

# Targeted Molecular Dynamics Simulations of Protein Unfolding

Philippe Ferrara, Joannis Apostolakis,<sup>†</sup> and Amedeo Caflisch\*

Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Received: December 13, 1999; In Final Form: February 22, 2000

The usefulness of targeted molecular dynamics (TMD) for the simulation of large conformational transitions is assessed in this work on the unfolding process of chymotrypsin inhibitor 2 (CI2). In TMD the force field is supplemented with a harmonic restraint which promotes either the increase of the conformational distance from the native state or the decrease of the distance from a target unfolded structure. As a basis of comparison, unfolding is also simulated by conventional, i.e., unrestrained, molecular dynamics at 375 and 475 K. In all simulations, an implicit approximation of solvation is used to adiabatically model the solvent response, which is appropriate for the nanosecond unfolding simulation method used here. In total, 44 TMD and 25 unrestrained high-temperature molecular dynamics simulations of CI2 unfolding were performed with an implicit solvation model that allowed more than 150 ns to be sampled. Qualitative agreement is found between the results of the TMD and unrestrained molecular dynamics at high temperature. The energies of the conformations sampled during TMD unfolding at 300 and 475 K are comparable to the ones obtained by conventional molecular dynamics at 375 and 475 K, respectively. The sequence of events, i.e., secondary and tertiary structure disruption, is similar in all unfolding simulations, despite the diversity of the pathways. Previous simulations of CI2 performed with different force fields and solvation models showed a similar sequence of events. This indicates that the TMD pathways are realistic even for very large conformational transitions such as protein unfolding.

## 1. Introduction

Many biological processes rely on the transitions between conformational states of macromolecules.<sup>1</sup> For the study of conformational changes, molecular dynamics simulations offer a convenient alternative to experimental approaches because they can treat a single macromolecule at an atomic level of detail.<sup>2</sup> The main drawback of simulation methods is the limited time scale. Most large-scale conformational changes have to be accelerated to be observed in the time scale of a computationally feasible molecular dynamics simulation. Targeted molecular dynamics (TMD) has been introduced to calculate reaction paths between two conformations of a molecule, by continuously decreasing the distance to the target conformation with the help of a constraint.<sup>3</sup> It has been used to predict reaction paths for the conformational changes in *ras* p21.<sup>4,5</sup> For the alanine dipeptide it was shown that most of the TMD paths connecting the (meta)stable states follow the bottom of the free energy valleys.<sup>6</sup> Another conclusion of the alanine dipeptide study was that, even in such a simple system, different paths with similar free energy barriers exist, which points out the necessity of generating many trajectories. Recently, a half-harmonic restraining potential has been used to study the forced unfolding of titin<sup>7</sup> and the unfolding of lysozyme *in vacuo* with the radius of gyration as reaction coordinate.<sup>8</sup>

The folding of chymotrypsin inhibitor 2 (CI2), a 64-residue protein whose folding–unfolding equilibria and kinetics follow a two-state model, has been the subject of experimental<sup>9,10</sup> and theoretical<sup>11,12</sup> works in the past. In this paper we evaluate the

robustness of TMD and the significance of the TMD pathways by a simulation study of CI2 unfolding at different values of the temperature with and without the harmonic restraint. The unfolding of CI2 is a very interesting test case for TMD for two reasons. First, unfolding is a very large and topologically nontrivial conformational change. Second, unlike other conformational changes that have been studied with TMD,<sup>4,5</sup> it can also be simulated by conventional molecular dynamics by simply increasing the temperature. This provides a basis of comparison that is well suited for the assessment of the TMD results.

An implicit solvation model based on the solvent-accessible surface and a distance-dependent dielectric function allows a total of 69 unfolding trajectories to be sampled. The sequence of events (secondary and tertiary structure disruption) and energetics during TMD unfolding are essentially the same as in the unrestrained trajectories, which indicates that the bias due to the restraint does not significantly perturb the dynamics. Moreover, the pathways of CI2 unfolding presented here are consistent with the large body of experimental results as well as simulation data performed by others with different force fields and solvation models.

## 2. Model and Methods

The CHARMM force field was used with a united-atom description of the protein.<sup>13</sup> The solvent was taken into account through an implicit model because it would be computationally prohibitive to perform many unfolding simulations with explicit water molecules. Moreover, the implicit model provides a mean field description of the solvent which avoids the problems related to the relaxation of explicit water molecules around the protein. This is particularly important for both TMD and unfolding at

\* Corresponding author. Phone: (41 1) 635 55 21. Fax: (41 1) 635 57 12. E-mail: caflisch@bioc.unizh.ch.

