

# Computational Approaches to Study Protein Unfolding: Hen Egg White Lysozyme as a Case Study

P.H. Hünenberger, A.E. Mark, and W.F. van Gunsteren

*Laboratorium für Physikalische Chemie, ETH-Zentrum, CH-8092 Zürich, Switzerland*

**ABSTRACT** Four methods are compared to drive the unfolding of a protein: (1) high temperature (T-run), (2) high pressure (P-run), (3) by imposing a gradual increase in the mean radius of the protein using a penalty function added to the physical interaction function (F-run, radial force driven unfolding), and (4) by weak coupling of the difference between the temperature of the radially outward moving atoms and the radially inward moving atoms to an external temperature bath (K-run, kinetic energy driven unfolding). The characteristic features of the four unfolding pathways are analyzed in order to detect distortions due to the size or the type of the applied perturbation, as well as the features that are common to all of them. Hen egg white lysozyme is used as a test system. The simulations are analyzed and compared to experimental data like  $^1\text{H}$ -NMR amide proton exchange-folding competition, heat capacity, and compressibility measurements.

© 1995 Wiley-Liss, Inc.

**Key words:** protein folding, computer simulation, molecular dynamics, high pressure, compressibility, heat capacity, amide proton exchange

## INTRODUCTION

### Study of Protein Folding and Unfolding by Computer Simulation

Elucidation of the mechanism of protein folding in terms of atomic interactions remains a major challenge in molecular biology, as despite the large body of experimental data on proteins in their native fold, the factors that determine the uniqueness of the native structure<sup>1–3</sup> and the nature of the process that leads from the potentially huge number of unfolded structures to this unique native state are poorly understood.<sup>4–12</sup> This is because, though the folded state of a protein has a well-defined structure, which can be determined in atomic detail, the multiple conformations that make up the partially or fully denatured states cannot be determined experimentally at atomic resolution using either X-ray diffraction or NMR spectroscopic techniques. In addition, though protein folding may occur along a limited number of highly cooperative pathways, the very low concentration of intermediate species makes detailed ex-

perimental characterization of folding intermediates virtually impossible.

In view of the difficulty of the experimental study of protein stability and folding, one may choose to turn to theoretical computational methods. Molecular dynamics (MD) simulation, for example, combines detailed information at the atomic level with high resolution in time, which allows for an analysis of structural and energetic changes that occur during the simulated process.<sup>13,14</sup> Because the structure of the unfolded state is not known, to study protein folding using computer simulation, one may start from the native fold and follow the process of unfolding. The implicit assumption is that the early stages of unfolding will approximate the later stages of protein folding. Unfolding simulations can also be used to select productive starting conformations for refolding studies.<sup>15</sup>

The study of protein unfolding by computer simulation is primarily limited by the accuracy of the atomic interaction function used to describe the protein and the time-scale that can be simulated. Using current supercomputers, the maximum period a protein in solution can be stimulated for is in the order of nanoseconds. This is about 6–9 orders of magnitude shorter than the time-scale of protein folding in vitro. Therefore, in practice, a large perturbation must be applied to the system to greatly accelerate the process of unfolding.

In vitro, proteins can be destabilized or denatured by changes in temperature, pressure,<sup>16,17</sup> or in the nature of the medium in which the protein is immersed (pH, ionic strength, solvent). Such physical perturbations can also be used to induce protein unfolding in computer simulations. Temperature-induced unfolding in aqueous solution has been studied for lysozyme,<sup>18</sup> apomyoglobin,<sup>19</sup> and bovine pancreatic trypsin inhibitor, BPTI.<sup>20,21</sup> Molecular dynamics studies of proteins at high pressure have been performed for BPTI,<sup>22,23</sup> though unfolding was not observed during these simulations. In computer simulation, nonphysical perturbations may also be used to induce the unfolding process. For example,

Received July 26, 1994; revision accepted November 14, 1994.

Address reprint requests to P.H. Hünenberger, Laboratorium für Physikalische Chemie, ETH-Zentrum, CH-8092 Zürich, Switzerland.

an unfolding pathway of BPTI in the gas phase has been obtained by forcing the protein to expand.<sup>15</sup>

Irrespective of the nature of the perturbation, in a computer simulation of unfolding, extreme conditions must be used to dramatically increase the rate of the unfolding process. Any method that is used to force a protein to unfold in less than a nanosecond is very likely to lead to a distorted pathway. Therefore, we have undertaken a detailed comparison of four different ways to induce unfolding in MD simulations: (1) high temperature (T-run), (2) high pressure (P-run), (3) by imposing a gradual increase in the mean radius of the protein (F-run, radial force driven unfolding), and (4) by coupling the difference between the temperature of the radially outward moving atoms and the radially inward moving atoms to an external temperature bath (K-run, kinetic energy driven unfolding).

The unfolding of hen egg white (HEW) lysozyme in aqueous solution was used as a test case. This 129 residue protein contains  $\alpha$ -helices and a triple-stranded  $\beta$ -sheet in two structural domains.<sup>24</sup> The crystal structure is known and a wealth of kinetic and thermodynamic data with respect to its folding is available to which results from the different unfolding schemes can be compared.

### Experimental Data on the Folding and Unfolding of Hen Egg White Lysozyme

The thermodynamics and kinetics of HEW lysozyme denaturant-induced unfolding under equilibrium conditions are consistent with a cooperative two state process.<sup>25,26</sup> Under nonequilibrium conditions, however,<sup>25,27–31</sup> a triphasic kinetics is observed: (1) Within a few milliseconds,<sup>30</sup> the protein acquires an almost fully native-like far ultraviolet (UV) circular dichroism (CD) spectrum (peptide region). However, no change in near UV CD ellipticity or protection of amide protons from exchange is observed. This indicates that though secondary structure is formed in this stage, there is neither a well-defined tertiary structure nor the formation of stable backbone hydrogen bonds. (2) In a second relatively fast ( $\tau \approx 10$ –20 ms) phase, an “overshooting”<sup>30</sup> of the far UV CD spectrum is observed. Near UV CD ellipticity begins to build up at this stage,<sup>30,31</sup> corresponding to the packing of the aromatic side chains. In this phase, the protection of amide protons in both the  $\alpha$ -helical and the  $\beta$ -sheet domain<sup>24,32,33</sup> follows the same kinetics. The amplitude of the former ( $40 \pm 12\%$ ) is, however, greater than that of the latter ( $24 \pm 6\%$ ). (3) In the final phase, asynchronous protection of protons in the two domains is observed (50–100 ms  $\alpha$ -domain, 150–800 ms  $\beta$ -domain). Both CD spectra reach their native form in this phase ( $\tau \approx 300$  ms). <sup>1</sup>H-NMR exchange-folding competition measurements<sup>34</sup> clearly indicate that the backbone amide protons of the  $\alpha$ -domain of lysozyme are protected much earlier than

those of the  $\beta$ -domain, which has been interpreted to mean that the  $\alpha$ -domain folds first.

Denaturation of HEW lysozyme at high pressure was studied by high resolution <sup>1</sup>H-NMR spectroscopy.<sup>35</sup> The evolution of the local environment of five residues was monitored upon denaturation at pressures ranging from 1 bar to 5 kbar, at  $T = 68.5^\circ\text{C}$  and pH 3.9. The maximum degree of denaturation observed was about 53% and the results were compatible with a two state model of folding. A previous fluorescence study over a range of pressure up to 11 kbar<sup>36</sup> indicated, however, a plurality of denatured forms of the protein.

## METHODS

### Molecular Model and Computational Procedures

The simulations were performed using the same interatomic interaction function and simulation parameters as in earlier MD simulation of lysozyme in aqueous solution at  $T = 300$  K<sup>37,38</sup> and at  $T = 500$  K.<sup>18</sup> Nonpolar hydrogens attached to carbon atoms were incorporated into the latter, forming united atoms. Polar hydrogen atoms were explicitly treated. The four disulfide bonds were not reduced. Atomic charges were chosen appropriate to pH 7.0, resulting in a total charge of  $+8e$  for the protein. His-15 was taken as protonated at the  $N_\epsilon$  nitrogen based on hydrogen bonding considerations with respect to its environment in the crystal structure. The GROMOS87 force field<sup>39</sup> was used for all the simulations.

The crystal structure of triclinic hen egg white lysozyme,<sup>40</sup> entry 2LZT of the Brookhaven Protein Data Bank,<sup>41</sup> served as starting structure. Water was explicitly included in all simulations in order to correctly account for the interplay between protein–protein and protein–solvent interactions. The protein was placed at the center of a rectangular periodic box of dimensions  $x = 4.88$  nm,  $y = 5.31$  nm, and  $z = 6.81$  nm (no protein atoms lying within less than 0.95 nm distance from the wall), containing 5345 simple point charge (SPC) water molecules.<sup>42</sup> The total system contained 17,299 atoms. Bond lengths were constrained by application of the SHAKE procedure<sup>43</sup> with a relative tolerance of  $10^{-4}$ . Nonbonded interactions were handled using a twin-range method,<sup>44</sup> with a short and long cut-off radius of 0.8 and 1.4 nm, respectively. The nonbonded pair list was updated every 5 fs (P-run, F-run, K-run) or 10 fs (equilibration, T-run). A time step of 1 fs was used for all simulations, except during the 300 K equilibration where it was equal to 2 fs. The temperature was maintained by weakly coupling the protein and the water separately to an external temperature bath<sup>45</sup> with a coupling constant  $\tau_T = 0.1$  ps. Except for the P-run, no pressure maintaining scheme was applied. The four simulations were all branched from a 50 ps equilibration run at 300 K. The coordinates were saved every 0.05 ps for

analysis. Secondary structure assignments were based on the DSSP program<sup>46</sup> and checked for consistency with the program Ramplus.<sup>47</sup> The protocol for the four different unfolding runs is described below.

### Temperature-Induced Unfolding: T-Run

After the 50 ps equilibration at 300 K, the temperature of the heat bath was switched to 500 K and the dynamics was followed for a further 180 ps. This simulation has already been briefly described previously<sup>18</sup> and is presented here primarily for comparison to the alternative unfolding schemes.

### Pressure-Induced Unfolding: P-Run

For compatibility with the available experimental data<sup>35</sup> the temperature was held at 341.7 K. A pressure of 10 kbar, instead of 5 kbar, was chosen however to increase the rate and degree of denaturation. This pressure was maintained for 210 ps using an isotropic coordinate scaling<sup>45</sup> with a coupling constant  $\tau_P = 0.5$  ps. An isothermal compressibility of  $45.63 \times 10^{-6} \text{ atm}^{-1}$  was used<sup>48</sup> for water. For compatibility with the other unfolding simulations, the protonation state of the amino acids was not adapted to the experimental low pH conditions<sup>35</sup> (effective  $\text{pH} \leq 3.9$ ). It is nevertheless likely that the protonated state of the Asp, Glu, and His residues ( $\text{pK}_a \approx 3.9, 4.3, 6.0$ ) would dominate under such conditions.

### Constant Radial Force-Induced Unfolding: F-Run

A penalty term was added to the standard GRO-MOS87 force field potential energy function  $V_{\text{phys}}$

$$V(\{\mathbf{r}_i\}) = V_{\text{phys}}(\{\mathbf{r}_i\}) - k_R R(\{\mathbf{r}_i\}) \quad (1)$$

where  $k_R$  is the (positive) force constant associated to the penalty function and  $R(\{\mathbf{r}_i\})$  is the mean radius of the protein configuration  $\{\mathbf{r}_i\}$ , defined by

$$R(\{\mathbf{r}_i\}) = \frac{1}{N} \sum_{i=1}^N \|\mathbf{r}_i - \mathbf{r}_{\text{CM}}\|. \quad (2)$$

Here,  $\mathbf{r}_i$  is the position vector of atom  $i$  and  $\mathbf{r}_{\text{CM}}$  that of the center of mass of the protein consisting of  $N$  atoms. Taking the negative of the derivative of expression (1) with respect to the coordinates of particle  $i$  yields the force

$$\mathbf{F}_i = -\frac{\partial V}{\partial \mathbf{r}_i} = \mathbf{F}_{\text{phys},i} + \frac{k_R}{N} \{\mathbf{e}_{i,\text{CM}} - \frac{m_i}{M} \sum_{j=1}^N \mathbf{e}_{j,\text{CM}}\} \quad (3)$$

where  $\mathbf{F}_{\text{phys},i}$  is the standard physical force field term,  $m_i$  is the mass of atom  $i$ ,  $M$  the total mass of the protein, and

$$\mathbf{e}_{i,\text{CM}} = \frac{\mathbf{r}_i - \mathbf{r}_{\text{CM}}}{\|\mathbf{r}_i - \mathbf{r}_{\text{CM}}\|}. \quad (4)$$

Except for a (small) center of mass correction [i.e.,

the second term in the braces in Eq. (3)], due to the use of a mean radius definition that does not involve atomic masses, the added force is basically a radial force of constant size for all atoms at any time and for any configuration. The use of the radius of gyration, instead of the mean radius in expression (1) would lead to a force directly proportional to the distance between the atom and the center of mass. Such a force would bias the unfolding in favor of residues at the surface of the protein. The linear function in (1) has been chosen as a simple way to uniformly drive the unfolding process. An alternative is to restrain the mean radius to a value which is slowly increased during the simulation. Such a driving force results in the same type of behavior as induced by the linear potential energy term in (1) (results not shown).

