

Atomically Detailed Description of the Unfolding of α -Lactalbumin by the Combined Use of Experiments and Simulations

Tomotaka Oroguchi¹, Mitsunori Ikeguchi², Kimiko Saeki¹
Kiyoto Kamagata¹, Yoriko Sawano³, Masaru Tanokura³, Akinori Kidera²
and Kunihiro Kuwajima^{1*}

¹Department of Physics
Graduate School of Science
University of Tokyo, 7-3-1
Hongo, Bunkyo-ku, Tokyo
113-0033, Japan

²Graduate School of Integrated
Science, Yokohama City
University, 1-7-29 Suehirocho
Tsurumi-ku, Yokohama
230-0045, Japan

³Department of Applied
Biological Chemistry, Graduate
School of Agricultural and Life
Sciences, University of Tokyo
1-1-1 Yayoi, Bunkyo-ku, Tokyo
113-8657, Japan

The recombinant form of goat α -lactalbumin has a significantly faster unfolding rate compared to the authentic form, although the two molecules differ only in an extra methionine at the N terminus of the recombinant. The mechanism of the destabilization caused by this residue was investigated through the combined use of kinetic experiments and molecular dynamics simulations. Unfolding simulations for the authentic and recombinant forms at 398 K (ten trajectories of 5 ns for each form, 100 ns total) precisely reproduced the experimentally observed differences in unfolding behavior. In addition, experiments reproduced the destabilization of a mutant protein, T38A, faithfully as predicted by the simulations. This bidirectional verification between experiments and simulations enabled the atomically detailed description of the role of the extra methionine residue in the unfolding process.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: equilibrium experiments; kinetic experiments; molecular dynamics simulations; protein unfolding; unfolding dynamics

*Corresponding author

Introduction

The recombinant form of wild-type goat α -lactalbumin (designated RWT) expressed in *Escherichia coli* has an extra methionine residue at the N terminus, Met0. The presence of Met0 destabilizes the native state, despite no appreciable change in the three-dimensional structure (Figure 1).^{1,2} This destabilization is due primarily to an increase in unfolding rate, which is significantly faster than that of the authentic form (AWT).² The goal of this study is to obtain an atomically detailed description

about the role of Met0 in the unfolding process using unfolding experiments together with molecular dynamics (MD) simulation at high temperature.^{3–10}

Experiments and simulations provide complementary information. The former yields reliable data but does not provide high spatial and temporal resolution, whereas the latter provides high resolution but is not always perfectly reliable. Therefore, we can exploit trajectories derived from the simulations to supplement the experimental information for the elucidation of the molecular mechanisms at atomic resolution, after confirming that the simulations reproduce detailed results of the experiments.

Here, we performed two series of unfolding MD simulations of goat α -lactalbumin, one for AWT and the other for RWT, at 398 K. Each series consisted of ten trajectories of 5 ns simulation, or 100 ns total. Because the unfolding process is highly stochastic,^{5,11,12} a single trajectory is not sufficient to distinguish between AWT and RWT. However,

Abbreviations used: MD, molecular dynamics; CD, circular dichroism; GdnHCl, guanidine hydrochloride; AWT, the authentic form of the wild-type goat α -lactalbumin; RWT, the recombinant form of the wild-type goat α -lactalbumin; AT38A, the authentic form of the T38A mutant; RT38A, the recombinant form of the T38A mutant.

E-mail address of the corresponding author:
kuwajima@gagliano.phys.s.u-tokyo.ac.jp

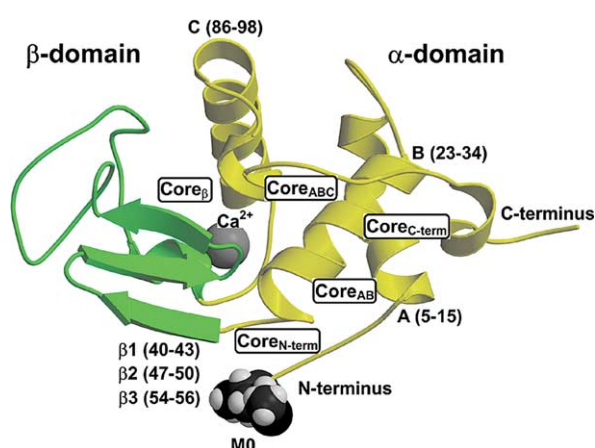


Figure 1. The crystal structure of goat α -lactalbumin (recombinant form, PDB ID: 1HMK); the three missing residues were added at the C terminus in the model. The structure of goat α -lactalbumin consists of the α -domain (yellow) and β -domain (green). The α -domain consists of four α -helices and a 3_{10} -helix. The β -domain consists of a three-stranded anti-parallel β -sheet and a 3_{10} -helix. The Ca^{2+} -binding site is located near the border between the β -domain and the C-helix (residues 86–98). The extra methionine residue at the N terminus, M0, in the recombinant protein is shown by CPK. The five core regions, $\text{Core}_{\text{N-term}}$, Core_{ABC} , Core_{AB} , $\text{Core}_{\text{C-term}}$, and Core_{β} , are defined in terms of native contacts.

averaging ten trajectories for each form successfully provided a clear manifestation of the expected differences, and enabled the identification of important interactions within the protein molecule that were responsible for the differences in unfolding behaviors of AWT and RWT.

Furthermore, the simulation results allowed the construction of mutants, T38A of AWT and of RWT, in which the important interactions predicted to yield higher stability in AWT were eliminated. Equilibrium and kinetic unfolding experiments of the authentic and recombinant forms of T38A confirmed the theoretical prediction.

Based upon synergistic and bidirectional studies using experiments and simulations, we characterized the initial stage of the unfolding process from the viewpoint of protein dynamics. Results of the analysis revealed that the initial stage of unfolding can be represented by a simple amplification of equilibrium fluctuations that occur in the protein in the native state, and that weakened interactions in an N-terminal core region in RWT has resulted in more amplified motions of the protein, and hence the more rapid unfolding of RWT.