<sup>†</sup> Current address: GMD-SCAI, Schloss Birlinghoven, D-53754 St. Augustin, Germany.

**TABLE 1: Simulations Performed**

simulation	U300Rt <sup>a</sup>	U300Rrt <sup>b</sup>	U300R <sup>c</sup>	U375	U475R <sup>c</sup>	U475
temperature (K)	300	300	300	375	475	475
length of each run (ns)	1.2	1.2	6	6	0.3	0.3
number of simulations	12 (12 <sup>d</sup> )	25 (12 <sup>d</sup> )	10	10	10	15
use of the restraint	yes	yes	yes	no	yes	no

<sup>a</sup>Unfolding simulations with harmonic restraint based on the decreasing distance from a target unfolded structure (see the text and Figure 1a). The target structure is the final conformation of one of the U475 runs. This structure has none of the 52 native contacts of CI2, and its RMSD from the native state is 15.0 Å. Furthermore, no common features with the native state were found by DSSP.<sup>29</sup> The 12 trajectories were generated by changing the initial assignment of the velocities. <sup>b</sup> Same as in footnote *a* with 25 random conformations used as target unfolded structures. Five hundred structures were generated by randomizing the dihedral angles of the rotatable bonds, followed by 1000 steps of energy minimization. The 25 structures with the most favorable effective energies were retained as starting conformations. They were equilibrated for 5 ps at 300 K, and their average RMSD from the native state is 13.9 Å. <sup>c</sup> Unfolding simulations with harmonic restraint based on the increasing distance from the native state (see the text and Figure 1b). <sup>d</sup> Number of simulations which reached the target structure within a 0.2 Å all-atom RMSD.

high temperature, since the conformational change is strongly accelerated and explicit solvent would show significantly increased friction.

To approximate the screening effects of the electrostatic interactions, the ionizable amino acids were neutralized<sup>12,14</sup> and a distance-dependent dielectric function,  $\epsilon(r) = 2r$ , was used. The CHARMM PARAM19 default cutoffs for long-range interactions were employed; i.e., a shift function<sup>13</sup> was used with a cutoff at 7.5 Å for both the electrostatic and van der Waals terms. Solvation effects were approximated with a model based on the solvent-accessible surface (SAS).<sup>15</sup> Previous studies have shown that the SAS model can be used in MD simulations of different proteins to avoid the main difficulties which arise in *in vacuo* simulations.<sup>16</sup> The mean solvation term is given by

$$V_{\text{solv}}(\mathbf{r}) = \sum_{i=1}^N \sigma_i A_i(\mathbf{r}) \quad (1)$$

for a protein having  $N$  atoms with Cartesian coordinates  $\mathbf{r} = (\mathbf{r}_1, \dots, \mathbf{r}_N)$ .  $A_i(\mathbf{r})$  is the solvent-accessible surface area of atom  $i$ , computed by an approximate analytical expression<sup>17</sup> and using a 1.4 Å probe radius. The atomic solvation parameters  $\sigma_i$  were determined by performing MD simulations at 300 K on six proteins: crambin (1crn, 46 residues), trypsin inhibitor (1bpi, 58 residues), CI2 (2ci2, 64 residues), ubiquitin (1ubq, 76 residues), SH3 domain of the p85 $\alpha$  subunit of bovine phosphatidylinositol 3-kinase (1pht, 83 residues), and histidine-containing phosphocarrier protein (1hdn, 85 residues). Optimization of the atomic solvation parameters based on RMSD from the native conformation, radius of gyration, and number of hydrogen bonds yielded the same  $\sigma_i$  values as in a previous work<sup>16</sup> (0.012 kcal/(mol Å<sup>2</sup>) for carbon and sulfur atoms, -0.060 kcal/(mol Å<sup>2</sup>) for nitrogen and oxygen atoms and zero for the remaining atoms). With these  $\sigma_i$  values, in a 1 ns MD simulation at 300 K, the root mean square deviation (RMSD) from the native structure for the C $\alpha$  atoms averaged over the last 0.5 ns was 1.5 Å (1crn), 1.9 Å (1bpi), 1.8 Å (1ubq), 1.6 Å (1pht), and 2.5 Å (1hdn).