### Kinetic Energy Gradient Driven Unfolding: K-Run

The method applied to drive unfolding by maintaining a constant uniform expansion kinetic energy was inspired by a method proposed by Scheraga.<sup>15</sup> It differs from the latter in two basic aspects: first, instead of an energy minimization technique we use an MD technique; second, our method explicitly includes solvent, while in the work cited above, vacuum simulations were performed. The basic idea is to define two contributions to the atomic velocities at a given time: one corresponding to a uniform expansion mode and another to a uniform contraction mode. A fixed kinetic energy difference (i.e., a temperature difference,  $\Delta\theta$ ) between these modes is then maintained by means of a first-order weak coupling to an “unfolding thermostat.” The method is essentially a velocity rescaling scheme. Inward moving atoms are slowed whereas outward moving atoms are accelerated. This is in contrast to the constant radial force induced unfolding technique, in which a term depending solely on the atomic mass (radial acceleration) is added every time step to each atomic velocity. The algorithm consists of 8 steps that are to be incorporated in the leap-frog algorithm for the integration of Newton’s equations of motion and is described in detail in Appendix A.

## RESULTS

### Temperature-Induced Unfolding: T-Run

The behavior of the protein when the temperature is raised to 500 K has already been briefly described.<sup>18</sup> The radius of gyration (Fig. 1a) increases from a value of 1.45 nm in the equilibrated structure almost steadily during 70 ps. Then it stabilizes for 40 ps at 1.58 nm (plateau A), before rising again, this time more rapidly. Two additional smaller plateaus are visible, plateau B (1.69 nm) between 134 and 150 ps and plateau C (1.77 nm) between 156 and 174 ps. Analysis of the behavior of the two structural domains of lysozyme<sup>32,33</sup> shows that up to 70

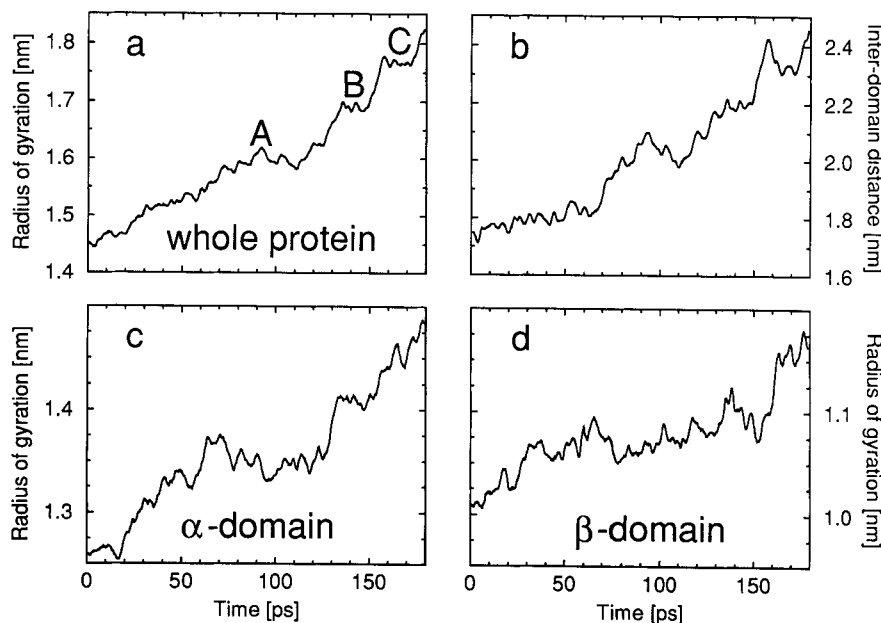


Fig. 1. (a) Radius of gyration, (b) distance between the centers of mass of the  $\alpha$ -domain and  $\beta$ -domain, (c) radius of gyration of the  $\alpha$ -domain, and (d) radius of gyration of the  $\beta$ -domain, as a

function of time for the T-run. The  $\alpha$ -domain comprises<sup>33</sup> residues 1 to 39 and 88 to 129, whereas the  $\beta$ -domain comprises residues 40 to 87.

ps, the increase in the radius of gyration reflects the expansion of the individual domains (Fig. 1c and d). From this time point on, the interdomain distance begins to increase (Fig. 1b). However, a simultaneous contraction of the  $\alpha$ -domain (Fig. 1c) counteracts the molecular expansion, leading to plateau A. After 110 ps, the main contribution to the radius of gyration is the separation of the two domains. Both structural domains do however undergo substantial expansion during the whole simulation.

The root mean square (rms) atomic positional deviations from the X-ray structure calculated using only  $C_{\alpha}$  atoms or all atoms (Table I) start from 0.12 and 0.20 nm and tend asymptotically to values of 0.39 and 0.51 nm respectively, at 110 ps, before rising almost linearly thereafter (see<sup>18</sup> Fig. 1b). The all atom deviation curve markedly diverges from the  $C_{\alpha}$  one along plateau A, indicating that during this period, extensive side chain rearrangement occurs, accompanied by a relatively small backbone motion.

The solvent accessible surface area<sup>49</sup> (SASA) increases rapidly from 64.8 nm<sup>2</sup> (Table I) to about 76 nm<sup>2</sup> during the first 25 ps. Of this increase 75% is due to exposure of apolar surface area. This is followed by a quasilinear evolution. After 180 ps, the SASA reaches 102.7 nm<sup>2</sup>. The initial rapid increase in the solvent accessibility of the apolar surface is likely to be in part an artifact. Since these simulations were performed, it has been shown that the GROMOS87 force field overestimates the van der Waals interaction between neutral carbon groups and SPC water.<sup>50</sup> Based on the polar and apolar ex-

posed surface area, the heat capacity  $C_p$  of the conformations corresponding to plateau A has been estimated to be 0.8 kcal mol<sup>-1</sup> K<sup>-1</sup> below that of a model of the unfolded state.<sup>51</sup> Using the same formalism relating  $C_p$  to the exposed surface area<sup>52</sup> we find that these conformations and the end state of the simulation have a  $\Delta C_p$  value of 0.65 and 1.00 kcal mol<sup>-1</sup> K<sup>-1</sup>, respectively, compared to the X-ray structure (Table I). Experimentally<sup>53</sup> the  $\Delta C_p$  between the folded and unfolded state is 1.6  $\pm$  0.1 kcal mol<sup>-1</sup> K<sup>-1</sup>. Thus, the end state of the simulation, despite an almost total loss of secondary structure, is not completely unfolded with respect to apolar side chain exposure.

The time evolution of secondary structure (Fig. 2) gives further insight into the unfolding process. Helix D (residues 108–115) and the C-terminal 3<sup>10</sup> helix (residues 120–125) are quite distorted and labile. Rapid interchange of  $i$  to  $i+3$ ,  $i$  to  $i+4$ , and  $i$  to  $i+5$  hydrogen bonds on a picosecond time-scale is observed. Both helices progressively shrink and become less stable throughout the simulation. Analogous fluctuations are seen for the 3<sup>10</sup> helix of the  $\beta$ -domain (residues 80–84). However, this helix is more stable, surviving in the form of an  $\alpha$ -helix until the end of the simulation. The fluctuating nature of the above three helices is not surprising. They are short, distorted, and irregular helical elements and are the least protected elements of secondary structure in H/D NMR exchange experiments.<sup>54</sup> By contrast, the long amphipathic helix C (residues 88–99) undergoes no major fluctuations during the simula-

TABLE I. Selected Properties Monitored Along the Four Unfolding Simulations\*

	Equil.		T-run		P-run	F-run	K-run
Time point (ps)	0	70	110	180	210	210	180
$R_{\text{gyr}}$ (nm)	1.453	1.581	1.586	1.834	1.454	2.081	1.883
Mean radius (nm)	1.383	1.494	1.511	1.758	1.376	1.959	1.781
Principal axes radii (nm)	0.660	0.716	0.718	0.895	0.747	0.694	0.797
	0.833	0.900	0.986	1.016	0.748	1.318	1.036
	0.991	1.084	1.014	1.200	0.999	1.454	1.356
$R$ ( $\alpha$ - $\beta$ ) (nm)	1.740	1.802	1.978	2.485	1.759	3.324	2.789
$R_{\text{gyr}}$ ( $\alpha$ ) (nm)	1.260	1.382	1.354	1.485	1.245	1.434	1.403
$R_{\text{gyr}}$ ( $\beta$ ) (nm)	1.017	1.084	1.078	1.163	1.026	1.074	1.106
$E_{\text{pot}}$ total ( $10^6$ J mol $^{-1}$ )	-240.3	-194.1	-194.7	-193.6	-283.5	-240.8	-240.5
$E_{\text{pot}}$ solv.-solv. ( $10^6$ J mol $^{-1}$ )	-215.4	-171.4	-171.6	-170.7	-211.7	-213.7	-213.7
$E_{\text{pot}}$ prot.-solv. ( $10^6$ J mol $^{-1}$ )	-16.4	-18.5	-18.7	-19.3	-19.6	-20.5	-20.8
$E_{\text{pot}}$ prot.-prot. ( $10^6$ J mol $^{-1}$ )	-12.0	-10.3	-9.4	-7.8	-10.6	-10.1	-9.4
All atom rms (nm)	0.196	0.441	0.513	0.836	0.301	0.982	0.791
$C_{\alpha}$ rms (nm)	0.124	0.357	0.385	0.724	0.202	0.930	0.689
SASA polar (nm $^2$ )	20.0	24.2	26.6	29.6	20.1	26.8	27.6
SASA apolar (nm $^2$ )	44.8	61.6	62.6	73.1	45.1	62.9	62.4
SASA total (nm $^2$ )	64.8	85.8	89.2	102.7	65.2	89.7	90.0
$\Delta C_p$ (kcal mol $^{-1}$ K $^{-1}$ )	0	0.65	0.63	1.03	0.01	0.74	0.70

\*Equil., equilibrated (50 ps, 300 K) starting structure; T-run (three time points), temperature (500 K) driven unfolding; P-run, pressure-driven unfolding; F-run, constant radial force-driven unfolding; K-run, kinetic energy gradient-driven unfolding;  $R_{\text{gyr}}$ , radius of gyration; mean radius, as defined in Eq. (2); principal axes radii, as defined in Eq. (B.8);  $R(\alpha\text{-}\beta)$ ,  $\alpha$ - $\beta$  interdomain distance (center of mass to center of mass); domain  $\alpha$  is defined<sup>33</sup> as residues 1–39 and 88–129 and domain  $\beta$  as 40–87;  $R_{\text{gyr}}(\alpha)$  and  $R_{\text{gyr}}(\beta)$ , radii of gyration of the two individual domains;  $E_{\text{pot}}$  total, solv.-solv., prot.-solv., and prot.-prot., total potential energy (all force-field terms), then solvent-solvent, protein-solvent, and protein-protein Coulomb plus van der Waals potential energy, all averaged over 4 ps; all atom rms and  $C_{\alpha}$  rms, root mean square atomic positional deviations from the X-ray structure<sup>40</sup> for all atoms and  $C_{\alpha}$  atoms; SASA polar, apolar, and total, polar component, apolar component, and total of the solvent accessible surface area calculated using Connolly's program based on the Lee and Richards algorithm,<sup>49</sup> using a probe radius of 1.4 Å and a density of about 5 dots/Å $^2$ ;  $\Delta C_p$ , heat capacity changes relative to the equilibrated starting structure determined using an increment of 0.45 cal K $^{-1}$  mol $^{-1}$  Å $^{-2}$  for apolar and -0.26 cal K $^{-1}$  mol $^{-1}$  Å $^{-2}$  for polar surface area.<sup>52</sup>

tion, but slowly shrinks from both termini before disappearing. This shrinking from both ends corresponds with the observed lower protection of the helix termini in H/D exchange experiments. Most interesting is the behavior of helices A and B. During the equilibration period, helix B (residues 24–36) shrinks to residues 25 to 29. Up to 60 ps, helix A (residues 4–15) shrinks slowly. This is followed, however, by the reappearance from 60 to 85 ps of two ordered A and B helices, similar to those observed in X-ray structure. Both helices then shrink and shift away from each other, forming a large loop of residues 12–26. The triple-stranded  $\beta$ -sheet stays close to the native conformation up to 120 ps, at which point contact with the third strand (residues 58–59) is lost. It then progressively reduces to a single hydrogen bonded bridge between residues 44 and 52. Of the three initial  $\beta$ -bridges, bridge 3–39 disappears after 10 ps, bridge 20–23 after 90 ps, while bridge 65–79 remains stable until the end. This is interesting, because Trp-63, Cys-64, and Ile-78 were found in NMR H/D exchange experiments to be much more protected than the average  $\beta$ -domain residues during refolding<sup>31,34</sup> and to maintain some protection in the unfolded state.<sup>54</sup>

The global path of unfolding is illustrated in Figure 3. Figure 3a shows the equilibrated starting

structure. Figure 3b and c shows intermediate structures corresponding to average structures over <110–120 ps> and <170–180 ps>, respectively. In addition to the progressive loss of secondary structure, it can be seen in Figure 3b that the unwinding of helices A and B convert the  $\alpha$ -domain hairpin turn into a large loop, which moves toward the unwound helix D. The packing of this loop against the protein is probably responsible for the decrease in the radius of gyration of the  $\alpha$ -domain during plateau A. In Figure 3c, the complete disruption of helix C and the conversion of the  $\beta$ -sheet to a double elongated  $\beta$ -strand can be seen.