Results

Experimental investigation of unfolding and refolding of wild-type proteins AWT and RWT

Figure 2(a) shows the equilibrium thermal unfolding transition curves of AWT and RWT

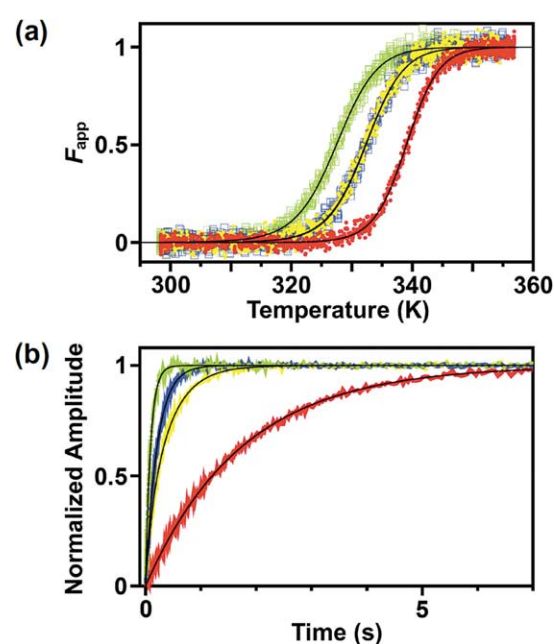


Figure 2. (a) The thermal transition curves of AWT (red dots), AT38A (yellow dots), RWT (blue open squares), and RT38A (green open squares). The apparent fractional extent of unfolding, F_{app} , was plotted against temperature. Continuous curves represent the theoretical thermal transition curves obtained by non-linear least squares fitting of the experimental data. (b) The kinetic unfolding reaction curves of AWT (red), AT38A (yellow), RWT (blue), and RT38A (green). The smooth continuous curves (black) represent the theoretical unfolding reaction curves obtained by non-linear least squares fitting of the experimental data.

measured by the ellipticity at 270 nm. The transition curves are represented by the apparent fractional extent, F_{app} , of unfolding as a function of temperature. The melting temperatures of AWT and RWT were 339.0 K and 332.2 K, respectively, indicating that RWT is less stable than AWT during thermal unfolding. This result is consistent with our previous observation that RWT is less stable than AWT in the GdnHCl unfolding.² Thermodynamic parameters for the two proteins are summarized in Table 1.

In the kinetic measurements, the two proteins showed single-exponential unfolding kinetics with significantly different rate constants: 8.5-fold larger for RWT than for AWT (Figure 2(b)). In contrast,

Table 1. Thermodynamic parameters of unfolding ($\Delta C_p = 1.8$ kcal/mol/K)

Protein	T_m (K)	$\Delta H(T_m)$ (kcal/mol)	$\Delta G(332 \text{ K})$ (kcal/mol)
AWT	339.0 ± 0.1	85.1 ± 1.3	1.6
AT38A	332.2 ± 0.1	68.8 ± 1.4	0.0
RWT	332.2 ± 0.1	69.5 ± 1.3	0.0
RT38A	327.3 ± 0.2	62.5 ± 1.5	−1.0

Table 2. Kinetic parameters of unfolding, $\Delta A'_j = (\Delta A_j / \sum_i \Delta A_i) \times 100$

Protein	A'_1 (%)	k_1 (s ⁻¹)	A'_2 (%)	k_2 (s ⁻¹)
AWT	100	0.55 ± 0.01		
AT38A	79	2.40 ± 0.02	21 ^a	9.09 ± 0.26
RWT	100	4.74 ± 0.01		
RT38A	100	10.94 ± 0.02		

^a 21% of the protein was formylated by CNBr/formic acid cleavage and exhibited more rapid unfolding.

the refolding kinetics was tri-phasic with essentially identical rate constants for AWT and RWT. Therefore, the destabilization in RWT can be ascribed solely to the acceleration of the unfolding rate in RWT. The kinetic parameters of the unfolding and refolding processes for the two proteins are summarized in Tables 2 and 3, respectively.

In the above experiments, we used two different preparations for AWT, one prepared from fresh goat milk, and the other prepared from RWT by cleaving the N-terminal methionine with cyanogen bromide (CNBr)/formic acid. Both preparations gave the same results, although a small fraction (10–20%) of a byproduct, or the formylated form of AWT in the CNBr/formic-acid preparation¹³ exhibited faster unfolding. This formylated form of AWT was identified by MALDI-TOF MS (data not shown).

Unfolding simulations of AWT and RWT

During the equilibrium simulations at 298 K, prior to the high-temperature unfolding simulation, the native structures of AWT and RWT reached an equilibrium state after 2 ns, in which the root-mean-square displacement (RMSD) of C^α atoms from each crystal structure was 1.2 Å.

We carried out two series of unfolding simulations at 398 K, one for AWT and the other for RWT. Figure 3(a)–(c) provide comparisons of these simulation results, each of which consists of ten trajectories of 5 ns simulations. The trajectories were monitored in terms of RMSD from the crystal structure, the fractional number of the native contacts (Q) (see Materials and Methods for the definition), and the radius of gyration (R_g) calculated for all non-hydrogen atoms. Although these parameters for individual trajectories fluctuated greatly with simulation time and trajectory, clear differences between AWT and RWT were observed after averaging the parameters for the ten trajectories. The RMSD and R_g values were larger, and the Q was smaller for RWT than for AWT. There-

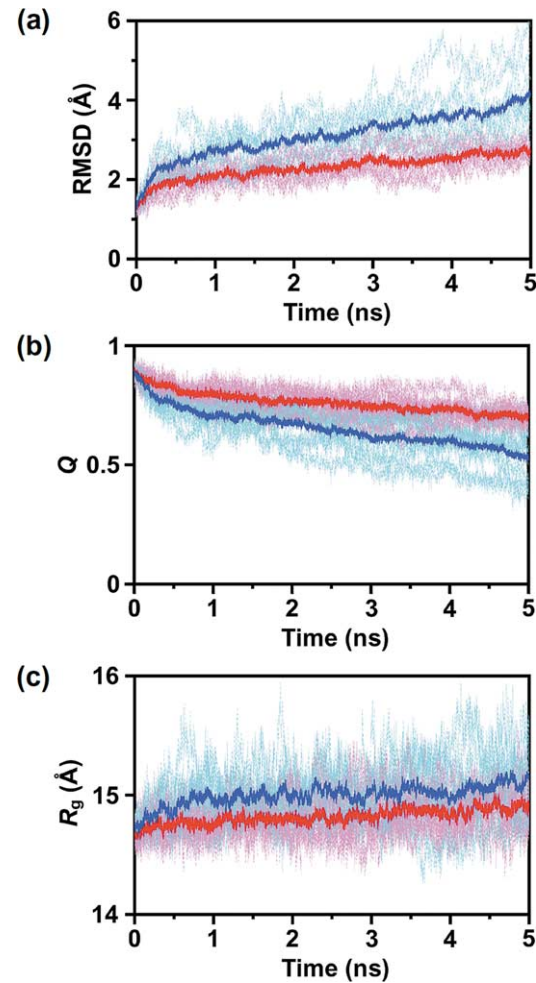


Figure 3. The time-course of the unfolding process during simulation at 398 K. (a) The root-mean-square displacement, RMSD, from the crystal structure; (b) the fractional number of the native contacts, Q ; and (c) the radius of gyration, R_g . The pink and cyan curves are those of the ten trajectories for AWT and RWT, respectively. The red and blue curves represent the average of the data for AWT and RWT, respectively.