The first 19 residues of CI2 are disordered in the crystal structure and were neglected in the simulations. The conformation obtained after 50 ps of equilibration dynamics at 300 K started from the minimized (by 400 steps of conjugate gradient) crystal structure has a C $\alpha$  RMSD of 1.3 Å and was taken as the initial point for the unfolding simulations. To assess the stability of CI2 with the force field and solvation model, a 10 ns simulation was performed at 300 K starting from the minimized crystal structure. The final value of the RMSD of the C $\alpha$  atoms was 1.4 Å, and the average over the last 2 ns was 1.6 Å.

The details of the targeted molecular dynamics method are described in ref 6. The TMD unfolding simulations were

performed with an additional time-dependent harmonic restraint:

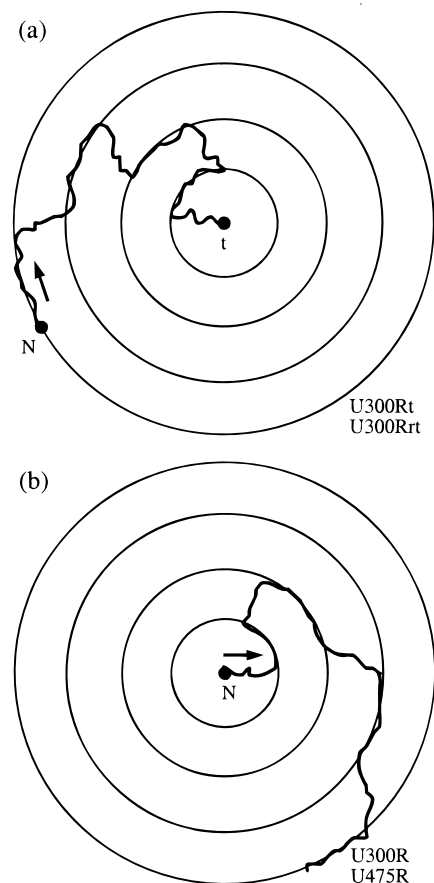
$$U_{\text{res}}(\mathbf{r}(t), t) = KN(\text{RMSD}(\mathbf{r}(t)) - \rho(t))^2 \quad (2)$$

where  $K$  is the force constant (a somewhat arbitrary value of 25 kcal/(mol Å<sup>2</sup>) was used in this study) and  $N$  the number of atoms.  $\text{RMSD}(\mathbf{r}(t))$  is the all-atom RMSD at time  $t$  from the reference state which can be either a target conformation or the initial state. Hence, the harmonic restraint is zero for all structures with an RMSD from the reference state equal to  $\rho(t)$ . At constant  $\rho(t)$ , the restraint keeps the protein at a distance of approximately  $\rho(t)$  from the reference conformation. There is an average bias toward (away from) the reference state only after a change in  $\rho(t)$ . If  $\rho(t)$  is changed continuously, as is the case in previous work,<sup>4,5</sup> there is constantly a bias toward the target conformation and the system is never allowed to relax. Therefore, in this study  $\rho(t)$  is updated 300 times by adding (or subtracting) 0.05 Å, and the system is allowed to equilibrate after each update of  $\rho(t)$ .

In the simulations U300Rt and U300Rrt (where Rt stands for restraint to target and Rrt for restraint to random targets, Table 1) the all-atom RMSD to a target unfolded structure was used.  $\rho(t)$  was initially set to the value of the RMSD between the equilibrated crystal conformation and the unfolded structure and was then discontinuously and linearly decreased to zero in 300 intervals (Figure 1a). This does not guarantee that the target conformation will be reached, because the simulation can be trapped before (see below).

For the simulations U300R and U475R the all-atom RMSD from the native structure was employed instead of the deviation to a target unfolded conformation. At the beginning of the U300R and U475R runs,  $\rho(t)$  was set to 0 Å and was then increased by 0.05 Å 300 times to 15 Å (Figure 1b and Table 1).