In Figure 4a (equilibrated starting structure), 4b (<110–120 ps>), and 4c (<170–180 ps>), the van der Waals surface of hydrophobic residues is shown. The colors have been chosen in order to emphasize the splitting of the hydrophobic core. The bulk of the core, viz. Leu-17, Tyr-20, Tyr-23, Trp-28, Ile-98, Met-105, Trp-108, and Trp-111, as well as some other smaller hydrophobic residues are colored in yellow. Trp-123 and residues clustering around it are colored in blue. Phe-3, His-15, Phe-38, and other residues in the N-terminal region of the protein are colored in green. With the exception of Phe-3, all these residues belong to the  $\alpha$ -domain. The hydrophobic residues of the  $\beta$ -domain and those at the

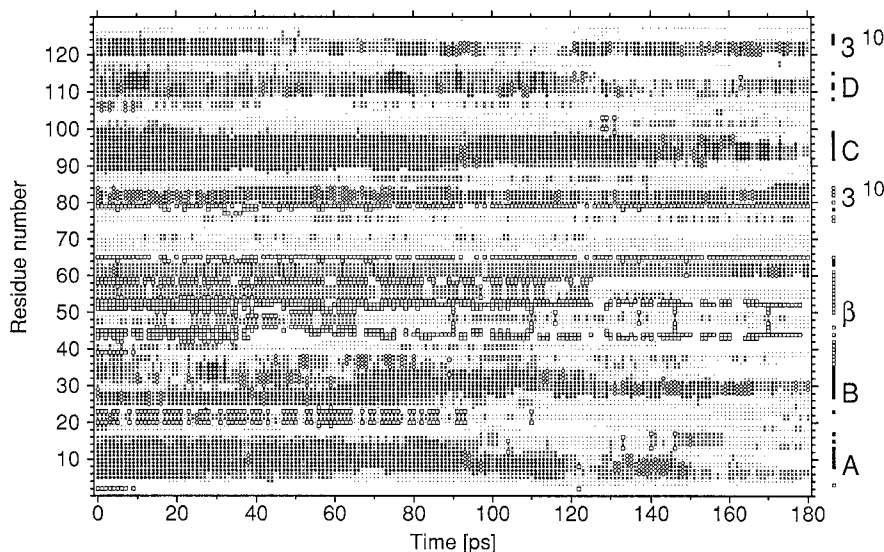


Fig. 2. Secondary structure as a function of time, for the T-run, as given by the "SUMMARY" entry of DSSP program output.<sup>46</sup> (●)  $\alpha$ -helix, (□)  $\beta$ -bridge or  $\beta$ -sheet, (◇)  $3^{10}$  helix, (◆)  $\pi$  helix, (×) hydrogen-bonded turn, and (–) bend. On the right hand side, a

symbol indicates location of helices A, B, C, and D, the two  $3^{10}$  helices and the  $\beta$ -sheet. On this side, (□) indicates observed but rapidly exchanging protons upon folding<sup>34</sup> and (■) slowly exchanging ones.

beginning of the helix C (88–95) are colored in violet. The hydrophobic core splits into four hydrophobic clusters between which water penetrates. As mentioned earlier, however, the force field used in these simulations has since been shown to overestimate the interaction between neutral carbons and water.<sup>50</sup> This results in a higher than expected exposure to water of some hydrophobic residues (e.g., Trp-101 and Met-105) and means that details in regard to the precise mechanism of unfolding illustrated in Figure 4 should be interpreted with caution.

### Pressure-Induced Unfolding: P-Run

The pressure was increased to a value of 10 kbar in 1.5 ps and stayed at this value with an rms deviation of 190 bar (1.9%) during the rest of the P-run. The box volume decreased asymptotically from 177 to 146 nm<sup>3</sup> within 55 ps and remained stable thereafter. As can be seen in Figure 5a, no net unfolding is achieved within 210 ps. The radius of gyration shrinks very rapidly to a value of 1.42 nm, stays at this value for approximately 60 ps, and then slowly increases toward the starting value. The evolution of the radius of gyration of the molecule as a whole strongly parallels that of the  $\alpha$ -domain to  $\beta$ -domain distance (Fig. 5b). The main difference is that the radius of gyration approaches its starting value after 210 ps whereas the interdomain distance remains slightly larger than in the equilibrated starting structure, indicating that parts of the protein have contracted. Comparing the radii of gyration of the individual domains (Fig. 5c and d), the  $\beta$ -do-

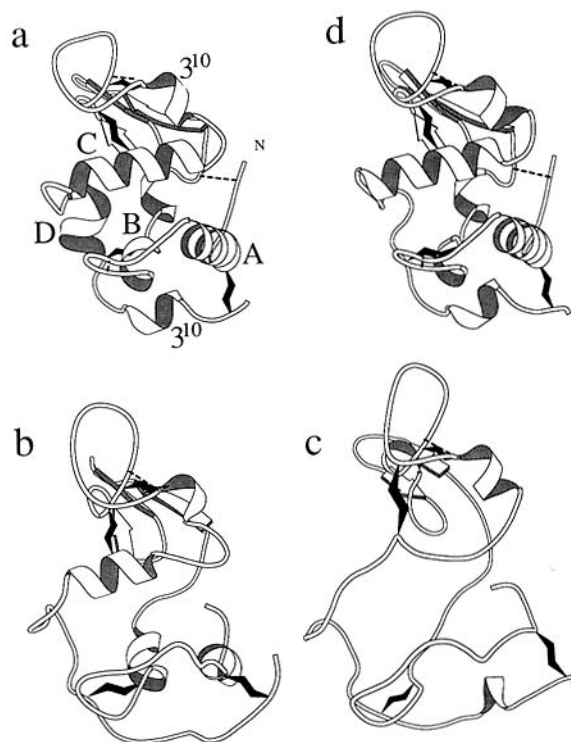
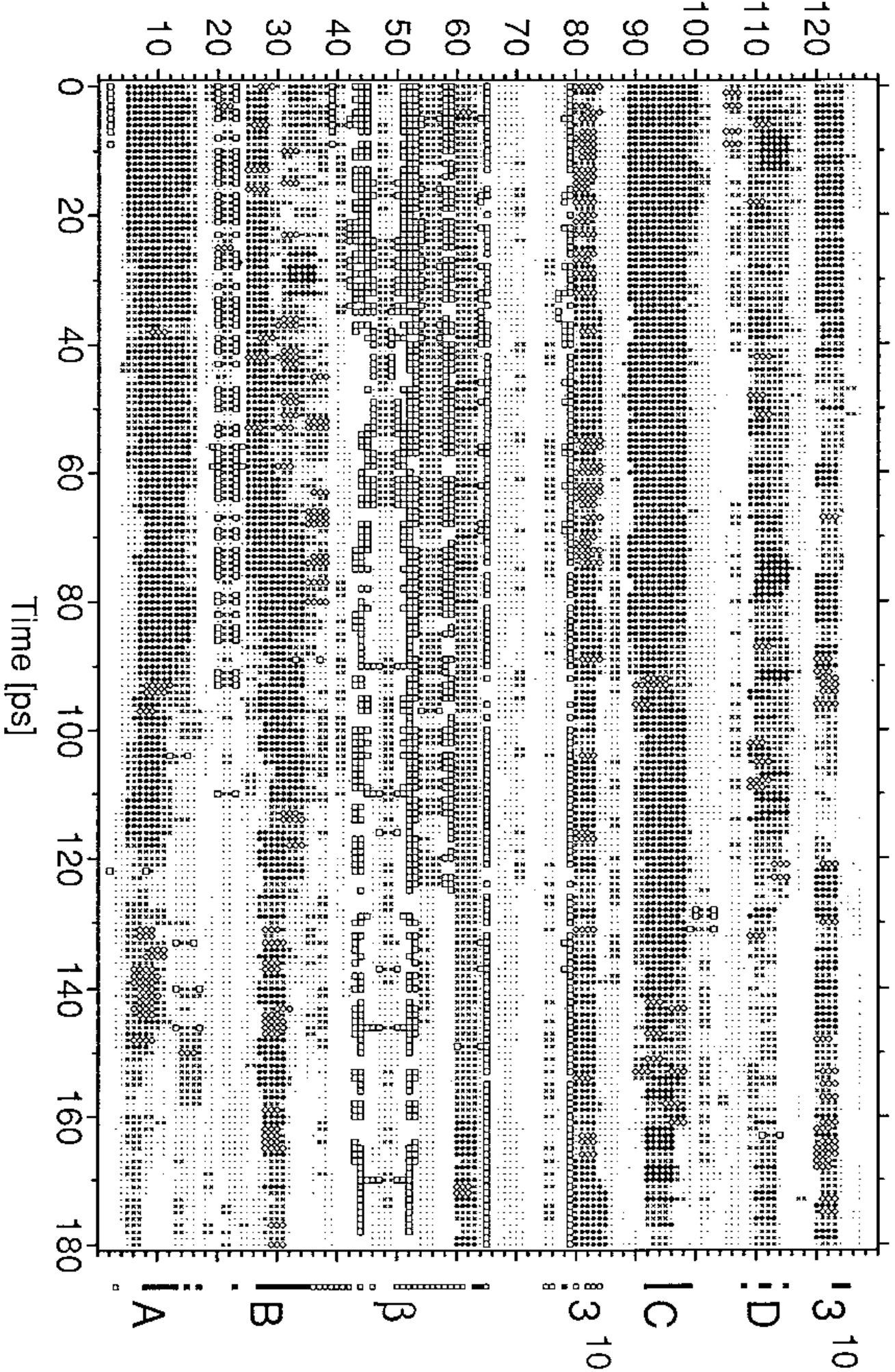


Fig. 3. Ribbon models for the different structures (a) averaged over the last 10 ps of the 300 K equilibration run, (b) T-run <110–120 ps>, (c) T-run <170–180 ps>, (d) P-run <180–210 ps>, drawn according to Richardson's representation. Dashed lines represent  $\beta$ -bridges. The ribbon structures have been drawn using the program Molscript<sup>60</sup> according to DSSP secondary structure assignment.

Residue number





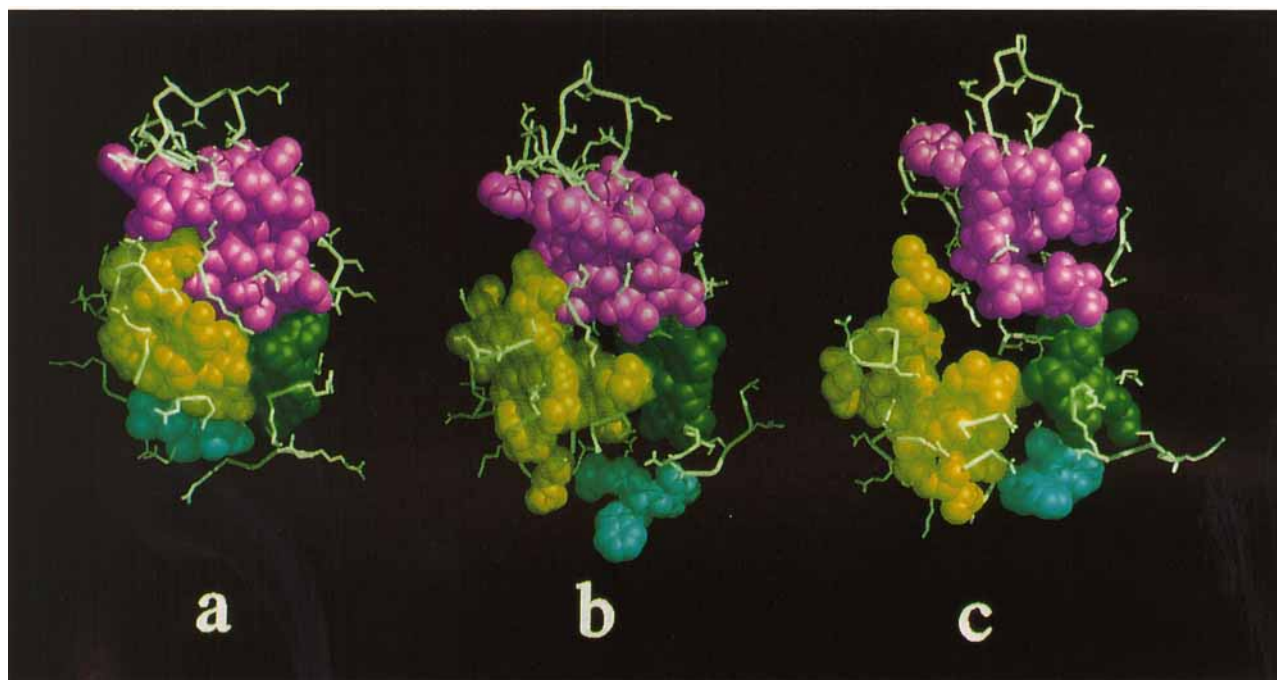


Fig. 4. van der Waals surface of hydrophobic residues for (a) the equilibrated starting structure, (b) T-run <110–120 ps>, and (c) T-run <170–180 ps>. The surface is drawn for all Ala, Val, Leu, Ile, Pro, Trp, Phe, Tyr, His, Cys, and Met except Pro-70, Cys-127, and Leu-129 because they do not pack closely with the

body of the other mentioned residues. Color differentiates residues according mostly to sequence number. Green, hydrophobic residues between 2–15 and 38; yellow, 17–34 and 98–115; violet, 42–95; and blue, 120–124. The plot was drawn using UCSF Midas.<sup>61</sup>

main, after the initial very rapid compression due to the sudden increase of the pressure, slowly reexpands, reaching a maximum radius after 100 ps. In contrast, the  $\alpha$ -domain continues to contract for the first 60 ps. Unlike the  $\beta$ -domain, the  $\alpha$ -domain does not regain its initial radius during the simulation. This is in agreement with results from an X-ray diffraction study of lysozyme crystals at a hydrostatic pressure of 1 kbar<sup>65</sup> which showed a differential compressibility between the two domains, the  $\beta$ -domain being essentially incompressible while the  $\alpha$ -domain and the interdomain hinge-region contracted under pressure.