fore, RWT was less stable, and unfolded more quickly than did AWT, which agrees well with the experimental observations of low stability and faster unfolding in RWT.

The average R_g values of the AWT and RWT proteins during the last 1 ns were 14.88 Å and 15.07 Å, respectively. These values are larger than the value (14.65 Å) observed in the native state at 298 K for both proteins, but smaller than the value (16.06 Å) observed experimentally in the molten

Table 3. Kinetic parameters of refolding, $\Delta A'_j = (\Delta A_j / \sum_i \Delta A_i) \times 100$

Protein	A'_1 (%)	k_1 (s ⁻¹)	A'_2 (%)	k_2 (s ⁻¹)	A'_3 (%)	k_3 (s ⁻¹)
AWT	82	5.30 ± 0.05	15	1.24 ± 0.07	3	0.13 ± 0.01
AT38A	79	4.91 ± 0.11	16	1.84 ± 0.35	5	0.46 ± 0.06
RWT	88	4.36 ± 0.04	11	0.89 ± 0.09	1	0.11 ± 0.16
RT38A	85	4.32 ± 0.03	15	0.85 ± 0.03		

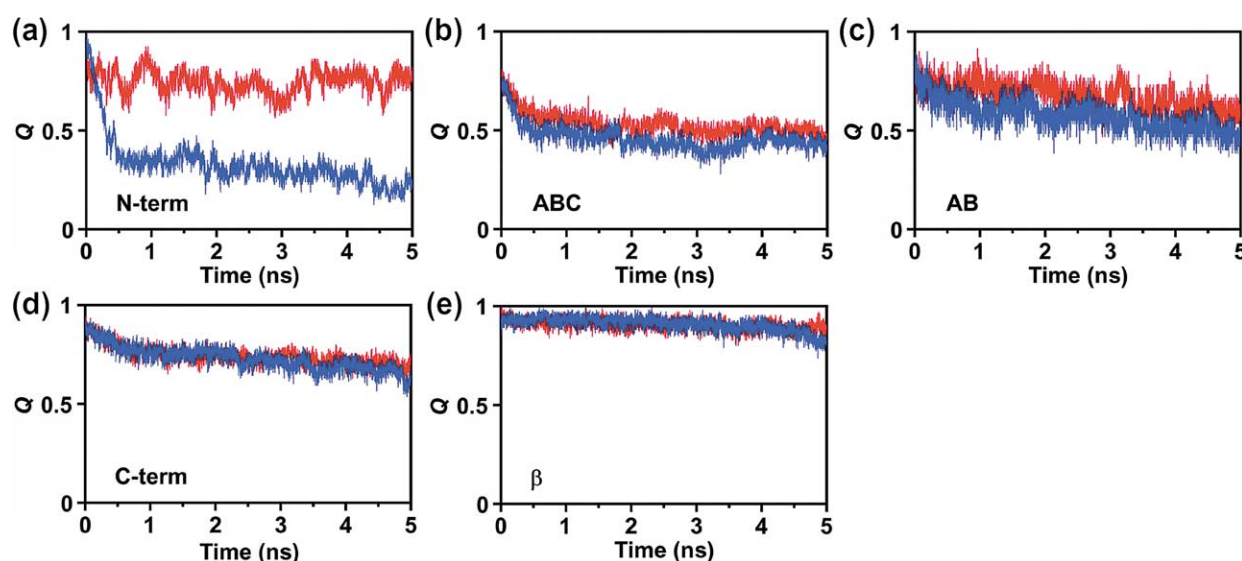


Figure 4. The Q value calculated for each core region: (a) Core_{N-term}, (b) Core_{ABC}, (c) Core_{AB}, (d) Core_{C-term}, and (e) Core _{β} . The red and blue curves represent the average of the data for AWT and RWT, respectively.

globule state.^{14,15} These R_g values indicate that the structure of the two proteins did not reach the molten globule state, but remained at the initial stage of unfolding within a simulation time of 5 ns. Thus, the unfolding simulation at 398 K is considered to represent the initial stage of the unfolding process. It was confirmed in another series of unfolding simulations at 498 K, which reached the molten globule state, that the characteristics in the 398 K simulations were also found within the first 1 ns of the 498 K simulations, i.e., larger RMSD, larger R_g , and smaller Q values of RWT than those of AWT (unpublished results).

Structural differences between AWT and RWT observed in unfolding simulations

To examine the structural changes along the trajectories in detail, we defined five core regions in goat α -lactalbumin, i.e., Core_{N-term}, Core_{ABC}, Core_{AB}, Core_{C-term}, and Core _{β} , where subscripts denote the sites at which residues contact each other (Figure 1). These core regions were uniquely defined based on the atomic coordinates of the native protein structure (see Materials and Methods for the definitions). The Q values monitored structural changes along the trajectories for each core region (Figure 4). Figure 4(a)–(c) show more rapid degradations in Core_{N-term}, Core_{ABC}, and Core_{AB} of RWT than in the corresponding core regions of AWT. Particularly, Core_{N-term} of RWT was disrupted within the first few hundreds of ps, followed by additional gradual disruption of the native contacts up to $Q \sim 0.2$. On the other hand, the other two regions, Core_{C-term} and Core _{β} , did not exhibit any clear differences between the two proteins (Figure 4(d) and (e)). Core _{β} was extremely stable in both proteins, indicating that Core _{β} is the most stable region in both proteins. Based on plots of these Q values, it appears that the unfolding

process was initiated at Core_{N-term} in RWT, followed by propagation of unfolding to Core_{ABC} and Core_{AB}.

