperature is the transient re-collapse of the core at 12.75 ns after its initial solvent exposure at 1.89 ns. This re-collapse of the core is accompanied by a slight decrease in the C^α RMSD and is indicative of an increased reversibility of the unfolding pathway at lower temperatures.

At 398 K, the exchange between sliding and recovery happens more frequently during the unfolding. A coupled sliding and breathing motion appeared to drive unfolding. This coupling between the sliding and breathing motion was also observed at 373 K between 2.7 and 3.9 ns. As the temperature decreased, the amplitude of the motion decreased.

In addition to the obvious effects on the amount of thermal energy driving unfolding, increasing the temperature is accompanied by a decrease in the density of water. As one would expect, this leads to a drop in the number of water molecules associated with the protein as the temperature increases (Table 2). Additionally, differences in these values are seen as the protein approaches and enters the transition state ensemble. When normalized by the density, however, these values are all seen to fall within the range of water molecules associated with the native protein at 298 K. The one exception to this is the transition state ensemble at 373 K, which has significantly more water associated with it than would be expected based on the other transition states.

The lowest temperature at which unfolding events were observed was 348 K, which is very close to the T_m of the protein. In general, the protein appeared to unfold over the first 25 ns and then it refolded from 45 ns to 80 ns (Figure 3). During the first 25 ns of the simulation, the protein underwent exchange motions between sliding, breathing, and recovery. After this, the protein recovered native secondary and tertiary structures and the core re-collapsed. The last event observed was an increase in the dynamics of the N terminus and loop, which is the first unfolding event in all simulations.
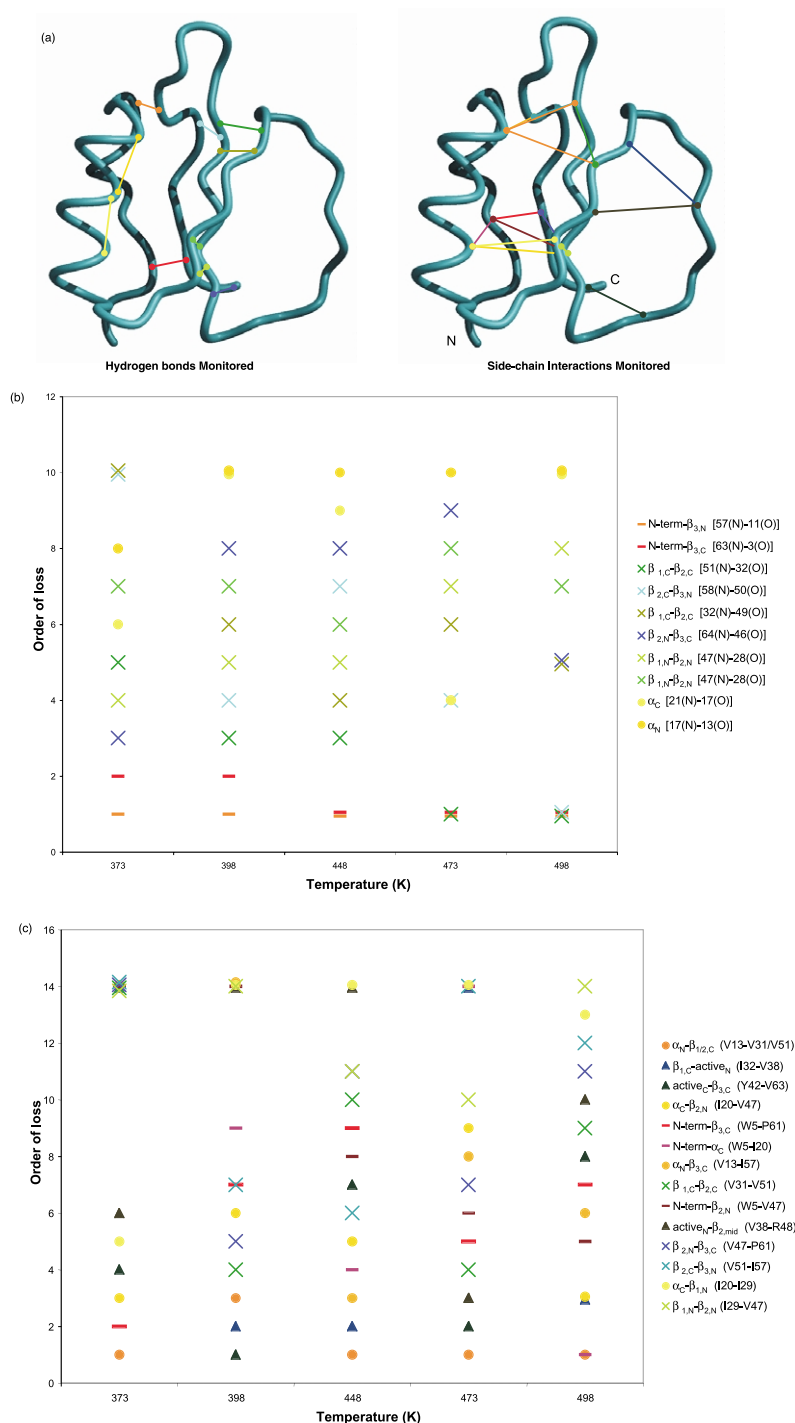
Figure 10. Order of loss of representative native main-chain and side-chain contacts. (a) Location of contacts monitored. Main-chain hydrogen bonds are mapped in the Figure on the left, and side-chain contacts on the right. Coloring is as in the legends to (b) and (c). (b) Order of loss of main-chain hydrogen bonds. Loss is defined by the point at which the bond is intact for less than 50% of a 200 ps interval and does not re-form over the course of the rest of the simulation. (c) Order of loss of side-chain contacts. Loss was determined by comparison to the contact distance in the 298 K simulation. The point of loss is the point at which the contact is broken and does not reform over the course of the simulation. At least one contact at each temperature was not completely lost by this definition. These contacts are assigned a rank of 14. These contacts are lost and reformed many times over the course of the protein's unfolding.