The solvent accessible surface area (Table I) decreases slightly (about 5 nm<sup>2</sup>) during the first few picoseconds, then increases almost linearly (slope 0.03 nm<sup>2</sup>/ps), reaching a value slightly above the starting value by the end of the simulation. This indicates that the system is not in equilibrium and unfolding may be beginning to occur, despite little overall change in the radius of gyration. However, assuming that this slow linear rate of change of accessible surface area is maintained, full unfolding ( $\Delta C_p$  equal to the experimental value) would require 2 ns of simulation at 10 kbar. This is supported by the behavior of the total potential energy (Table I) which still slowly decreases at the end of the simulation. The end conformation of the P-run (Fig. 3d)

shows that, despite slight motions of the loop, of the hairpin turn, the  $\beta$ -domain <sup>310</sup> helix, and the loop connecting C and D helices, the global shape has changed little compared to the equilibrated starting structure (Fig. 3a).

#### Constant Radial Force-Induced Unfolding: F-Run

The behavior of the system when submitted to the penalty function defined in Eqs. (1) and (2) is very dependent on the force constant  $k_R$  that is chosen. As shown in Figure 6a, if the force constant  $k_R$  is chosen equal to 5 kJ mol<sup>-1</sup> nm<sup>-1</sup> (—), a very rapid expansion leads to full unfolding within a few tens of picoseconds. Choosing a smaller force constant like 3 kJ mol<sup>-1</sup> nm<sup>-1</sup> (---) affects only the initial phase of the expansion and the curves join each other after about 10 ps. If  $k_R$  is decreased to 2 kJ mol<sup>-1</sup> nm<sup>-1</sup> (—), a biphasic behavior is observed. A slow expansion phase (about 0.088 nm within 90 ps) is followed by a rapid phase (rate 0.0047 nm/ps). Only this last run (F-run) will be discussed in more detail below. The observation that the fast phase is essentially linear in water suggests that the main effect of the solvent is to exert friction, limiting the rate of expansion. The increase in the radius of gyration of the whole molecule (Fig. 6a) primarily reflects an increase in the distance between the centers of mass of



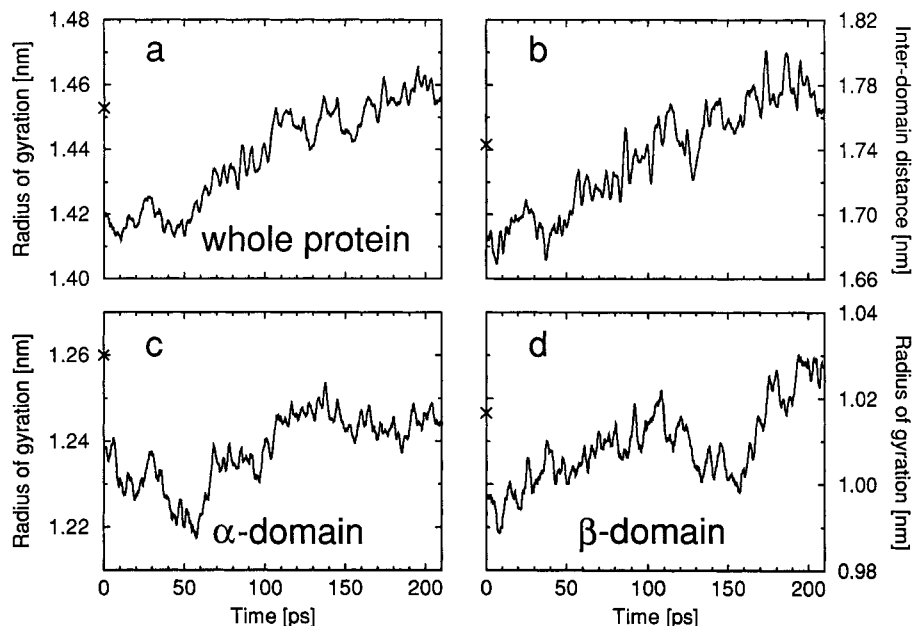


Fig. 5. (a) Radius of gyration, (b) distance between the centers of mass of the  $\alpha$ -domain and  $\beta$ -domain, (c) radius of gyration of the  $\alpha$ -domain, and (d) radius of gyration of the  $\beta$ -domain, as a function of time for the P-run. See also Figure 1. ( $\times$ ) indicates initial values.

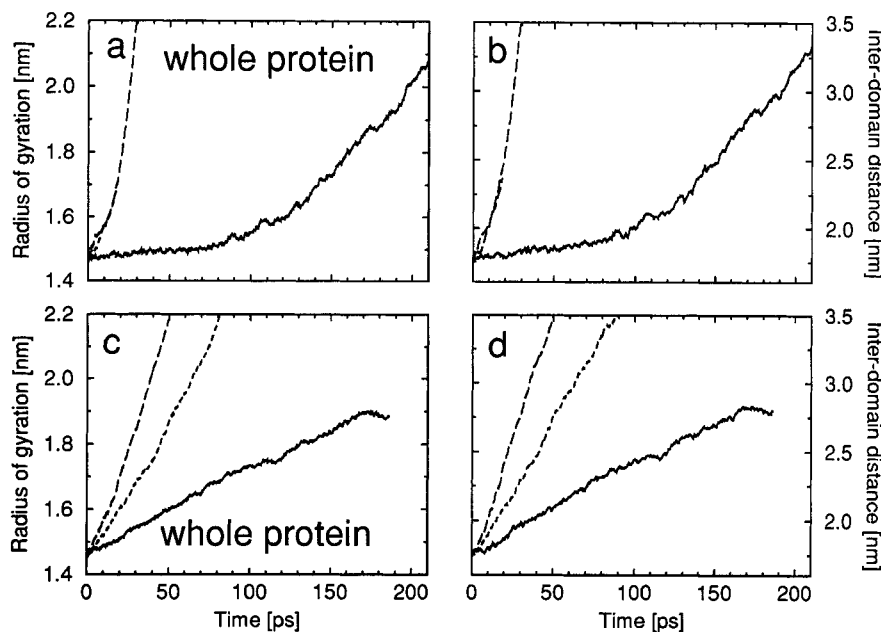


Fig. 6. (a) Radius of gyration, (b) distance between the centers of mass of the  $\alpha$ -domain and  $\beta$ -domain as a function of time, for the F-run; (—)  $k_R = 2 \text{ kJ mol}^{-1} \text{ nm}^{-1}$  (210 ps); (---)  $k_R = 3 \text{ kJ mol}^{-1} \text{ nm}^{-1}$  (20 ps); (---)  $k_R = 5 \text{ kJ mol}^{-1} \text{ nm}^{-1}$  (60 ps).

(c) Radius of gyration, (d) distance between the centers of mass of the  $\alpha$ -domain and  $\beta$ -domain as a function of time, for the K-run; (—)  $\Delta\theta = 20 \text{ K}$  (190 ps); (---)  $\Delta\theta = 30 \text{ K}$  (100 ps); (---)  $\Delta\theta = 40 \text{ K}$  (60 ps).

the  $\alpha$ - and  $\beta$ -domains (Fig. 6b). The radius of gyration of the  $\alpha$ -domain also increases significantly whereas the  $\beta$ -domain in contrast hardly expands (Table I).

The protein appears at first to remain trapped in a low energy state and is searching the local conformational space with relatively small structural

changes, before a "break point" occurs, characterized by disruption of short-range van der Waals interactions close to the interdomain region and intrusion of solvent molecules between the two domains. At this point, the relatively large radial force drives the protein to expand. This is illustrated in Figure

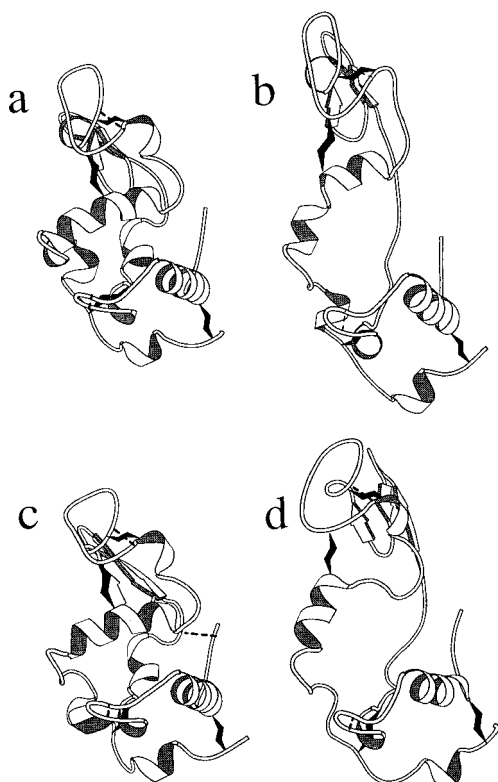


Fig. 7. Ribbon models for the different structures (a) F-run <110–120 ps>, (b) F-run <200–210 ps>, (c) K-run <40–50 ps>, and (d) K-run <170–180 ps>, according to Richardson's representation. Dashed lines represent  $\beta$ -bridges. See also Figure 3.

7a and b. At the end of the slow phase (Fig. 7a), the major backbone distortions involve primarily the  $\alpha$ -domain. These mainly consist of a bending and shifting away of helix C from the hydrophobic core. After the rapid phase (Fig. 7b), the  $\beta$ -domain  $3^{10}$  helix has disappeared, helix B is almost completely unrolled and helix C and D are largely elongated. These extended regions connect two relatively intact portions of the protein.

Estimates of the heat capacity value based on exposed surface area (Table I) indicate that the protein is far from unfolded. The calculated value for  $\Delta C_p$  upon unfolding of  $0.74 \text{ kcal mol}^{-1} \text{ K}^{-1}$  is well below the experimental value of  $1.6 \text{ kcal mol}^{-1} \text{ K}^{-1}$ . Interestingly, the solvent accessible surface area properties of the final conformation reached in this run are very similar to those of the metastable intermediate observed in the T-run. They correspond, however, to very different structures. Thus, the calculated  $\Delta C_p$  value is not a very stringent criterium for judging the validity of an unfolding model.

The breakup of the hydrophobic core of lysozyme in the initial phase of the run is illustrated in Figure 8a. (<110–120 ps>) An opening of the active site cleft is observed, though not exactly along the pro-

posed hinge bending region,<sup>33</sup> the N-terminal end of the C-helix moving together with the  $\beta$ -domain. Although the hydrophobic packing remains tighter in this case than in the T-run (Fig. 4b), in both cases a hole near the center of mass of the molecule appears, which allows water molecules to penetrate.

### Kinetic Energy Gradient Driven Unfolding: K-Run

The uniform expansion kinetic energy driven unfolding gives a much better control of the evolution of unfolding than the constant radial force method (F-run). Three temperature differences  $\Delta\theta$  were used for the "unfolding thermostat" [Eq. (A.17)]. In all cases, the relaxation time  $\tau_\theta$ , which controls the strength of the coupling to the temperature difference bath, was set to 0.02 ps. As can be seen from the radius of gyration evolution (Fig. 6c), the expansion is quasilinear for  $\Delta\theta = 20 \text{ K}$  (—, slope 0.0024 nm/ps), 30 K (- - -, slope 0.0095 nm/ps), or 40 K (- - -, slope 0.0150 nm/ps). Uniform behavior is not unexpected, as the velocity rescaling scheme was designed to conserve expansion rates. Here again, the radius of gyration evolution mostly reflects the increase in the interdomain distance (Fig. 6d). Only the run with  $\Delta\theta = 20 \text{ K}$  (K-run) will be discussed in more detail below.