Equilibration intervals of 4 and 20 ps at 300 K were employed for total simulation times of 1.2 and 6 ns, respectively. At 475 K, the length of each equilibration interval was 1 ps (total simulation time of 0.3 ns) because equilibration is much faster than at 300 K. The Eckart conditions, with the native state as reference in the U300R and U475R runs and the target conformation as reference in the U300Rt and U300Rrt runs, were applied to avoid rigid body motions when the harmonic restraint was used.<sup>18</sup> The integration time step was 1 fs for the simulations that used the Eckart conditions and 2 fs (with SHAKE<sup>19</sup> on covalent bonds involving hydrogen atoms) for the high-temperature unfolding simulations without the harmonic restraint. In all simulations the temperature was kept almost constant by coupling to an external bath<sup>20</sup> with coupling constants of 0.05 ps (and 0.1 ps for the high-temperature



**Figure 1.** Schematic drawing of TMD unfolding. The circles represent the minima of the harmonic restraint. (a) Restraint based on the decreasing RMSD to a target unfolded conformation labeled by the letter t. (b) Restraint based on the increasing RMSD from the native structure labeled by the letter N.

simulations without the harmonic restraint). The nonbonded interaction list was updated every 10 steps. Coordinates were saved every 0.5 ps in the 0.3 ns simulations and every 2 ps in the 1.2 and 6 ns simulations.

The method applied here is similar to the one used by Diaz *et al.*,<sup>4</sup> and Ma and Karplus,<sup>5</sup> to simulate large conformational changes in *ras* p21. The main difference is that a harmonic restraint is employed here, whereas in the *ras* p21 studies the following holonomic constraint was used to control the sampling:

$$\Phi(\mathbf{r}) = |\mathbf{r}(t) - \mathbf{r}_{\text{target}}|^2 - \rho(t)^2 = 0 \quad (3)$$

The difference between eqs 2 and 3 is that  $|\mathbf{r}(t) - \mathbf{r}_{\text{target}}|$  has to be equal to  $\rho(t)$  in eq 3, whereas in eq 2 RMSD( $\mathbf{r}(t)$ ) can oscillate around  $\rho(t)$ .

### 3. Results and Discussion

**3.1. Overall Behavior.** Three kinds of unfolding simulations for a total time of 156.3 ns were carried out (Table 1). First, CI2 was unfolded by TMD using as target conformation either a structure obtained by conventional unfolding at 475 K (U300Rt), or different random conformations (U300Rrt). Second, CI2 was unfolded at 300 and 475 K by periodically increasing the RMSD from the native state in TMD and without specifying the final conformation (U300R and U475R). Third, denaturation was simulated by conventional molecular dynamics at moderate (375 K) and high (475 K) temperature.

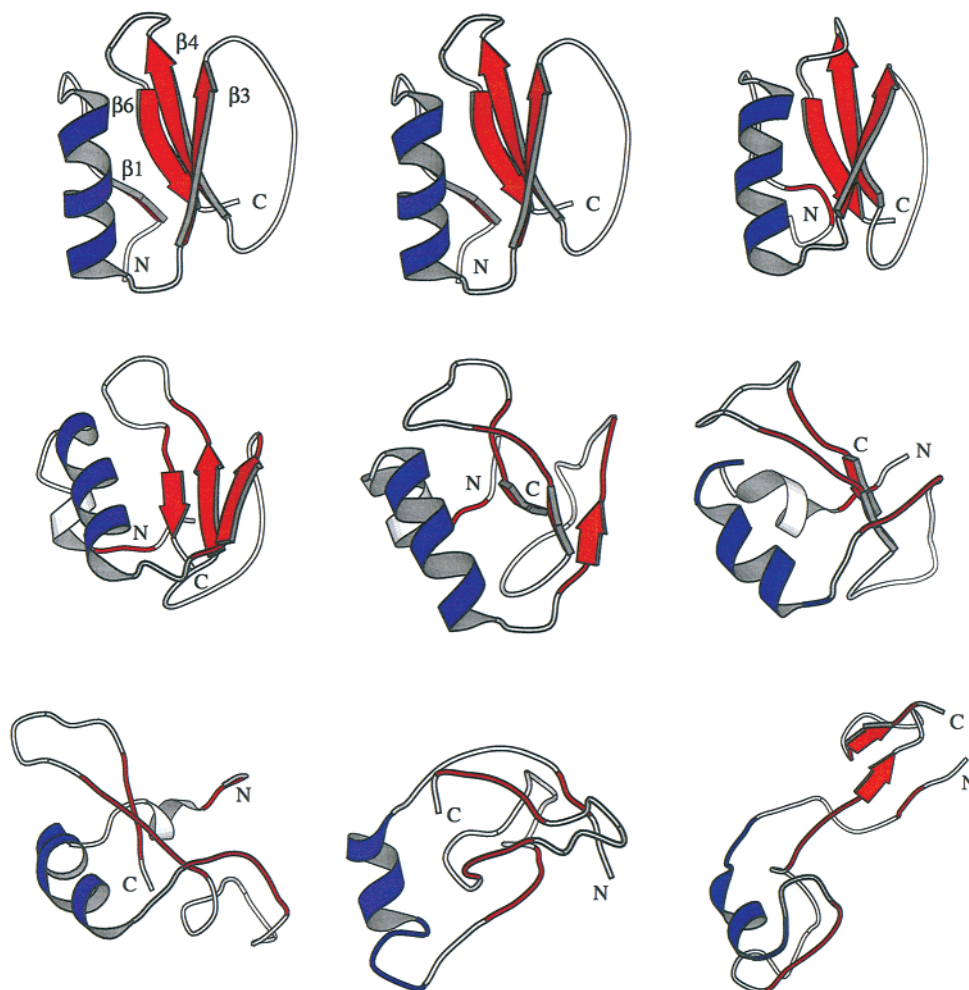
The U300Rt simulations were always successful, leading to a final all-atom RMSD from the target conformation of less than 0.2 Å, whereas the U300Rrt runs reached the final structure in about 50% of the cases. The final RMSD of the unsuccessful simulations is 1.4 Å on average and is due in most cases to a local entanglement in the main chain. The backbone topology of an unfolded conformation at high temperature is simpler and easier to reach than the one of a random conformation, which explains the difference in the success rate.