Conclusions

Systematic investigation of the unfolding of CI2 as a function of temperature indicates that unfolding is essentially an activated process. Thus, the pathway is not substantially changed across a wide range of temperatures. At all temperatures, the protein began unfolding by expanding slightly
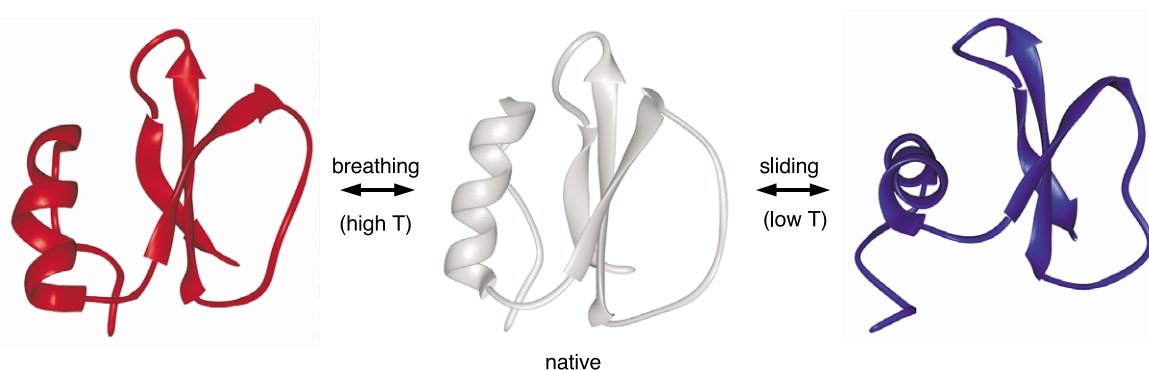
Figure 11. Breathing and sliding-type motions, typical in the early unfolding at high and low temperatures, respectively, are shown. At high temperature direct breakdown of the compact tertiary packing is possible between regions 1 and 2. The sliding movement involves breaking fewer tertiary contacts, making it more favorable at the lower temperatures.

with a corresponding disruption of the core packing. The initial expansion of the protein was followed by either a sliding or a breathing motion of the C-terminal helix relative to the N-terminal strand and active site loop. After undergoing this motion for some time, the protein reached the unfolding transition state, characterized by a weakened hydrophobic core and some loss of secondary structure. Once it passed through the transition state, the protein core rapidly became fully disrupted and the active site loop was highly distorted. In time, the protein reached a fully denatured state in which there was virtually no native structure, although fluctuating secondary and non-native tertiary structures were populated. At lower temperatures, a sliding motion of the helix relative to the sheet became more prevalent in the pre-transition state ensemble, but the primary effect of lowering the temperature was merely to increase the length of time that the protein spends in each of these states.