The overall temperature stabilized at 302.5 K ( $\pm 1.7 \text{ K}$ ). The expansion and contraction temperatures were 344.1 K and 333.2 K with rms deviations of 15.4 and 15.0 K respectively. The average temperature difference could not be maintained at 20 K, which is probably due to the fact that the unfolding thermostat drives an energetically unfavorable process. The temperature difference is only held positive on average with high amplitude variations. Expansion occurs only "on average" as is reflected in the deformation rates of Equation (A.1); all three fluctuate highly between positive and negative values (of the order of  $\pm 0.2 \text{ ps}^{-1}$  with an average of about  $0.002 \text{ ps}^{-1}$ ). The lambda factors of Eq. (A.7) are effectively close to one. Each of their six average values is in the range  $1.000 \pm 0.012$ . The standard deviations along the run are of the order of 3.5%.

Apolar solvent accessible surface area jumps by  $5 \text{ nm}^2$  during the first 5 ps, then increases steadily. Polar surface area increases almost linearly from the beginning. The estimated change of the heat capacity  $\Delta C_p$  for the end structure was  $0.7 \text{ kcal mol}^{-1} \text{ K}^{-1}$ , less than half of the experimental value. Although the radius of gyration of this structure is less than that of the end structure of the F-run, the estimated heat capacity is very similar. Possibly, the smoother unfolding pathway used here allows for better equilibration of water around the protein, resulting in a better energetic balance and a better selectivity in hydrophilic/hydrophobic atom exposure.

Ribbon models display some analogies between

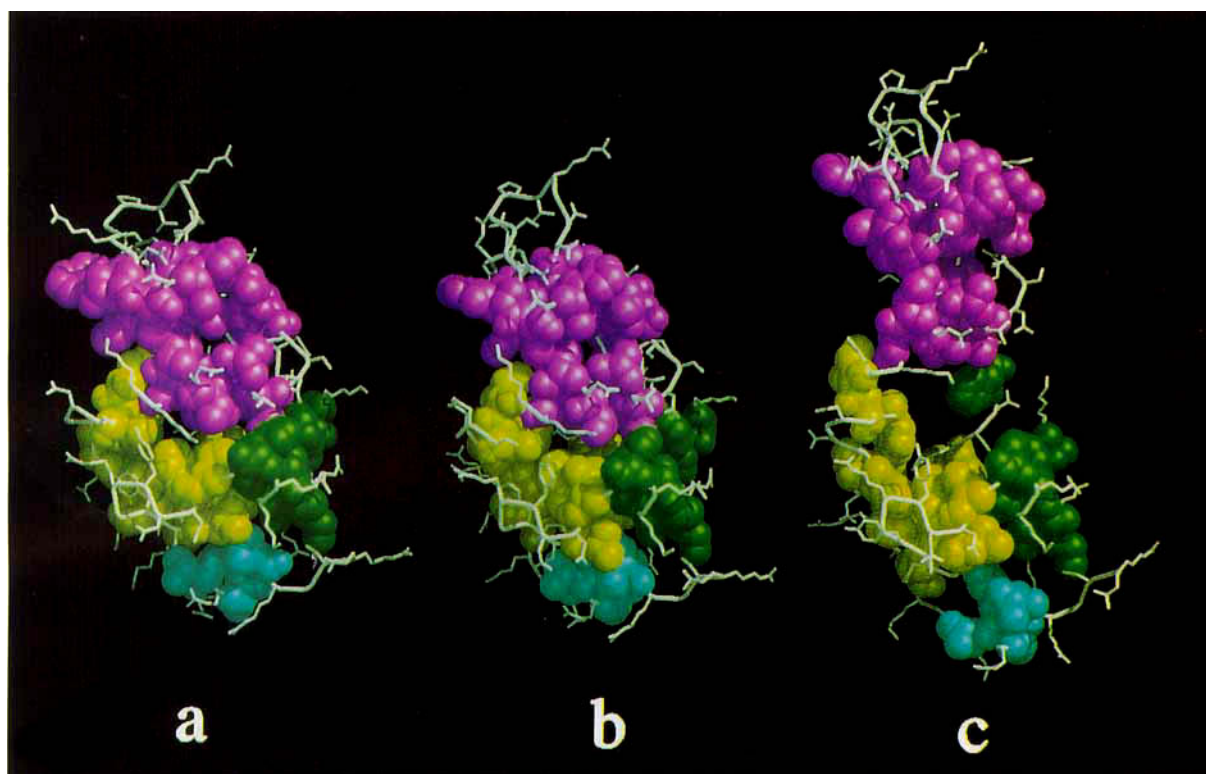


Fig. 8. van der Waals surface of hydrophobic residues for (a) F-run  $<110\text{--}120\text{ psec}>$ , (b) K-run  $<40\text{--}50\text{ ps}>$ , and (c) K-run  $<170\text{--}180\text{ ps}>$ . The color code is the same as in Figure 4.

the structures of the previous F-run (Fig. 7a and b) and those of this K-run (Fig. 7c and d), especially at an early stage of the unfolding process. In later stages, the main difference in the overall shape is that the latter shows expansion along all principal axes of the protein, while the former essentially involves expansion along only two of them (Table I). In both cases, disruption of helix B and elongation of helix C is observed, while the two domains drift away from each other.

Figure 8b ( $<40\text{--}50\text{ ps}>$ ) and 8c ( $<170\text{--}180\text{ ps}>$ ) shows the behavior of hydrophobic residues. At later stages (Fig. 8c) the splitting occurs mainly into two relatively compact blocks ( $\beta$ -domain plus helix C on the one hand, the remainder of the  $\alpha$ -domain on the other). This is in contrast to the T-run (Fig. 4c) where the global shape is much more spherical and splitting occurs into four regions. However, at an early stage (Fig. 4b versus Fig. 8b), interesting analogies can be seen such as the opening of the active site cleft (on the left of the picture) and the formation of an open space near the center of the protein.

#### Peptide Amide Hydrogen Bonding to Water

As a first approximation, the proton/deuterium exchange pattern should correlate with the time each amide proton spends hydrogen bonded to water. This time has been computed as an integral over

the whole simulation for the T-run and is presented in Figure 9 together with experimental results.<sup>34</sup> The time evolution of this protein–water hydrogen bonding was also determined to localize in which part of the trajectory contact with water occurs. Although some clear correlations are evident, many discrepancies can be seen. In the 91 residues with a peptide–water hydrogen bond percentage lower than 52%, 64 of the 65 experimentally observed protected protons are found, the exception being Phe-3. The best agreement with experiment is found among the 37 C-terminal residues, with two exceptions Trp-111 and Arg-125. Thus, the transient hydrogen-bonding pattern of the highly fluctuating helix D and the C-terminal  $3^{10}$  helix still constitutes an effective shield from the solvent. The time evolution of the accessibility of this protein fragment to water correlates with the secondary structure evolution for the T-run (Fig. 2). For instance, the amide protons of residues 122–124 are first in contact with water after 50 ps in the T-run to then display very clear plateaus between 65 and 110 ps and between 120 and 160 ps, corresponding to the phases with more structure in the C-terminal  $3^{10}$  helix (Fig. 2). The low hydrogen bond percentage for Ile-124 is consistent with this amide being also considerably protected in the unfolded state.<sup>54</sup> Residues 30–34, although in contact with the solvent both at the be-

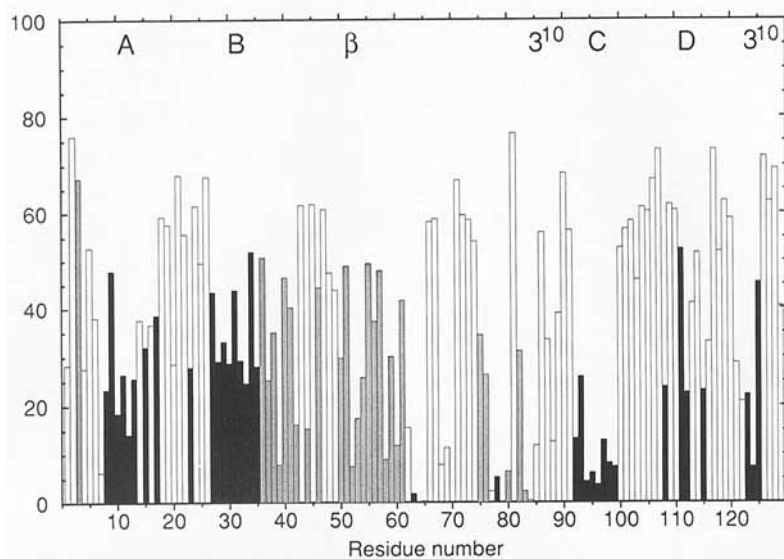


Fig. 9. Percentage of time each backbone amide proton spends being hydrogen-bonded to a water molecule as a function of the residue sequence number for the T-run. These times have been integrated over the whole trajectory and are given as percentage of the total simulation time. A hydrogen bond has been assumed to exist when the H–O distance is less than 0.25 nm and the N–H–O angle is greater than 135°. The experimentally mea-

sured protection against H/D exchange for the peptide NH groups upon refolding<sup>34</sup> is indicated by shading of the bars: black indicates strong protection (exchange <30%), gray indicates weak protection. Unshaded bars correspond to protons not observed experimentally. Symbols at the top indicate location of helices A, B, C, and D, the two 3<sup>10</sup> helices and the  $\beta$ -sheet.

ginning and the end of the simulation, lose contact with water between 60 and 120 ps. This time period matches the secondary structure rearrangements (Fig. 2) giving rise to plateau A (Fig. 1a) and suggests that protection of amide hydrogens might be caused by the formation of transient secondary structure along the folding/unfolding pathway.<sup>56</sup>

A clear difference in the behavior between the experimentally fast and slowly exchanging protons upon folding<sup>34</sup> is not found here. The average peptide–water hydrogen-bond percentage for the two type of protons is 28% for rapidly exchanging and 24% for slowly exchanging. Residues 39, 42, 44, 52, 53, 58, 60, 65, 80, 83, and 84, which belong to the fast exchanging class, are less than 20% of the time hydrogen-bonded to water. Most of these residues belong to the triple stranded  $\beta$ -sheet (residues 42–46, 50–53, and 58–60) and the  $\beta$ -domain 3<sup>10</sup> helix (residues 79–84). Interestingly, the amides of residues 52, 53, 56–58, 63, 65, 83, and 84 are significantly protected in the thermally unfolded state of lysozyme.<sup>54</sup>

The protection of other residues (Asn-39, Asn-65, Leu-75) from water in the simulation is likely to be due to hydrogen bonds between the secondary structure elements (tertiary hydrogen bonds). There are three main  $\beta$ -bridges in the structure. Bridge 2–39 (involving NH 3/CO 38 and NH 40/CO 1) correlates with protection of Phe-3 and Asn-39. The bridge 20–23 (involving NH 23/CO 20 and NH 19/CO 23) accounts for the similar hydrogen-bonding to water propensities of Tyr-20 and Tyr-23 in the simulation.

Bridge 65–79 (involving NH 65/CO 78, NH 80/CO 65 and NH 76/CO 63), although belonging to the irregular loop region, is conserved until the end of the simulation, indicating very high stability. This stability is probably due to the presence of the disulfide bond 64–80, adjacent to Pro-79, reducing the accessible conformations for the residues involved and allowing the hydrogen-bonded partners to stay in close proximity. Surprisingly, Asn-65 and Cys-80 are not among the experimentally most protected residues, but the exchange rates of the amide protons of the adjacent residues 63, 64, and 78 are very low. In the simulation, all five amides are approximately equally shielded from solvent. These protons are also protected in the thermally unfolded state,<sup>54</sup> which could be a consequence of the stability and rigidity of this cluster.

## DISCUSSION

### Methodology: Temperature-Induced Unfolding

The examination of protein folding and unfolding using molecular dynamics simulations is limited by the time scale that can be simulated. In order to speed up the unfolding process, a temperature of 500 K has been used in this study. Within a period of 180 ps the protein largely unfolds. The change in heat capacity (estimated from the accessible surface area) does not, however, reach the value observed experimentally for the denatured state. The exact effect of the high temperature on the dynamics of the unfolding process is unclear, but likely to be substantial. For one, entropic effects will be enhanced. Such bi-

ases could be reduced by using a temperature only slightly above the melting temperature. However, in that case, only the onset of protein unfolding could be simulated within 180 ps.

### Methodology: Pressure-Induced Unfolding

The reason why a pressure of 10 kbar is unable to drive unfolding within 210 ps in the present simulation might be thermodynamic or kinetic. Thermodynamically, the shift in chemical equilibrium upon pressure change is given by

$$\left( \frac{\partial \Delta G_{N \rightarrow U}}{\partial p} \right)_{T, \text{composition}} = \Delta V_{N \rightarrow U} \quad (5)$$

where  $\Delta G_{N \rightarrow U}$  and  $\Delta V_{N \rightarrow U}$  are the free energy and volume changes upon unfolding at a given temperature and composition of the medium. As shown elsewhere<sup>17</sup> the perturbation induced by even very high pressure (several kbar) on the folding equilibrium is relatively small and largely dependent on the extent of reaction at equilibrium. Sizable effects can be expected only if the molecule is near its folding/unfolding transition (typically >1% unfolded form). In our case, it cannot be excluded that the perturbation is too small to drive unfolding.