Interactions in the N terminus core observed in unfolding simulations

To investigate the more rapid degradation in Core_{N-term} of RWT, we created a list of atomic pairs of positively and negatively charged atoms between residues involved in Core_{N-term} (see Materials and Methods). We then calculated the time-courses of the distances between these atoms along the averaged trajectories. Clear differences were observed in time-courses of the distances between the N-terminal NH_3^+ group of Glu1 and the side-chain oxygen atoms of Asp37 and Thr38. In Figure 5(a) and (b), RWT shows a rapid increase in the distances between N of the N terminus and O^δ of Asp37, and also between N of the N terminus and O^γ of Thr38, whereas the corresponding sites of AWT maintained original hydrogen bond distances. It is of note that the rapid increase in the distance between N of the N terminus and O^γ of Thr38 within the first few hundred ps coincided precisely with a decrease in the Q value of Core_{N-term} as shown in Figure 4(a).

Figure 5(d) and (e) show views of the average N-terminal structures in the 298 K simulations of AWT and RWT, respectively, in which the atoms stabilizing Core_{N-term} are shown. In AWT, the N-terminal NH_3^+ group of Glu1 formed three hydrogen bonds with the side-chain oxygen atoms of Asp37, Thr38, and Gln39. The side-chain hydroxyl group of Thr38 also formed a hydrogen bond with the main-chain oxygen atom of Asp82. Among them, the interaction with Asp37 exhibited the characteristics of the ionic interaction between $-\text{NH}_3^+$ and $-\text{COO}^-$. On the other hand, the presence of the extra Met0 in RWT resulted in

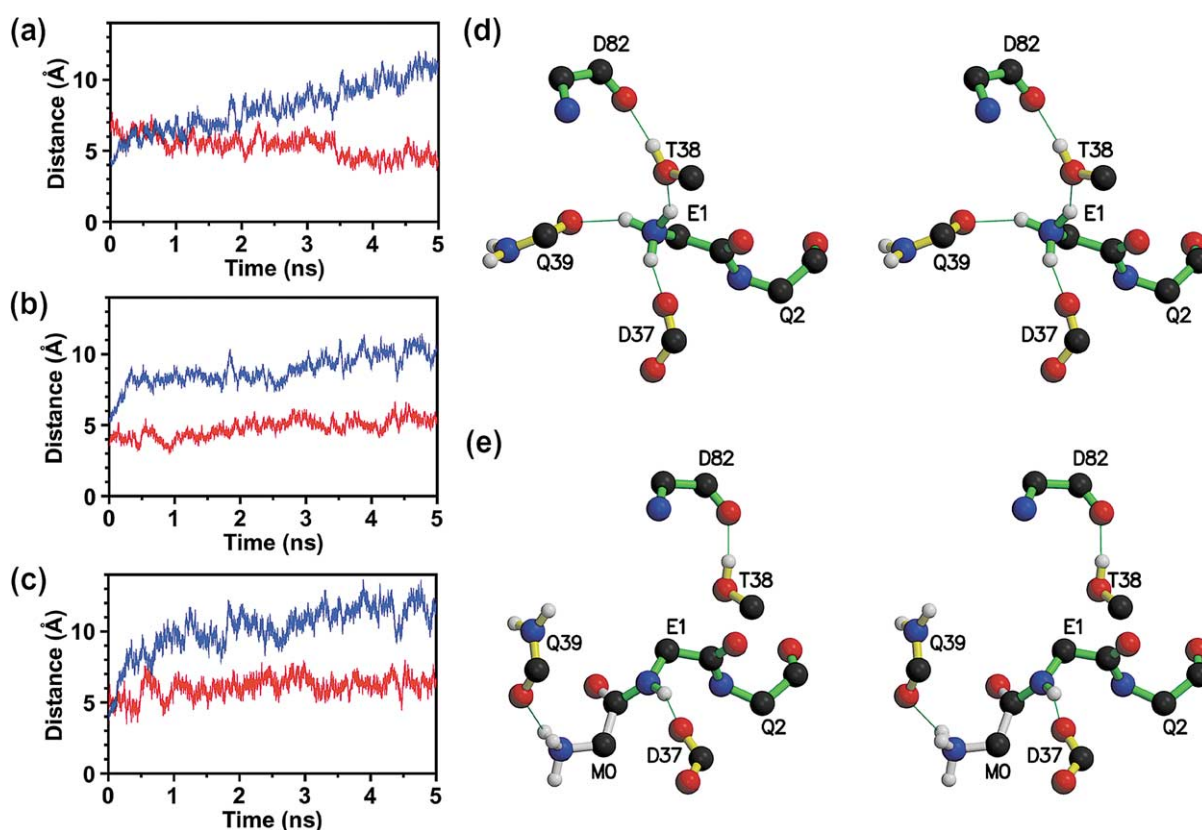


Figure 5. The distance between N of the N terminus and (a) O^δ of Asp37, (b) O^γ of Thr38, or (c) N^ε of Gln39, during simulation at 398 K, averaged for ten trajectories. The red and blue curves represent those for AWT and RWT, respectively. Close views of interaction atoms in Core_{N-term} for (d) AWT and (e) RWT.

the loss of the two hydrogen bonds in Glu1 with Thr38 and Gln39, although the hydrogen bond distance with Asp37 was maintained; however, this bond was no longer an ion pair, because the NH group of Glu1 has no net charge. Instead, the N-terminal NH₃⁺ of Met0 formed a hydrogen bond with Gln39. On the other hand, hydrogen bond distance between Thr38 and Asp82 was maintained. Therefore, the difference in stability between the two proteins can be attributed to the two hydrogen bonds of the NH₃⁺ group of Glu1 with Thr38 and Asp37 in AWT. In RWT, the former bond was lost and the latter was weakened, as demonstrated in the time-course of the distances between these hydrogen bond pairs shown in Figure 5(a) and (b). In addition, the hydrogen bond between Met0 and Gln39 in RWT was not necessarily strong enough to maintain Core_{N-term} (Figure 5(c)).