No single force emerged as a dominant contributor to the thermal behavior of the protein and the order of loss of specific native contacts was not conserved across these temperatures. The total number of contacts formed in the transition state ensemble and order of global unfolding events was, however, essentially the same at all temperatures. These observations suggest that the thermal denaturation of proteins is an activated process taking place on an energy landscape that is not grossly changed by elevated temperatures. The barriers to unfolding on this energy landscape can be thought of as the sum of the interaction energies of each contact in a given element of structure. The order in which these contacts are broken changes from one simulation to the next, but the protein can be expected to cross the lower barriers before higher ones, regardless of temperature.

Table 2. Hydration properties in the simulations

Temperature (K)	Density	At t_{start}[a]	Sigma[b]	Normalized[c]	At $t_{tse}/2$	Sigma	Normalized	At t_{tse}	Sigma	Normalized
(A) Number of water molecules in the first hydration shell[d]										
298[e]	0.997	155.1	6.5	155.6	155.5	5.3	156.0	175.4	5.6	175.9
373	0.958	154.7	6.2	161.5	167.0	7.0	174.3	194.2	7.5	202.7
398	0.939	142.0	6.3	151.2	152.3	6.5	162.2	168.2	7.6	179.1
448	0.89	137.7	7.7	154.7	137.5	6.7	154.5	145.6	6.4	163.6
473	0.861	133.2	5.1	154.7	155.1	8.0	180.1	142.4	5.6	165.4
498	0.829	136.7	7.3	164.9	137.9	5.3	166.3	139.3	6.6	168.0
(B) Number of protein−water hydrogen bonds										
298	0.997	129.3	5.0	129.7	127.7	4.1	128.1	138.6	4.4	139.0
373	0.958	125.7	5.2	131.2	133.0	4.2	138.8	137.0	4.5	143.0
398	0.939	114.1	5.7	121.5	119.4	4.2	127.2	130.4	5.1	138.9
448	0.89	108.9	5.8	122.4	112.9	5.0	126.9	116.4	7.6	130.8
473	0.861	104.4	5.1	121.3	118.7	5.3	137.9	110.8	6.5	128.7
498	0.829	109.7	6.0	132.3	114.4	6.7	138.0	109.5	5.6	132.1

[a] t_{start} was considered to be 25 ps to allow the system to come to temperature. The number of water molecules or hydrogen bonds was averaged over a 10 ps interval about the specified time point.

[b] Sigma is $\sqrt{|\langle x^2 \rangle - \langle x \rangle^2|}$ over the 10 ps interval.

[c] The number of water molecules or hydrogen bonds was normalized by dividing by the density.

[d] The first hydration shell is defined as all water molecules within 3.5 Å of any polar side-chain atom or 4.5 Å of any non-polar side-chain atom.[23]

[e] The $t_{tse}/2$ and t_{tse} points were taken at 25 ns and 50 ns, respectively, for the 298 K simulation. For the other temperature t_{tse}, the time for the transition−state ensemble is given in Figures 3−8.

Methods

All molecular dynamics (MD) simulations were performed using the program ENCAD.[28] The potential energy function and the protocols for molecular dynamics are described elsewhere.[19,28,29] The initial structure used in all simulations was the crystal structure solved at 1.7 Å resolution (1YPC, Harpaz *et al.*[30]). Seven simulations were performed at different temperatures: 498, 473, 448, 398, 373, 348, 298 K. All simulations were carried out at neutral pH.