There are three main contributions to the change in molar volume<sup>17</sup> (solute plus solvent) upon unfolding,  $\Delta V_{N \rightarrow U}$ : (1) The exposure of apolar residues to water induces a contraction due to stronger intermolecular forces and shorter intermolecular distances in the polar solvent, (2) the exposure of charged or polar groups to solvent can result in electrostriction, (3) the change from a restricted packing of residues in the globular protein, imposed by the rigid covalently linked backbone cage, to an unrestricted packing with solvent, with the removal of packing defects or void volumes. The degree of electrostriction will be pH dependent. The total charge of +8e for lysozyme used in the present simulation, instead of +18e for the protein with protonated Asp, Glu, and His residues<sup>35</sup> (pH ≤ 3.9) in the experimental work would be one factor disfavoring unfolding.

Failure to observe unfolding in the simulation may, however, be simply due to kinetic effects. Even if the overall volume change upon unfolding,  $\Delta V_{N \rightarrow U}$ , is negative, any conformational rearrangement in the protein is likely to involve a transient volume expansion ( $\Delta V^\ddagger > 0$ ). This hypothesis is supported by the fact that the kinetics of protein unfolding/refolding is dramatically slowed by high pressure.<sup>57</sup> The restriction of side chain motion at high pressure has also been observed in molecular dynamics simulations of BPTI.<sup>23</sup>

### Methodology: Artificial Driving Forces

In contrast with the T-run, the F- and K-run mainly result in the splitting and drifting away of the  $\alpha$ - and  $\beta$ -domain, each keeping a compact hydro-

phobic core as well as most of its secondary structure. Unlike in the former simulation, the biases, in the latter cases, are quite evident. In the F-run, the added force is radially directed. On a fully symmetrical body, this would drive symmetrical unfolding. On an elongated body like lysozyme, however, the overall stretching force along the longest principal axis will dominate. As the two domains drift away, the forces exerted on each atom in one of the domains will tend to be parallel, and aligned on the main elongation axis. An additional bias is due to the relatively rigid nature of the covalent bonds and bond angles along the polypeptide chain. The forces exerted on the atoms at the protein surface will tend to sum up along the chain, resulting in a much higher effective strain on the atoms in the middle of the protein. Thus, helices B and C (Fig. 3a) feel at their termini a force virtually equal to the sum of the forces exerted on the whole  $\alpha$ - and  $\beta$ -domain, and thus are elongated very quickly. A third bias is that the effect of the force on an element of secondary structure depends on its position relative to the center of mass. An  $\alpha$ -helix situated far from the center of mass (e.g., helix A) will feel a nearly parallel force on each of its atoms, resulting in translation of the helix, while the same helix if passing through the center of mass will feel a strong unfolding force. Thus, to minimize the penalty potential energy while keeping the total potential energy of the system as small as possible (within the constraint of disulfide bonds), helix B is sacrificed and helix C largely extended, resulting in separation of the two domains and a large increase in the mean radius, while the secondary structure and the hydrophobic packing in both domains are mostly preserved.

The K-run was designed to improve this behavior: The mean radius is replaced by the principal radii, in order to take account of the aspherical shape of lysozyme. It is successful in the sense that significant expansion is now observed along the three principal axes (Table I). The expansion is also more controlled as indicated by a quasilinear increase in the radius of gyration (Fig. 6c). In contrast to the F-run, in which the atomic velocity of each atom  $i$  was augmented at every step by a radial component of constant magnitude (only function of  $m_i$ ), in the K-run, the velocities are scaled by different factors. The increment is therefore not imposed in a given direction and its magnitude is determined by the atomic velocities themselves. Expansion occurs on average but not in a concerted way, as indicated by the fact that the contraction temperature and the expansion one stay close. Both methods nevertheless give rise to similar results. This is most likely because a helix far away from the center of mass may still undergo a concerted translational motion, while for a helix near the center of mass, both termini will acquire velocities in opposite directions, forcing it to unfold.

### Peptide Amide Proton Protection

The time a peptide amide proton spends hydrogen-bonded to water clearly shows some correlation with the experimental H/D exchange rate. The implications of such correlations or the lack of them are, however, difficult to interpret. In cases where the simulation indicates poor contact to water for experimentally rapidly exchanging protons, it is difficult to conclude anything. First, because the simulation period is so much shorter than that of the experiment, and second, because any proton unprotected in the native state will not be observed experimentally even if it is transiently protected upon folding. Where an experimentally highly protected proton is often seen hydrogen bonded to water in the simulation (such as Phe-3 or Trp-111), this can most likely be attributed to force-field inaccuracies. Nevertheless, the simulations led to the following observations in regard to the mechanism of proton protection. First, protection of an amide proton upon unfolding might result from transient formation of nonnative secondary structure elements.<sup>56</sup> Second, protection of a proton does not necessarily result from direct involvement in a secondary (or tertiary) hydrogen bond. For example, in the case of Ile-78, Trp-63, and Cys-64, the rigidification of a peptide region by disulfide bonds, hydrogen bonds, or relatively rigid proline residues may result in inaccessibility of a given proton to water, even though it is not directly involved in a stable hydrogen bond. For these protons, protection in the unfolded state is often also reported experimentally.

Amide proton protection is a relatively late event in folding (of the order<sup>31</sup> of 10 to a few 100 ms), occurring when the protein has already assumed a compact almost native-like structure (almost native-like far UV CD spectrum is achieved within a few milliseconds). Therefore, it should not necessarily be concluded that earlier protection of the  $\alpha$ -domain compared to the  $\beta$ -domain corresponds to an earlier formation of the secondary structure elements, but that it may indicate their earlier stabilization with tight packing of the sidechains, exclusion of water, and formation of tertiary contacts. Likewise, with respect to unfolding, one should not necessarily expect an unfolding simulation to result in disruption of the  $\beta$ -domain before the  $\alpha$ -domain. In summary, protection may correlate more with correct tertiary structure and sidechain packing than with secondary structure formation.

### Models for Folding and Unfolding

Two points should be kept in mind when trying to extrapolate from the simulations to a model for protein folding. First, in the simulations, only one unfolding pathway of a unique molecule is monitored. Second, the extent to which a nonequilibrium unfolding simulation can be used to model the equilib-

rium folding process is unclear. For instance, to form a  $\beta$ -sheet, partners located at different sites along the chain must become correctly aligned to form hydrogen bonds. The entropic barrier to the formation of  $\alpha$ -helix is probably lower, because the hydrogen bonding partners are only four residues apart, limiting the conformational space to be searched. On the other hand, upon unfolding, the barrier for the disruption of an element of secondary structure is probably mainly enthalpic. It is therefore not surprising that (unlike in folding experiments) the  $\beta$ -sheet unfolds relatively late in all the simulations, considering the larger intrinsic stability of a  $\beta$ -sheet respective to an  $\alpha$ -helix.<sup>58</sup>

Although the results of the T-run and those of the F- and K-run diverge markedly by the end of the simulations (see Figs. 4c and 8c, as well as 3c, 7b, and 7d), the early stages display striking similarities (see Figs. 4b, 8a, and 8b) with respect to the packing of hydrophobic residues and the global shape of the molecule. In all cases, splitting between the  $\alpha$ - and  $\beta$ -domain is observed, resulting in an increase in the interdomain distance (Figs. 1b, 6b, and 6d) and in the longest principal radius (Table I). This results in the three runs in an opening of the active-site cleft and formation of a water accessible void in the center of the molecule. If the global shape and hydrophobic packing at an early stage display similarities, this is not the case for the secondary structure content. As can be seen from Figures 3c, 7b, and 7d, increasing the temperature to 500 K has a dramatic disruptive effect on the secondary structure, while the two other driving forces preserve a large amount of it.

At later stages of the simulations, the divergence is clear (see Figs. 4c and 8c, as well as 3c, 7b, and 7d). The T-run ends up with a relatively spherical structure with large water accessible voids, almost lacking any secondary structure. The unfolding gives rise to four relatively well packed hydrophobic clusters (four colors in Fig. 4c). On the other hand, the K-run ends up with an extended protein consisting of two well packed domains (roughly, the  $\alpha$ - and  $\beta$ -domain of lysozyme), each with a high content of secondary structure, plus an interdomain region lacking both close packing and most of its secondary structure.

### CONCLUSION

Simulating the unfolding of a protein within a reasonable amount of computer time necessarily requires a strong perturbation, which in turn introduces unavoidable biases into the results. In the present work, four different methods to accelerate the unfolding process have been applied to hen egg white lysozyme and compared.

At high temperature (500 K, T-run), the dynamics of the system is probably significantly affected, among others by entropic effects. These effects are, however, difficult to quantify. Applied to lysozyme, this unfolding method resulted in an almost complete

loss of secondary structure and the splitting of the hydrophobic core into four main clusters. The overall shape of the protein, however, remained relatively compact. The folding model supported by this simulation is a hydrophobic collapse prior to, or concomitant with, secondary and tertiary structure formation.<sup>59</sup>

A pressure of 10 kbar (P-run) proved insufficient to cause unfolding within 210 ps. Both the protein and the solvent motions are dramatically slowed under such conditions. However, using the radii of gyration of the two domains as an indicator, our data agree well with the experimental higher compressibility of the  $\alpha$ -domain compared to the  $\beta$ -domain.<sup>55</sup>

Constant radial force (F-run) and kinetic energy gradient driven (K-run) unfolding suffer from similar biases as will any unfolding scheme depending on the center of mass of the protein. Using such approaches, the unfolding of a secondary structure element will always depend on its location relative to the center of mass, and not only on its inherent stability. The results of both runs are comparable, though in practice the kinetic energy gradient approach proved to be the smoother method. The K- and F-run suggest a folding mechanism involving the docking of preformed  $\alpha$ - and  $\beta$ -domains, both with almost native secondary and tertiary structure, concomitant with the formation of secondary and tertiary structure (hydrophobic packing) in the interdomain region. In all cases, however, the final structures from the simulations do not correspond to fully unfolded structures, based on  $\Delta C_p$  values estimated from solvent accessible surface areas.<sup>52,53</sup> The  $\Delta C_p$  value proved in fact to be of limited use to assess the validity of a simulated structure, as in our simulations very different structures had almost the same calculated heat capacity.

In the thermal unfolding simulation (T-run), the percentage of time that an amide proton spends hydrogen-bonded to water is to some extent correlated with experimental H/D exchange rates upon folding.<sup>34</sup> The simulation suggests that protection of an amide proton can result from a transient (un-)folding conformation and does not necessarily result from direct involvement in secondary structure.

As the power of computers increases, more studies of protein unfolding will be undertaken. However, as the present work has clearly shown, the conclusions drawn from such simulations will strongly depend on the type and size of the perturbation used to drive unfolding. This does not mean that such studies should be discouraged, but rather that the analysis of possible artifacts must form an integral part of the work.

#### ACKNOWLEDGMENTS

Financial support from Unilever Ltd., Vlaardingen, The Netherlands is gratefully acknowledged, as

is the work of Dr. Lorna Smith in identifying artifacts in the high temperature simulation.