Another factor that may be responsible for the thermodynamic destabilization of RWT is the conformational entropy effect caused by the presence of an additional Met0 residue, which should result in comparatively greater conformational entropy of RWT compared to that of AWT in the unfolded state.² However, it is difficult to estimate such entropy effects quantitatively based on the present simulations, as they sampled only an initial stage of the unfolding process.

Experimental investigation of unfolding and refolding of T38A mutant proteins AT38A and RT38A

To investigate the role of the hydrogen bond between NH₃⁺ of Glu1 and O^{γ1} of Thr38 in the thermodynamic stabilization of AWT relative to that of RWT, we constructed two mutant proteins, authentic and recombinant forms of the T38A mutant, AT38A and RT38A, respectively, in which Thr38 was replaced by alanine. This mutation was intended to prevent hydrogen bond formation between Glu1 and Thr38 in AWT, which are absent in RWT. As shown in the simulation results, this hydrogen bond is a key interaction to maintain Core_{N-term}. At the same time, the mutation breaks the hydrogen bond between Thr38 and Asp82, but gives the same effects to both AT38A and RT38A. Therefore, it was expected that the T38A mutation would destabilize AWT more than RWT.

We investigated the equilibrium thermal unfolding transitions of AT38T and RT38T, as well as the kinetic unfolding and refolding reactions, using the approach described above for AWT and RWT. Figure 2(a) shows a comparison of the thermal unfolding transition curves of AT38A and RT38A with those of AWT and RWT. As expected from the simulation results, the T38A mutation caused greater changes in AWT than in RWT. At the

melting temperature of RWT, 332.3 K, the T38A mutation destabilized AWT and RWT by 1.6 kcal/mol and 1.0 kcal/mol, respectively. This difference (0.6 kcal/mol) should be attributed to the deletion of the hydrogen bond between Glu1 and Thr38 in AWT. The remaining 1.0 kcal/mol common to AT38A and RT38A would be mostly caused by the deletion of the hydrogen bond between Thr38 and Asp82.

Figure 2(b) shows the kinetic unfolding curves of AT38A and RT38A as compared with those of AWT and RWT. Tables 2 and 3 summarize the kinetic parameters for the unfolding and refolding reactions of these proteins. The kinetic results were completely consistent with the simulation results. AT38A and RT38A unfolded 4.4 and 2.3 times faster than AWT and RWT, respectively, while no significant differences were found in the rate constants of the refolding reactions among the four proteins. In AT38A, a minor kinetic unfolding reaction phase was caused by the formylated species, which amounted to 10–20% of the protein, under the standard protocol of the CNBr/formic acid cleavage.

Discussion

This study took a bidirectional approach: reproduction of experimental results by MD simulations, followed by experimental verification of the theoretical prediction. First, experimental results showing a faster unfolding rate in RWT than in AWT were reproduced by MD simulations. Agreement between experimental and simulation results allowed identification of the atomic interactions destabilizing RWT in the MD trajectories. The extra Met0 prevented hydrogen-bond formation between Glu1 and Thr38/Asp37, destabilizing RWT. Secondly, detailed analyses of the trajectories predicted that the T38A mutation would destabilize AWT more than RWT. According to the prediction, we constructed the mutants AT38A and RT38A to prevent the formation of those hydrogen bonds. Equilibrium and kinetic unfolding experiments on AT38A and RT38A confirmed that the experimental results were fully consistent with the prediction.

These bidirectional analyses strengthened the reliability of the simulation results, and encouraged further investigation into the details of the unfolding trajectories. Thus, the initial stage of the unfolding process was characterized from the viewpoint of protein dynamics.

Dynamic view of the initial stage of unfolding

We explored the correlation between unfolding dynamics at 398 K and equilibrium dynamics at 298 K, and then described how changes in the local atomic interactions at the N terminus amplify the global structural motions associated with unfolding.

We considered the hinge-bending motions between rigid-body domains as the most charac-

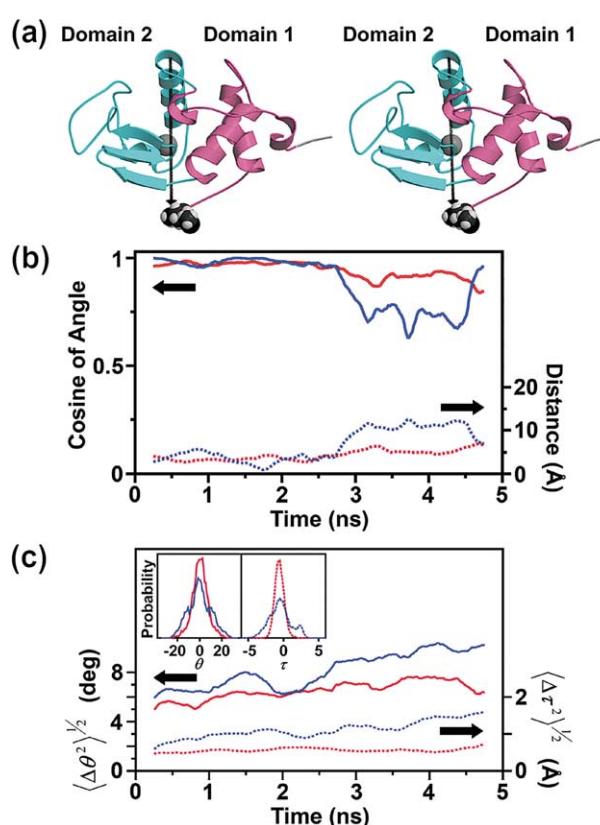


Figure 6. (a) The dynamic domains of α -lactalbumin, domain 1 and domain 2, and the screw axis of the inter-domain motions. (b) Continuous curves: the cosine of the angle between the screw axes of the 298 K simulation and that of the 398 K simulation. Dotted curves: the distance between the two screw axes calculated by the equation, $\Sigma_{i=1}^N |\mathbf{r}_i^s(T=398\text{ K}) - \mathbf{r}_i^s(T=298\text{ K})|/N$, where the summation is taken for N C^α atoms in domain 1 (see Materials and Methods for the definition of $\mathbf{r}_i^s$). Red and blue curves are for AWT and RWT, respectively. (c) The root-mean-square fluctuations of rotation, $\langle \Delta\theta^2 \rangle^{1/2}$ (continuous curves), and those of translation, $\langle \Delta\tau^2 \rangle^{1/2}$ (dotted curves), of the screw motion calculated from the trajectories at 398 K. The screw axis was defined at 298 K. The inset shows the distribution of θ and τ during the last 1 ns of the ten simulations at 398 K, where the origin was set to those values for each crystal structure.

teristic dynamics of α -lactalbumin. The domains, the elements of the rigid-body motions, were determined for AWT and RWT from their unfolding trajectories as described by Hayward *et al.*¹⁶ (see Materials and Methods for the definitions). The analyses gave the same domain partition for AWT and RWT (Figure 6(a) illustrates domain 1 and domain 2). The rigid-body motions of the two domains accounted for the dominant portion of dynamics observed in the unfolding simulations; 69% and 70% variance of the fluctuations for AWT and RWT, respectively. Also at 298 K, the rigid-body domain motions accounted for the dominant part of dynamics; 54% variance of the fluctuations for both AWT and RWT.