The protocols for the MD simulations were as follows. The starting structure was minimized for 1000 steps. After minimization, water molecules were added to solvate the protein in a rectangular box extending at least 8 Å in all directions resulting in 2596 water molecules. The water density was set to the experimental value for the temperature in question by adjusting the volume of the box. The final densities were: 0.829 g/ml at 498 K, 0.861 g/ml at 473 K, 0.890 g/ml at 448 K, 0.939 g/ml at 398 K, 0.958 g/ml at 373 K, 0.975 g/ml at 348 K, and 0.997 g/ml at 298 K.[31,32] The solvent water was then subjected to the conjugate gradient minimization of 1000 cycles followed by 1000 steps of molecular dynamics. The water was then minimized again for 1000 cycles. Finally, the protein was minimized for 1000 steps, followed by 1000 steps of minimization of the entire protein–water system. After these preparatory steps, the system was heated to the desired temperature and the simulations were allowed to evolve over time using 2 fs integration time step. Initial atomic velocities were assigned from a Maxwellian distribution. Atoms were allowed to move according to Newton's equations of motion and the velocities of the atoms were adjusted intermittently until the system reached the desired temperature. No further velocity scaling was needed after that point. In all calculations, an 8 Å non-bonded interaction cut-off was used and the non-bonded list was updated every five cycles, except for at 298 K in which the non-bonded list was updated every two cycles from 15 ns to 50 ns time period. The simulation times were 20 ns at 498 and 473 K, 40 ns at 448 and 398 K, 94 ns at 373 K, 80 ns at 348 K, and 50 ns at 298 K. Figures were created using Midas.[33]

Acknowledgments

References

1. Jackson, S. E. & Fersht, A. R. (1991). Folding of chymotrypsin inhibitor 2. 1. Evidence for a two-state transition. *Biochemistry*, **30**, 10428–10435.
2. Otzen, D. E., Itzhaki, L. S., elMasry, N. F., Jackson, S. E. & Fersht, A. R. (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.
3. Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1995). Detailed structure of the transition state for folding of chymotrypsin inhibitor 2 analyzed by protein engineering. *J. Mol. Biol.* **254**, 260–288.
4. Itzhaki, L. S., Neira, J. L., Ruiz-Sanz, J., de Prat Gay, G. & Fersht, A. R. (1995). Search for nucleation sites in smaller fragments of chymotrypsin inhibitor 2. *J. Mol. Biol.* **254**, 289.
5. de Prat Gay, G., Ruiz-Sanz, J., Neira, J. L., Itzhaki, L. S. & Fersht, A. R. (1995). Folding of a nascent polypeptide chain in vitro: cooperative formation of structure in a protein molecule. *Proc. Natl Acad. Sci. USA*, **92**, 3683–3686.
6. Itzhaki, L. S., Neira, J. L. & Fersht, A. R. (1997). Hydrogen exchange in chymotrypsin inhibitor 2 probed by denaturants and temperature. *J. Mol. Biol.* **270**, 89–98.
7. Ladurner, A. G., Itzhaki, L. S., Daggett, V. & Fersht, A. R. (1998). Synergy between simulation and experiment in describing the energy landscape of protein folding. *Proc. Natl Acad. Sci. USA*, **95**, 8473–8478.
8. Neira, J. L., Itzhaki, L. S., Otzen, D. E., Davis, B. & Fersht, A. R. (1997). Hydrogen exchange in chymotrypsin inhibitor 2 probed by mutagenesis. *J. Mol. Biol.* **270**, 99–110.
9. Neira, J. L., Itzhaki, L. S., Ladurner, A. G., Davis, B., de Prat Gay, G. & Fersht, A. R. (1997). Following cooperative formation of secondary and tertiary structure in a single protein module. *J. Mol. Biol.* **270**, 99–110.
10. Killick, T. R., Freund, S. M. V. & Fersht, A. R. (1998). Real-time NMR studies on folding of mutants of barnase and chymotrypsin inhibitor 2. *FEBS Letters*, **423**, 110–112.
11. Li, A. & Daggett, V. (1994). Characterization of the transition state of protein unfolding by use of molecular dynamics: chymotrypsin inhibitor 2. *Proc. Natl Acad. Sci. USA*, **91**, 10430–10434.
12. Daggett, V., Li, A., Itzhaki, L. S., Otzen, D. E. & Fersht, A. R. (1996). Structure of the transition state for folding of a protein derived from experiment and simulation. *J. Mol. Biol.* **257**, 430–440.
13. Pan, Y. P. & Daggett, V. (2001). Direct comparison of experimental and calculated folding free energies for hydrophobic deletion mutants of chymotrypsin inhibitor 2: free energy perturbation calculations using transition and denatured states from molecular dynamics simulations of unfolding. *Biochemistry*, **40**, 2723–2731.
14. De Jong, D., Riley, R., Alonso, D. O. V. & Daggett, V. (2002). Probing the energy landscape of protein folding/unfolding transition states. *J. Mol. Biol.* **319**, 229–242.
15. Li, A. & Daggett, V. (1996). Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulations. *J. Mol. Biol.* **257**, 412–429.
16. Lazaridis, T. & Karplus, M. (1997). New view of protein folding reconciled with the old through multiple unfolding simulations. *Science*, **278**, 1928–1931.
17. Kazmirski, S. L., Li, A. & Daggett, V. (1999). Analysis methods for comparison of multiple mulecular dynamics trajectories: applications to protein

unfolding pathways and denatured ensembles. *J. Mol. Biol.* **290**, 283–304.