#### REFERENCES

1. Matthews, B.W. Genetic and structural analysis of the protein stability problem. *Biochemistry* 26:6885–6888, 1987.
2. Dill, K.A. Dominant forces in protein folding. *Biochemistry* 29:7133–7155, 1990.
3. Creighton, T.E. Stability of folded conformations. *Curr. Opin. Struct. Biol.* 1:5–16, 1991.
4. Dill, K.A. Theory for the folding and stability of globular proteins. *Biochemistry* 24:1501–1509, 1985.
5. Creighton, T.E. Protein folding. *Biochem. J.* 270:1–16, 1990.
6. Baldwin, R.L. Pieces of the folding puzzle. *Nature (London)* 346:409–410, 1990.
7. Kim, P.S., Baldwin, R.L. Intermediates in the folding reactions of small proteins. *Annu. Rev. Biochem.* 59:631–660, 1990.
8. Ptitsyn, O.B. How does protein synthesis give rise to the 3D-structure? *FEBS Lett.* 285:176–181, 1991.
9. Jaenicke, R. Protein folding: Local structures, domains, subunits, and assemblies. *Biochemistry* 30:3147–3161, 1991.
10. Matthews, C.R. The mechanism of protein folding. *Curr. Opin. Struct. Biol.* 1:28–35, 1991.
11. Dobson, C.M. Characterization of protein folding intermediates. *Curr. Opin. Struct. Biol.* 1:22–27, 1991.
12. Dill, K.A. Folding proteins: Finding a needle in a haystack. *Curr. Opin. Struct. Biol.* 3:99–103, 1993.
13. Brooks, C.L., III Molecular simulations of peptide and protein unfolding: In quest of a molten globule. *Curr. Opin. Struct. Biol.* 3:92–98, 1993.
14. Daggett, V., Levitt, M. Protein folding <-> unfolding dynamics. *Curr. Opin. Struct. Biol.* 4:291–295, 1994.
15. Hao, M.-H., Pincus, M.R., Rackovsky, S., Scheraga, H.A. Unfolding and refolding of the native structure of bovine pancreatic trypsin inhibitor studied by computer simulation. *Biochemistry* 32:9614–9631, 1993.
16. Zipp, A., Kauzmann, W. Pressure denaturation of metmyoglobin. *Biochemistry* 12:4217–4228, 1973.
17. Weber, G., Drickamer, H.G. The effect of high pressure upon proteins and other biomolecules. *Q. Rev. Biophys.* 16:89–112, 1983.
18. Mark, A.E., van Gunsteren, W.F. Simulation of the thermal denaturation of hen egg white lysozyme: Trapping the molten globule state. *Biochemistry* 31:7745–7748, 1992.
19. Tirado-Rives, J., Jorgensen, W.L. Molecular dynamics simulations of the unfolding of apomyoglobin in water. *Biochemistry* 32:4175–4184, 1993.
20. Daggett, V., Levitt, M. A model of the molten globule state from molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* 89:5142–5146, 1992.
21. Daggett, V., Levitt, M. Protein unfolding pathways explored through molecular dynamics simulations. *J. Mol. Biol.* 232:600–619, 1993.
22. Kitchen, D.B., Reed, L.H., Levy, R.M. Molecular dynamics simulation of solvated protein at high pressure. *Biochemistry* 31:10083–10093, 1992.
23. Brunne, R.M., van Gunsteren, W.F. Dynamical properties of bovine pancreatic trypsin inhibitor from a molecular dynamics simulation at 5000 atm. *FEBS Lett.* 323:215–217, 1993.
24. McKenzie, H.A., White, F.H., Jr. Lysozyme and  $\alpha$ -lactalbumin: Structure, function and interrelationships. *Adv. Prot. Chem.* 41:174–315, 1991.
25. Tanford, C., Aune, K.C., Ikai, A. Kinetics of unfolding and refolding of proteins III. *J. Mol. Biol.* 73:185–197, 1973.
26. Dobson, C.M., Evans, P.A. Protein folding kinetics from magnetization transfer nuclear magnetic resonance. *Biochemistry* 23:4267–4270, 1984.
27. Kato, S., Okamura, M., Shimamoto, N., Utiyama, H. Spectral evidence for a rapidly formed structural intermediate in the refolding kinetics of hen egg-white lysozyme. *Biochemistry* 20:1080–1085, 1981.
28. Kuwajima, K., Hiraoka, Y., Ikeguchi, M., Sugai, S. Comparison of the transient folding intermediates in lysozyme and  $\alpha$ -lactalbumin. *Biochemistry* 24:874–881, 1985.
29. Ikeguchi, M., Kuwajima, K., Mitani, M., Sugai, S. Evi-



- dence for identity between the equilibrium unfolding intermediate and a transient folding intermediate: A comparative study of the folding reactions of  $\alpha$ -lactalbumin and lysozyme. *Biochemistry* 25:6965–6972, 1986.
30. Chaffotte, A.F., Guillou, Y., Goldberg, M.E. Kinetic resolution of peptide bond and side chain far-UV circular dichroism during the folding of hen egg white lysozyme. *Biochemistry* 31:9694–9702, 1992.
  31. Radford, S.E., Dobson, C.M., Evans, P.A. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature (London)* 358:302–307, 1992.
  32. McCammon, J.A., Gelin, B.R., Karplus, M., Wolynes, P.G. The hinge-bending mode in lysozyme. *Nature (London)* 262:325–326, 1976.
  33. Brooks, B., Karplus, M. Normal modes for specific motions of macromolecules: Application to the hinge-bending mode of lysozyme. *Proc. Natl. Acad. Sci. U.S.A.* 82:4995–4999, 1985.
  34. Miranker, A., Radford, S.E., Karplus, M., Dobson, C.M. Demonstration by NMR of folding domains in lysozyme. *Nature (London)* 349:633–636, 1991.
  35. Samarasinghe, S.D., Campbell, D.M., Jonas, A., Jonas, J. High-resolution NMR study of the pressure-induced unfolding of lysozyme. *Biochemistry* 31:7773–7778, 1992.
  36. Li, T.M., Hook, J.W., III, Drickamer, H.G., Weber, G. Plurality of pressure-denatured forms in chymotrypsinogen and lysozyme. *Biochemistry* 15:5571–5580, 1976.
  37. Smith, P.E., Brunne, R.M., Mark, A.E., van Gunsteren, W.F. Dielectric properties of trypsin inhibitor and lysozyme calculated from molecular dynamics simulations. *J. Phys. Chem.* 97:2009–2014, 1993.
  38. Smith, P.E., van Gunsteren, W.F. Translational and rotational diffusion of proteins. *J. Mol. Biol.* 236:629–636, 1994.
  39. van Gunsteren, W.F., Berendsen, H.J.C. Groningen Molecular Simulation (GROMOS) Library Manual, Biomos, Nijenborgh 4, 9747 AG Groningen, The Netherlands, 1987.
  40. Ramanadham, M., Sieker, L.C., Jensen, L.H. SRLSQ refinement of triclinic lysozyme. *Acta Crystallogr., Sect. A (Suppl.)* 43:13, 1987.
  41. Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T., Tasumi, M. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J. Mol. Biol.* 112:535–542, 1977.
  42. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., Hermans, J. Interaction models for water in relation to protein hydration. In: "Intermolecular Forces." Pullman, B., (ed.) Dordrecht: Reidel, 1981: 331–342.
  43. Ryckaert, J.-P., Ciccotti, G., Berendsen, H.J.C. Numerical integration of the Cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.* 23:327–341, 1977.
  44. van Gunsteren, W.F., Berendsen, H.J.C. Computer simulation of molecular dynamics: Methodology, applications and perspectives. *Angew. Chem. Int. Ed. Engl.* 29:992–1023, 1990.
  45. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., Haak, J.R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* 81:3684–3690, 1984.
  46. Kabsch, W., Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22:2577–2637, 1983.
  47. Fauman, E.B. (1993) "RamPlus." Ph.D. dissertation, University of California, San Francisco, Department of Biochemistry and Biophysics.
  48. Kell, G.S. Precise representation of volume properties of water at one atmosphere. *J. Chem. Eng. Data* 12:66–69, 1967.
  49. Lee, B., Richards, F.M. The interpretation of protein structures: Estimation of static accessibility. *J. Mol. Biol.* 55: 379–400, 1971.
  50. Mark, A.E., van Helden, S.P., Smith, P.E., Janssen, L.H.M., van Gunsteren, W.F. Convergence properties of free energy calculations:  $\alpha$ -Cyclodextrin complexes as a case study. *JACS* 116:6293–6302, 1994.
  51. Haynie, D.T., Freire, E. Structural energetics of the molten globule state. *Proteins* 16:115–140, 1993.
  52. Murphy, K.P., Bhakuni, V., Xie, D., Freire, E. Molecular basis of co-operativity in protein folding III. *J. Mol. Biol.* 227:293–306, 1992.
  53. Khechinashvili, N.N., Privalov, P.L., Tiktopulo, E.I. Calorimetric investigation of lysozyme thermal denaturation. *FEBS Lett.* 30:57–60, 1973.
  54. Radford, S.E., Buck, M., Topping, K.D., Dobson, C.M., Evans, P.A. Hydrogen exchange in native and denatured states of hen egg-white lysozyme. *Proteins* 14:237–248, 1992.
  55. Kundrot, C.E., Richards, F.M. Crystal structure of hen egg-white lysozyme at a hydrostatic pressure of 1000 atmospheres. *J. Mol. Biol.* 193:157–170, 1987.
  56. Creighton, T.E. Characterizing intermediates in protein folding. *Curr. Biol.* 1:8–10, 1991.
  57. Hawley, S.A., Mitchell, R.M. An electrophoretic study of reversible protein denaturation: Chymotrypsinogen at high pressures. *Biochemistry* 14:3257–3264, 1975.
  58. Tobias, D.J., Sneddon, S.F., Brooks, C.L., III Stability of a model  $\beta$ -sheet in water. *J. Mol. Biol.* 227:1244–1252, 1992.
  59. Chan, H.S., Dill, K.A. Origins of structure in globular proteins. *Proc. Natl. Acad. Sci. U.S.A.* 87:6388–6392, 1990.
  60. Kraulis, P. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structure. *J. Appl. Cryst.* 24:946–950, 1991.
  61. Ferrin, T.E., Huang, C.C., Jarvis, L.E., Langridge, R. The MIDAS display system. *J. Mol. Graphics* 6:13–27, 1988.

## APPENDIX A

Here we present the algorithm to carry out kinetic energy gradient driven unfolding. It consists of 8 steps. After each velocity leap-frog step,<sup>44</sup> the following sequence of operations is to be performed.

I. A new coordinate system is defined, with its origin at the center of mass of the protein, and its cartesian axes along the protein principal axes.

II. Two uniform deformation rates are calculated: one corresponding to expansion, the other to contraction. The expansion rate vector is defined by its components as follows (Appendix B)

$$\dot{D}_{\alpha}^{+} = \frac{1}{N r_{\alpha}^2} \sum_{i=1}^N x_{i,\alpha} \dot{x}_{i,\alpha} \quad (\alpha = a, b, c) \quad (\text{A.1})$$

where the dot over a variable denotes a time derivative,  $\alpha$  denotes the three principal axes ( $a$ ,  $b$ , and  $c$ ),  $N$  is the number of protein atoms, and  $r_{\alpha}$  is the principal radius along the  $\alpha$  axis, defined as

$$r_{\alpha} = \left( \frac{1}{N} \sum_{i=1}^N x_{i,\alpha}^2 \right)^{1/2} \quad (\alpha = a, b, c). \quad (\text{A.2})$$

The "+" at the summation sign in Eq. (A.1) indicates that only the positive terms, i.e., the ones for which the atomic velocity along the  $\alpha$  coordinate points away from the center of mass, are summed. The contraction vector is defined correspondingly as

$$\dot{D}_{\alpha}^{-} = -\frac{1}{N r_{\alpha}^2} \sum_{i=1}^N x_{i,\alpha} \dot{x}_{i,\alpha} \quad (\alpha = a, b, c) \quad (\text{A.3})$$

where now only the negative terms are summed.

III. The next step is to split the atomic velocities  $\{\dot{\mathbf{x}}_i\}$  into three terms, namely expansion,  $\{\dot{\mathbf{x}}_i^{+}\}$ , contraction,  $\{\dot{\mathbf{x}}_i^{-}\}$ , and residual velocities,  $\{\dot{\mathbf{x}}_i^{\text{res}}\}$ , in such a way that (A) the sum of the three components is equal to the overall velocities, (B) the sum of their kinetic energies is equal to the overall kinetic en-

ergy, (C) the expansion (respectively, contraction) component reproduces the corresponding deformation rate vector  $\dot{\mathbf{D}}^+$  (respectively  $\dot{\mathbf{D}}^-$ ), and (D) the residual velocities do not have any deformation component. Condition *B* ensures that we are not creating a new, spurious, velocity component in one of the terms that would be cancelled out by an opposite one in another term (e.g., an atom with zero velocity cannot be assigned two exactly cancelling expansion and contraction velocity vectors without an increase in the kinetic energy). If one defines the expansion, contraction and residual velocities by

$$\begin{aligned}\dot{\mathbf{x}}_{i,\alpha}^+ &= \lambda_\alpha^+ \frac{\dot{\mathbf{x}}_i \mathbf{k}_i^+}{(\mathbf{k}_i^+)^2} \mathbf{k}_{i,\alpha}^+ \quad \text{and} \\ \dot{\mathbf{x}}_{i,\alpha}^- &= \lambda_\alpha^- \frac{\dot{\mathbf{x}}_i \mathbf{k}_i^-}{(\mathbf{k}_i^-)^2} \mathbf{k}_{i,\alpha}^- \quad (\alpha = a, b, c)\end{aligned}\quad (\text{A.4})$$

where

$$\mathbf{k}_{i,\alpha}^+ = \begin{cases} \mathbf{x}_{i,\alpha} \dot{\mathbf{D}}_\alpha^+ & \text{if } \mathbf{x}_{i,\alpha} \dot{\mathbf{x}}_{i,\alpha} > 0 \\ 0 & \text{else} \end{cases} \quad (\alpha = a, b, c) \quad (\text{A.5})$$

and

$$\mathbf{k}_{i,\alpha}^- = \begin{cases} \mathbf{x}_{i,\alpha} \dot{\mathbf{D}}_\alpha^- & \text{if } \mathbf{x}_{i,\alpha} \dot{\mathbf{x}}_{i,\alpha} < 0 \\ 0 & \text{else} \end{cases} \quad (\alpha = a, b, c)$$

and

$$\dot{\mathbf{x}}_i^{\text{res}} = \dot{\mathbf{x}}_i - \dot{\mathbf{x}}_i^+ - \dot{\mathbf{x}}_i^- \quad (\text{A.6})$$

it is easy to prove that as long as all the parameters  $\lambda_\alpha$  [Eq. (A.4)] are equal to unity, condition *B* is fulfilled (Appendix C). Condition *A* is fulfilled by construction [Eq. (A.6)]. The  $\mathbf{k}_i^+$  and  $\mathbf{k}_i^-$  vectors basically define the ideal expansion and contraction velocity of particle *i*, if the overall motion was to be exactly uniform with deformation rates  $\dot{\mathbf{D}}^+$  and  $\dot{\mathbf{D}}^-$ , with the particularity that a velocity component pointing outward from the center of mass will be counted in  $\mathbf{k}_i^+$  using deformation rate  $\dot{\mathbf{D}}^+$  whereas if it points toward it,  $\mathbf{k}_i^-$  and  $\dot{\mathbf{D}}^-$  will be used. Equation (A.4) then corresponds to a projection of the actual velocities on these  $\mathbf{k}$  vectors.