It is noted that the C-helix was defined to be a part of domain 2 comprising the β -domain, unlike human lysozyme in which the C-helix belongs to the α -domain.¹⁶ This difference is due to Ca^{2+} in α -lactalbumin, which connects the β -domain with the C-helix to form the most stable core region, Core_β . In contrast, in human lysozyme, the absence of Ca^{2+} would free the C-helix from the β -domain.

The hinge-bending motion between the rigid-body domains can be represented by a screw motion, or a combination of translation and rotation of the two domains (see Materials and Methods for the definitions). The filled arrow shown in Figure 6(a) is the screw axis for the motions at 298 K; two axes for AWT and RWT almost overlap in the Figure, the cosine of the angle between them being 0.93. Figure 6(a) indicates that $\text{Core}_{\text{N-term}}$ and Core_{ABC} , the unstable regions in RWT, are located almost on the screw axis or in the hinge region of the inter-domain motion.

To investigate the time-course of the unfolding process, the screw axes were calculated for the time-window of 500 ps shifting from $t=0$ –5 ns. Figure 6(b) clearly shows that the two screw axes, one for the equilibrium simulation at 298 K and the other for the unfolding simulation at 398 K, were similar up to ~ 3 ns. These findings indicate that the initial stage of unfolding occurred as a simple amplification of equilibrium fluctuations, in particular those of the inter-domain hinge-bending motion. However, after degradation had proceeded (>3 ns), the fluctuations began to modulate.

To determine the differences in the unfolding dynamics between AWT and RWT in detail, we calculated the root-mean-square fluctuations of rotation around the screw axis, $\langle \Delta\theta(t)^2 \rangle^{1/2}$ [$\Delta\theta(t) = \theta(t) - \bar{\theta}$], and the root-mean-square fluctuations of translation along the screw axis, $\langle \Delta\tau(t)^2 \rangle^{1/2}$ [$\Delta\tau(t) = \tau(t) - \bar{\tau}$], where the angular brackets and the macron denote the average in terms of trajectory number (from 1 to 10) and time, respectively (see Materials and Methods for the definition). As expected, RWT fluctuated more than AWT (Figure 6(c)). The accumulated distributions of θ and τ during the last 1 ns indicated that the screw motions in both proteins occurred as the fluctuations centered at the corresponding crystal structures. The difference between the two proteins was seen only in the width of the distributions, indicating that the initial stage of unfolding was not unidirectional motions changing the center of fluctuations, but instead was the amplification of the hinge-opening and closing motions while maintaining the center of the fluctuations. The present findings confirmed that unfolding begins simply as an amplification of the equilibrium fluctuations.

In summary, the more rapid unfolding in RWT can be explained by: the weakened interactions in $\text{Core}_{\text{N-term}}$ (Figure 4(a)) cause amplification of the hinge-bending motions (Figure 6(c)) to induce rapid degradation of the core regions, $\text{Core}_{\text{N-term}}$ and Core_{ABC} , in the hinge region (Figure 4(a) and (b)).

Because $\text{Core}_{\text{N-term}}$ is located at the end of the screw axis of the hinge-bending motion between the two domains (Figure 6), the loss of a single hydrogen bond in $\text{Core}_{\text{N-term}}$ in RWT remarkably amplified the motion of the protein.

Materials and Methods

Protein preparation

The authentic form of wild-type α -lactalbumin was prepared from fresh goat milk by Kuwajima *et al.*¹⁷ The recombinant forms of the wild-type protein and the T38A mutant were purified as described.¹⁸ Concentrations of goat α -lactalbumin were determined spectrophotometrically using an extinction coefficient of $E_{1\text{cm}}^{1\%} = 20.1$ at 280 nm.¹⁹

The recombinant form of goat α -lactalbumin has an extra methionine residue at the N terminus. Because there is no methionine residue in the mature sequence of the protein, we removed the extra N-terminal methionine by CNBr/formic acid cleavage² to produce the authentic form of the T38A mutant. The removal of the N-terminal methionine residue was confirmed by mass spectrometric analysis carried out by MALDI-TOF MS with an Applied Biosystems model Voyager DE STR mass spectrometer. Sinapinic acid was used as the matrix.

Equilibrium thermal unfolding

Equilibrium CD measurements were carried out in a Jasco J-720 spectropolarimeter, using optical cuvettes of 1.0 cm path-length that were equipped with water circulating temperature control (NESLAB, RTE-111). Thermal transitions of the proteins were measured in a buffer (50 mM sodium cacodylate, 50 mM NaCl, and 1 mM CaCl_2 at pH 7.0). CD ellipticity at 270 nm was recorded as a function of temperature. The reversibility of the thermal transition was confirmed by cooling experiments. Protein concentration was 0.20–0.25 mg/ml.

Assuming a two-state transition, the observed ellipticity was calculated by a function of temperature T as follows:

$$[\theta]_{\text{obs}} = \frac{[\theta]_{\text{N}} + [\theta]_{\text{U}} \exp(-\Delta G(T)/RT)}{1 + \exp(-\Delta G(T)/RT)}$$

$$\Delta G(T) = \Delta H(T_{\text{m}})(1 - T/T_{\text{m}}) + \Delta C_p \{T - T_{\text{m}} - T \ln(T/T_{\text{m}})\}$$

$$[\theta]_{\text{N}} = a_{\text{N}} + b_{\text{N}}T$$

$$[\theta]_{\text{U}} = a_{\text{U}} + b_{\text{U}}T$$

where $[\theta]_{\text{N}}$ and $[\theta]_{\text{U}}$ are the ellipticity values of the native and the fully unfolded states, respectively, and they are assumed to linearly depend on T . $\Delta G(T)$ is the unfolding free energy, T_{m} the melting temperature, $\Delta H(T_{\text{m}})$ the unfolding enthalpy change at T_{m} , and ΔC_p the unfolding heat capacity change. We performed global fitting of the data to the equations by non-linear least squares, assuming that ΔC_p is independent of temperature and the same for all the four proteins, AWT, RWT, AT38A, and RT38A.