18. Kazmirski, S. L., Wong, K.-B., Freund, S. M. V., Tan, Y.-J., Fersht, A. R. & Daggett, V. (2001). Protein folding from a highly disordered denatured state: the folding pathway of chymotrypsin inhibitor 2 at atomic resolution. *Proc. Natl Acad. Sci. USA*, **98**, 4349–4354.

19. Levitt, M., Hirshberg, M., Sharon, R., Laidig, K. E. & 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. B*, **101**, 5051–5061.

20. Halle, B. & Wennerstrom, H. (1981). Interpretation of magnetic resonance data from water nuclei in heterogeneous systems. *J. Chem. Phys.* **75**, 1928.

21. Soper, A. K. (2000). The radial distribution functions of water and ice from 220 to 673 K and at pressures up to 400 MPa. *Chem. Phys.* **258**, 121–137.

22. Walser, R., Mark, A. E. & van Gusteren, W. F. (2000). On the temperature and pressure dependence of a range of properties of a type of water model commonly used in high-temperature protein unfolding simulations. *Biophys. J.* **78**, 2752–2760.

23. Beck, D. A. C., Alonso, D. O. V. & Daggett, V. (2002). A microscopic view of peptide and protein solvation. *Biophys. Chem.* In press.

24. Shaw, G. L., Davis, B., Keeler, J. & Fersht, A. R. (1995). Backbone dynamics of chymotrypsin inhibitor 2: effect of breaking the active site bond and its implications for the mechanism of inhibition of serine proteases. *Biochemistry*, **34**, 2225–2233.

25. Li, A. & Daggett, V. (1995). Investigation of the solution structure of chymotrypsin inhibitor 2 using molecular dynamics: comparison to X-ray crystallographic and NMR data. *Protein Eng.* **8**, 1117–1128.

26. Mayor, U., Johnson, C. M., Daggett, V. & Fersht, A. R. (2000). Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. *Proc. Natl Acad. Sci. USA*, **97**, 13518–13522.

27. Ferguson, N., Pires, J. R., Toepert, F., Johnson, C. M., Pan, Y. P., Volkmer-Engert, R. *et al.* (2001). Using flexible loop mimetics to extend phi-value analysis to secondary structure interactions. *Proc. Natl Acad. Sci. USA*, **98**, 13008–13013.

28. Levitt, M. (1990). *ENCAD–Energy Calculations and Dynamics*, Yeda, Rehovat, Israel.

29. 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. *Comput. Phys. Commun.* **91**, 215–231.

30. Harpaz, Y., Elmasry, N., Fersht, A. R. & Henrick, K. (1994). Direct observation of a better hydration at the N terminus of an α-helix with glycine rather than alanine as the N-cap residue. *Proc. Natl Acad. Sci. USA*, **91**, 3–15.

31. Kell, G. S. (1967). Precise representation of volume properties of water at one atmosphere. *J. Chem. Eng. Data*, **12**, 66–69.

32. Haar, L., Gallagher, J. S. & Kell, G. S. (1984). *NBS/NRC Steam Tables: Thermodynamic and Transport Properties and Computer Programs for Vapor and Liquid States of Water in SI Units*, Hemisphere Pub. Corp, Washington, DC.

33. Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. (1988). The Midas display system. *J. Mol. Graph.* **6**, 13–27.

34. Krynicki, K., Green, C. D. & Sawyer, D. W. (1978). Pressure and temperature dependence of self-diffusion in water. *Farad. Disc. Chem. Soc.* **66**, 199–208.

35. Soper, A. K. & Phillips, M. G. (1986). A new determination of the structure of water at 25 °C. *Chem. Phys.* **107**, 47–60.

36. Narten, A. H. & Levy, H. A. (1969). Observed diffraction pattern and proposed models of liquid water. *Science*, **165**, 447–454.

Edited by C. R. Matthews