If conditions *A* and *B* are satisfied with unit  $\lambda_\alpha$ , conditions *C* and *D* are generally not. In order to satisfy conditions *C* and *D*, we have introduced the scaling parameters  $\lambda_\alpha^+$  and  $\lambda_\alpha^-$ , defined as

$$\lambda_\alpha^+ = \frac{Nr_\alpha^2 \dot{D}_\alpha^+}{\sum_{i=1}^N \mathbf{x}_{i,\alpha} \dot{\mathbf{x}}_{i,\alpha}^+} \quad \text{and} \quad \lambda_\alpha^- = \frac{Nr_\alpha^2 \dot{D}_\alpha^-}{\sum_{i=1}^N \mathbf{x}_{i,\alpha} \dot{\mathbf{x}}_{i,\alpha}^-} \quad (\alpha = a, b, c). \quad (\text{A.7})$$

As these  $\lambda_\alpha$  parameters are usually very close to unity (see the Results Section), condition *B* is still almost fulfilled, while conditions *C* and *D* are now satisfied.

IV. The kinetic energies  $K^+$  and  $K^-$  as well as the

(fictitious) temperatures  $T^+$  and  $T^-$  associated with both deformation modes can then be computed,

$$K^+ = \frac{1}{2} \sum_{i=1}^N m_i (\dot{\mathbf{x}}_i^+)^2 \quad \text{and} \quad K^- = \frac{1}{2} \sum_{i=1}^N m_i (\dot{\mathbf{x}}_i^-)^2. \quad (\text{A.8})$$

And assuming an equal number of degrees of freedom for each of the 3 modes

$$T^+ = \frac{K^+}{1/6 k_B N_{\text{df}}} \quad \text{and} \quad T^- = \frac{K^-}{1/6 k_B N_{\text{df}}} \quad (\text{A.9})$$

where  $N_{\text{df}}$  is the total number of degrees of freedom of the protein.

V. The expansion velocities are scaled by a factor  $\mu \geq 1$  and the contraction velocities by a factor  $\nu \leq 1$ , in such a way that the sum of both kinetic energies,  $K$ , is conserved. If the scaled expansion and contraction velocities (asterisk denotes rescaled quantities) are given by

$$(\dot{\mathbf{x}}_i^+)^* = \mu \dot{\mathbf{x}}_i^+ \quad \text{and} \quad (\dot{\mathbf{x}}_i^-)^* = \nu \dot{\mathbf{x}}_i^- \quad (\text{A.10})$$

then the rescaled kinetic energies become

$$(K^+)^* = \mu^2 K^+ \quad \text{and} \quad (K^-)^* = \nu^2 K^- \quad (\text{A.11})$$

If  $K$  is to be preserved, and thus must be equal to  $K^*$ , Eq. (A.11) can be rewritten as

$$\begin{aligned}\nu^2 K^- &= (K^-)^* = K^* - (K^+)^* = K - \mu^2 K^+ \\ &= K^- + (1 - \mu^2) K^+\end{aligned} \quad (\text{A.12})$$

from which follows that

$$\nu = \left[ (1 - \mu^2) \frac{K^+}{K^-} + 1 \right]^{1/2}. \quad (\text{A.13})$$

The kinetic energy difference  $\Delta K$  between the expansion and the contraction kinetic energies  $K^+$  and  $K^-$  becomes

$$\begin{aligned}(\Delta K)^* &= (K^+)^* - (K^-)^* = \mu^2 K^+ - \nu^2 K^- \\ &= \mu^2 K^+ - \left[ (1 - \mu^2) \frac{K^+}{K^-} + 1 \right] K^- = (2\mu^2 - 1) K^+ - K^-\end{aligned} \quad (\text{A.14})$$

which corresponds to a change in  $\Delta K$  upon velocity scaling by  $\mu$  and  $\nu$  of

$$\Delta(\Delta K) = (\Delta K)^* - \Delta K = 2(\mu^2 - 1) K^+. \quad (\text{A.15})$$

The corresponding change in expansion versus contraction temperature difference is given by [using Eq. (A.9)]

$$\Delta(\Delta T) = \Delta(T^+ - T^-) = \frac{\Delta(\Delta K)}{1/6 k_B N_{\text{df}}}. \quad (\text{A.16})$$

The temperature difference  $\Delta T$  is then weakly coupled to a first order thermostat<sup>45</sup> that relaxes it to a set value  $\Delta\theta$  with a coupling constant  $\tau_\theta$

$$\frac{\Delta(\Delta T)}{\Delta t} = \tau_\theta^{-1} (\Delta\theta - \Delta T). \quad (\text{A.17})$$

Equations (A.15), (A.16), and (A.17) can then be combined and solved for  $\mu$  to give

$$\mu = \left[ 1 + \frac{\Delta t}{2K^+ \tau_\theta} \left( \frac{1}{6} k_B N_{df} \Delta\theta - \Delta K \right) \right]^{1/2}. \quad (\text{A.18})$$

As  $\nu$  is related to  $\mu$  by Eq. (A.13), the scaling factors  $\mu$  and  $\nu$  for the expansion and contraction atomic velocities are defined by this equation together with Eq. (A.18).

VI. The three rescaled atomic velocity components are summed up

$$(\dot{x}_i)^* = (\dot{x}_i^+)^* + (\dot{x}_i^-)^* + \dot{x}_i^{\text{res}}. \quad (\text{A.19})$$

VII. The protein atomic coordinates and velocities are converted back from the center of mass principal axes system to the original system.

VIII. The overall coupling of protein and solvent to the global heat bath is performed in the usual manner.<sup>44</sup>

## APPENDIX B

In the general case, any deformation of a continuous body can be described in the following way. If  $\mathbf{u}(\mathbf{x})$  is the displacement vector connecting a volume element at coordinate  $\mathbf{x}$  in the undeformed state to its position in the deformed state and  $\mathbf{u}(\mathbf{x} + d\mathbf{x})$  the same displacement vector for an adjacent volume element, then

$$d\mathbf{u}(\mathbf{x}) = \mathbf{u}(\mathbf{x} + d\mathbf{x}) - \mathbf{u}(\mathbf{x}) = \nabla \mathbf{u}(\mathbf{x}) d\mathbf{x} = \underline{\underline{\mathbf{D}}}(\mathbf{x}) d\mathbf{x} \quad (\text{B.1})$$

where  $\underline{\underline{\mathbf{D}}}(\mathbf{x})$  is a rank two tensor describing the strain in each point,

$$D_{\alpha,\mu}(\mathbf{x}) = \left( \frac{\partial u_\alpha}{\partial x_\mu} \right)_{\mathbf{x}} \quad (\alpha, \mu = 1, 2, 3) \quad (\text{B.2})$$

where  $\alpha$  and  $\mu$  span the three cartesian coordinates. The deformation is said to be uniform if  $\underline{\underline{\mathbf{D}}}(\mathbf{x})$  is constant in the space.

$$\underline{\underline{\mathbf{D}}}(\mathbf{x}) = \underline{\underline{\mathbf{D}}}. \quad (\text{B.3})$$

Equation (B.1) can then be integrated to

$$\mathbf{u}(\mathbf{x}) = \underline{\underline{\mathbf{D}}}(\mathbf{x} - \mathbf{x}_0) \quad (\text{B.4})$$

where  $\mathbf{x}_0$  is the invariant point of the deformation. In vacuum, this point should be chosen to coincide with the center of mass of the molecule which is fixed in the absence of external forces. In solvent, we shall also take this point as the most logical choice. The  $\underline{\underline{\mathbf{D}}}$  matrix basically contains a dilatation/contraction component, a shear component, and a rotation component. In the proposed algorithm, we are

interested in controlling the evolution of the first component. In the absence of the two latter components, the stress tensor will be diagonal and the whole deformation can be described by one single vector

$$\mathbf{D} = \{D_{\alpha,\alpha}; D_{\beta,\beta}; D_{\gamma,\gamma}\} = \left\{ \frac{\partial u_\alpha}{\partial x_\alpha}; \frac{\partial u_\beta}{\partial x_\beta}; \frac{\partial u_\gamma}{\partial x_\gamma} \right\}. \quad (\text{B.5})$$

Then, Eq. (B.4) reads

$$u_\alpha = D_\alpha (x - x_0)_\alpha \quad (\alpha = 1, 2, 3). \quad (\text{B.6})$$

Its time derivative is

$$\dot{x}_\alpha = \dot{u}_\alpha = \dot{D}_\alpha x_\alpha \quad (\alpha = 1, 2, 3). \quad (\text{B.7})$$

The choice of the cartesian coordinate frame is arbitrary. The best option would be to choose the frame in which the actual  $\underline{\underline{\mathbf{D}}}$  matrix is closest to diagonal (shear component as small as possible). This was not done here. Following Scheraga,<sup>15</sup> we decided to work in the principal axes coordinate frame. The advantage of this choice is a simple mathematical formalism. In this coordinate frame, the principal radii are expressed as

$$r_\alpha = \left( \frac{1}{N} \sum_{i=1}^N x_{i,\alpha}^2 \right)^{1/2} \quad (\alpha = 1, 2, 3) \quad (\text{B.8})$$

and their time derivative as

$$\dot{r}_\alpha = \frac{1}{N r_\alpha} \sum_{i=1}^N x_{i,\alpha} \dot{x}_{i,\alpha} \quad (\alpha = 1, 2, 3). \quad (\text{B.9})$$

Then, using Eq. (B.7)

$$\dot{r}_\alpha = \frac{1}{N r_\alpha} \dot{D}_\alpha \sum_{i=1}^N x_{i,\alpha}^2 = \dot{D}_\alpha r_\alpha \quad (\alpha = 1, 2, 3). \quad (\text{B.10})$$

Equations (B.9) and (B.10) can be combined to give Eq. (A.1).

## APPENDIX C

Here we show that the sum of the kinetic energies  $K^+$ ,  $K^-$ , and  $K^{\text{res}}$  defined by the atomic velocities according to Eqs. (A.4)–(A.6) is equal to the total kinetic energy (condition B) if the  $\lambda$  parameters of Eq. (A.4) are equal to one.

The sum of the kinetic energies of the three components ( $K$ ) is expressed as

$$\begin{aligned} K &= K^+ + K^- + K^{\text{res}} \\ &= \frac{1}{2} \sum_{i,\alpha} m_i [(\dot{x}_{i,\alpha}^+)^2 + (\dot{x}_{i,\alpha}^-)^2 + (\dot{x}_{i,\alpha} - \dot{x}_{i,\alpha}^+ - \dot{x}_{i,\alpha}^-)^2] \\ &= \frac{1}{2} \sum_i m_i [2(\dot{\mathbf{x}}_i^+)^2 + 2(\dot{\mathbf{x}}_i^-)^2 + (\dot{\mathbf{x}}_i)^2 - 2\dot{\mathbf{x}}_i^+ \dot{\mathbf{x}}_i - 2\dot{\mathbf{x}}_i^- \dot{\mathbf{x}}_i] \end{aligned} \quad (\text{C.1})$$

where Eq. (A.6) has been used. The first and the fourth term cancel, because, due to Eq. (A.4) with  $\lambda_\alpha^+ = 1$  for each  $\alpha = a, b, c$

$$(\dot{\mathbf{x}}_i^+)^2 = \frac{(\dot{\mathbf{x}}_i \mathbf{k}_i^+)^2}{(\mathbf{k}_i^+)^4} (\mathbf{k}_i^+)^2 = \frac{(\dot{\mathbf{x}}_i \mathbf{k}_i^+)^2}{(\mathbf{k}_i^+)^2} \quad (\text{C.2})$$

and

$$\dot{\mathbf{x}}_i^+ \dot{\mathbf{x}}_i = \frac{(\dot{\mathbf{x}}_i \mathbf{k}_i^+)}{(\mathbf{k}_i^+)^2} (\mathbf{k}_i^+ \dot{\mathbf{x}}_i) = \frac{(\dot{\mathbf{x}}_i \mathbf{k}_i^+)^2}{(\mathbf{k}_i^+)^2}. \quad (\text{C.3})$$

The second and last term in Eq. (C.1) cancel for sim-

ilar reasons (if all  $\lambda_\alpha^- = 1$ ), and the equation therefore simplifies to

$$K = \frac{1}{2} \sum_i m_i [(\dot{\mathbf{x}}_i)^2] \quad (\text{C.4})$$

which is the expression of the total kinetic energy.