Kinetic measurements

Refolding and unfolding reactions of the proteins were induced by guanidine hydrochloride (GdnHCl) concentration jumps from 0.0 to 5.0 M and from 5.5 M to 0.5 M, respectively, which were carried out using an SX.18MV stopped-flow fluorescence apparatus (25 °C). These reactions were monitored by the change in intrinsic tryptophan fluorescence above 320 nm with excitation at 295 nm. The dead-time of the apparatus was approximately 4 ms. The mixing ratio of the protein solution and the diluent was 1:10, and the final concentration of the protein after mixing was 0.08–0.11 mg/ml. The diluent solution contained the same buffer (50 mM sodium cacodylate, 50 mM NaCl, and 1 mM CaCl_2 , at pH 7.0) and an appropriate concentration of GdnHCl.

The data were fitted by the non-linear least-squares method to the following equation:

$$A(t) = A(\infty) + \sum_i \Delta A_i \exp(-k_i t)$$

where $A(t)$ and $A(\infty)$ are the observed fluorescence at time t and infinite time, respectively. ΔA_i and k_i are the fluorescence amplitude and the rate constant, respectively, of the i th phase.

Molecular dynamics simulations protocol

Two series of molecular dynamics (MD) simulations were carried out at 398 K, one for the authentic and the other for recombinant goat α -lactalbumin, using the program package MARBLE.²⁰ The CHARMM22 potential²¹ and the TIP3P model²² were used for the proteins and for water, respectively. The periodic boundary condition was applied with the particle mesh Ewald method.²³ The symplectic integrator for rigid bodies was used with a time step of 2 fs, in which water and CH_x , NH_x ($x=1,2,3$), SH, and OH groups were treated as rigid bodies.

The initial structures for the simulations were taken from the crystal structures of goat α -lactalbumin, AWT (PDBID: 1HFY¹) and RWT (1HMK²). The coordinates for the three C-terminal residues, which were lacking in the PDB coordinates, were modeled by referring to the crystal structure of bovine α -lactalbumin (1HFZ¹). The protein and crystal water molecules were immersed in a water box. The total numbers of atoms were 27,669 and 28,334 for AWT and RWT, respectively.

Energy minimization of 2000 steps was followed by equilibration for 1 ns under the NPT condition^{20,24} with $T=298.15$ K and $P=1.0$ atm, where the positions of all atoms, except those of the three C-terminal residues and added water molecules, were constrained. Then, the constraints were gradually removed over a period of 100 ps, followed by further equilibration without any constraints for 2 ns at 298.15 K and 1.0 atm. For the unfolding simulations, ten different initial coordinates were prepared as follows. To adjust the density of water to 0.938 g/ml (the experimental density of bulk water at 398.15 K), an NPT run at 398.15 K was performed for 100 ps with constraints on the positions of the protein and a Ca^{2+} , using each initial set of different velocities. Finally, an NVE run for 200 ps was performed while the constraints were gradually removed during the last 100 ps. The product runs of ten unfolding simulations at 398.15 K were carried out for 5 ns under the NVE condition. Total simulation length was 114.2 ns.

MD analysis: the definition of native contacts, Q , and the core regions

We defined the native contacts in the average structure during the last 1 ns of the equilibrium simulations at 298 K. Two residues were identified to be in contact when two non-hydrogen atoms, each belonging to a different residue, were within 4.5 Å of each other. The contacts between two residues less than five residues apart were not included in the present analysis. Q is the ratio of the number of native contacts remaining in the structure during a simulation at 398 K, $Q = (\text{the number of the native contacts in the structure at 398 K}) / (\text{the number of the native contacts in the native structure})$.

From the native contact map of goat α -lactalbumin, we defined five core regions, $\text{Core}_{\text{N-term}}$, Core_{AB} , $\text{Core}_{\text{C-term}}$, Core_{ABC} , and Core_{β} . $\text{Core}_{\text{N-term}}$ is composed of contacts between residues near the N terminus [1-(36,37,38,39), 2-(36,37), and 3-(36,38)]. Core_{AB} lies between the A-helix and the B-helix [5-27, 8-(27,30), and 9-(24,23,27)]. $\text{Core}_{\text{C-term}}$ lies between the N-terminal and C-terminal regions [5-(117,118,119), 9-(118,119), 24-(115,118,119), 27-118, 28-118, 31-118, and 32-(110,111,118)]. Core_{β} lies within the Ca^{2+} -binding site including the C-helix [50-(77,81), 53-(88,92), 55-(80), 57-77, 60-95, 72-94, 79-87, and 80-(88,91)]. Core_{ABC} lies between the N-terminal region of the α -domain and Core_{β} [3-85, 11-(83,85,89), 12-(85,89), 18-(93,96,97), 21-(96,101), 25-101, 26-(96, 101), and 29-101].

MD analysis: the definition of the dynamic domain and the screw axis of interdomain motion

We defined the dynamic domains and their screw axes according to Hayward *et al.*¹⁶ First, we carried out the principal component analysis using the coordinates of N, C^{α} , C, and C^{β} atoms for each residue, and those of N, C^{α} , C, and H^{α} for glycine. Using the eigenvector of the first principal mode, we determined the displacement vectors of these atoms from the averaged structure. The dot product of the curls of the displacement vectors between two residues determined whether or not these two residues belonged to the same dynamic domain. Two dynamic domains, domain 1 and domain 2, were identified as residues 1–35, 101–120 for domain 1 and residues 36–100 for domain 2. This definition is the same for AWT and RWT (Figure 6(a)).

The inter-domain motion can be described by the screw motion of the two dynamic domains, which is defined by the equation:

$$\Delta \mathbf{r}_i = \tau \mathbf{n} + \theta (\mathbf{n} \times \mathbf{r}_i^s)$$

where $\Delta \mathbf{r}_i$ is the displacement vector of atom i calculated by the principal component analysis described below, $\mathbf{n}$ is the unit vector representing the screw axis, τ is the translational amplitude along the axis, θ is the rotation angle around the axis, and $\mathbf{r}_i^s$ is the normal vector from atom i to the screw axis. The influence of the intra-domain motions was excluded from the definition of the screw axis. First, the snapshots of domain 2 in the MD trajectories were superimposed onto the crystal structure of domain 2. Then, the crystal structure of domain 1 was superimposed onto each of the structures of domain 1 generated above. The resultant set of the crystal structures of domain 1 is the one moving as a rigid body relative to domain 2. The results were subjected to a principal component analysis to give the six principal modes. Using the first principal mode for $\Delta \mathbf{r}_i$, we determined θ , $\mathbf{n}$, τ , and $\mathbf{r}_i^s$ according to the method of Hayward *et al.*¹⁶

Acknowledgements

We thank Ms Natsuko Ichizuka, Mr Yoshiyuki Ito, and Professor Tomoko Akashi of Yokohama City University for help with the mass spectrometric analysis. This study was supported by a Grant-in-Aid for Scientific Research on Priority Areas (Project number 15076201) from the Ministry of Education, Science and Culture of Japan.

References

- Pike, A. C. W., Brew, K. & Acharya, K. R. (1996). Crystal structures of guinea-pig, goat and bovine α -lactalbumin highlight the enhanced conformational flexibility of regions that are significant for its action in lactose synthase. *Structure*, **4**, 691–703.
- Chaudhuri, T. K., Horii, K., Yoda, T., Arai, M., Nagata, S., Terada, T. P. *et al.* (1999). Effect of the extra N-terminal methionine residue on the stability and folding of recombinant α -lactalbumin expressed in *Escherichia coli*. *J. Mol. Biol.* **285**, 1179–1194. Corrigendum (2004) 336, 825.
- Daggett, V. & Fersht, A. (2003). The present view of the mechanism of protein folding. *Nature Rev. Mol. Cell. Biol.* **4**, 497–502.
- Fersht, A. R. & Daggett, V. (2002). Protein folding and unfolding at atomic resolution. *Cell*, **108**, 573–582.
- Snow, C. D., Nguyen, H., Pande, V. S. & Gruebele, M. (2002). Absolute comparison of simulated and experimental protein-folding dynamics. *Nature*, **420**, 102–106.
- 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.
- Mayor, U., Guydosh, N. R., Johnson, C. M., Grossmann, J. G., Sato, S., Jas, G. S. *et al.* (2003). The complete folding pathway of a protein from nanoseconds to microseconds. *Nature*, **421**, 863–867.
- Lindorff-Larsen, K., Vendruscolo, M., Paci, E. & Dobson, C. M. (2004). Transition states for protein folding have native topologies despite high structural variability. *Nature Struct. Mol. Biol.* **11**, 443–449.
- Yoda, T., Saito, M., Arai, M., Horii, K., Tsumoto, K., Matsushima, M. *et al.* (2001). Folding–unfolding of goat α -lactalbumin studied by stopped-flow circular dichroism and molecular dynamics simulations. *Proteins: Struct. Funct. Genet.* **42**, 49–65.
- Lazaridis, T. & Karplus, M. (1997). “New view” of protein folding reconciled with the old through multiple unfolding simulations. *Science*, **278**, 1928–1931.
- Paci, E., Smith, L. J., Dobson, C. M. & Karplus, M. (2001). Exploration of partially unfolded states of human α -lactalbumin by molecular dynamics simulation. *J. Mol. Biol.* **306**, 329–347.
- Ota, M., Ikeguchi, M. & Kidera, A. (2004). Phylogeny of protein-folding trajectories reveals a unique pathway to native structure. *Proc. Natl Acad. Sci. USA*, **121**, 17658–17663.
- Duewel, H. S. & Honek, J. F. (1998). CNBr/formic acid reactions of methionine- and trifluoromethionine-containing λ lysozyme: probing chemical and positional reactivity and formylation side reactions by mass spectrometry. *J. Protein Chem.* **17**, 337–350.
- Kataoka, M., Kuwajima, K., Tokunaga, F. & Goto, Y. (1997). Structural characterization of the molten globule of α -lactalbumin by solution X-ray scattering. *Protein Sci.* **6**, 422–430.
- Arai, M., Ito, K., Inobe, T., Nakao, M., Maki, K., Kamagata, K. *et al.* (2002). Fast compaction of α -lactalbumin during folding studied by stopped-flow X-ray scattering. *J. Mol. Biol.* **321**, 121–132.
- Hayward, S., Kitao, A. & Berendsen, H. J. (1997). Model-free methods of analyzing domain motions in proteins from simulation: a comparison of normal mode analysis and molecular dynamics simulation of lysozyme. *Proteins: Struct. Funct. Genet.* **27**, 425–437.
- Kuwajima, K., Nitta, K. & Sugai, S. (1980). Intramolecular perturbation of tryptophans induced by the protonation of ionizable groups in goat α -lactalbumin. *Biochim. Biophys. Acta*, **623**, 389–401.
- Saeki, K., Arai, M., Yoda, T., Nakao, M. & Kuwajima, K. (2004). Localized nature of the transition-state structure in goat α -lactalbumin folding. *J. Mol. Biol.* **341**, 589–604.
- Kuwajima, K., Hiraoka, Y., Ikeguchi, M. & Sugai, S. (1985). Comparison of the transient folding intermediates in lysozyme and α -lactalbumin. *Biochemistry*, **24**, 874–881.
- Ikeguchi, M. (2004). Partial rigid-body dynamics in NPT, NPAT and NP γ T ensembles for proteins and membranes. *J. Comput. Chem.* **25**, 529–541.
- MacKerell, A. D., Jr, Bashford, D., Bellott, M., Dunbrack, R. L., Jr, Evanseck, J. D., Field, M. J. *et al.* (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
- Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *J. Chem. Phys.* **103**, 8577–8593.
- Martyna, G. J., Tobias, D. J. & Klein, M. L. (1994). Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189.

Edited by F. Schmid

(Received 17 June 2005; received in revised form 5 September 2005; accepted 20 September 2005)
Available online 5 October 2005