Nine snapshots saved every 0.75 ns along one of the 6.0 ns U300R runs are shown in Figure 2. The sequence of secondary and tertiary structure disruption is typical of most of the unfolding simulations. There is transient formation of small nonnative regular elements of secondary structure, the most evident of which is a one-turn helix at the N-terminus (three snapshots in the middle of Figure 2). The small two-stranded parallel sheet at the C-terminus in the last snapshot of Figure 2 is in a nonnative topology since the  $\beta_4$  and  $\beta_6$  strands are antiparallel in the folded state. The denaturation of the native  $\alpha$ -helix is the last unfolding event, which is preceded by the rupture of the  $\beta_3$ – $\beta_4$  contacts. Evidence from protein-engineering experiments indicates that the transition state is the same for folding and unfolding and that the  $\alpha$ -helix is the only relatively well formed secondary structure in the transition state.<sup>10</sup> Although it is not possible to accurately determine the position of the transition state along the TMD pathways, the behavior of the  $\alpha$ -helix is in agreement with the mutagenesis data. The contact between Ala16 and Leu49, which is known experimentally to be critical for nucleation,<sup>10</sup> is broken for the first time in the fourth snapshot (at 2.25 ns) and present for the last time in the seventh snapshot (at 4.5 ns) of Figure 2.

The behavior of the RMSD from the native state depends on the type of restraint applied. There is either an almost linear increase to about 14 Å for the U300R (solid lines in Figure 3b) and U475R runs, or a strong increase followed by a plateau region for the U300Rt and U300Rrt runs (solid lines in Figure 3a). The plateau is due to the restraint to a given target conformation, which results in an almost constant RMSD from the native state.

A large variability is seen for the behavior of the radius of gyration during unfolding not only among trajectories of different types but also within runs of the same type (dotted lines in Figure 3). For the U300Rt simulations (Figure 3a) this is due mainly to the different sizes of the random conformations used as target, which had an average radius of gyration of  $15.3 \pm 2.7$  Å. Interestingly, the behavior of the radius of gyration during the U300R runs shows some diversity (Figure 3b). This indicates that restraining the RMSD from the initial conformation has no major impact on the radius of gyration. In most of the U300R and about half of the U375 simulations, the radius of gyration is almost constant for more than the first third (i.e., about 2 ns) of the trajectories and on average more than 80% of the native contacts are still present. This indicates that the protein structure initially performs a local search to find convenient pathways for unfolding in agreement with high-temperature molecular dynamics studies of barnase<sup>21,22</sup> and lattice simulation results.<sup>23</sup> After the first 2 ns of the U300R runs, the behavior is not homogeneous and there are phases of expansion, which can be minor or drastic, sometimes followed by significant contractions. The behavior of the radius of gyration is very similar in all of the U300Rt runs with a small increase and a plateau region in the first third of the simulation followed by an almost linear increase (not shown).





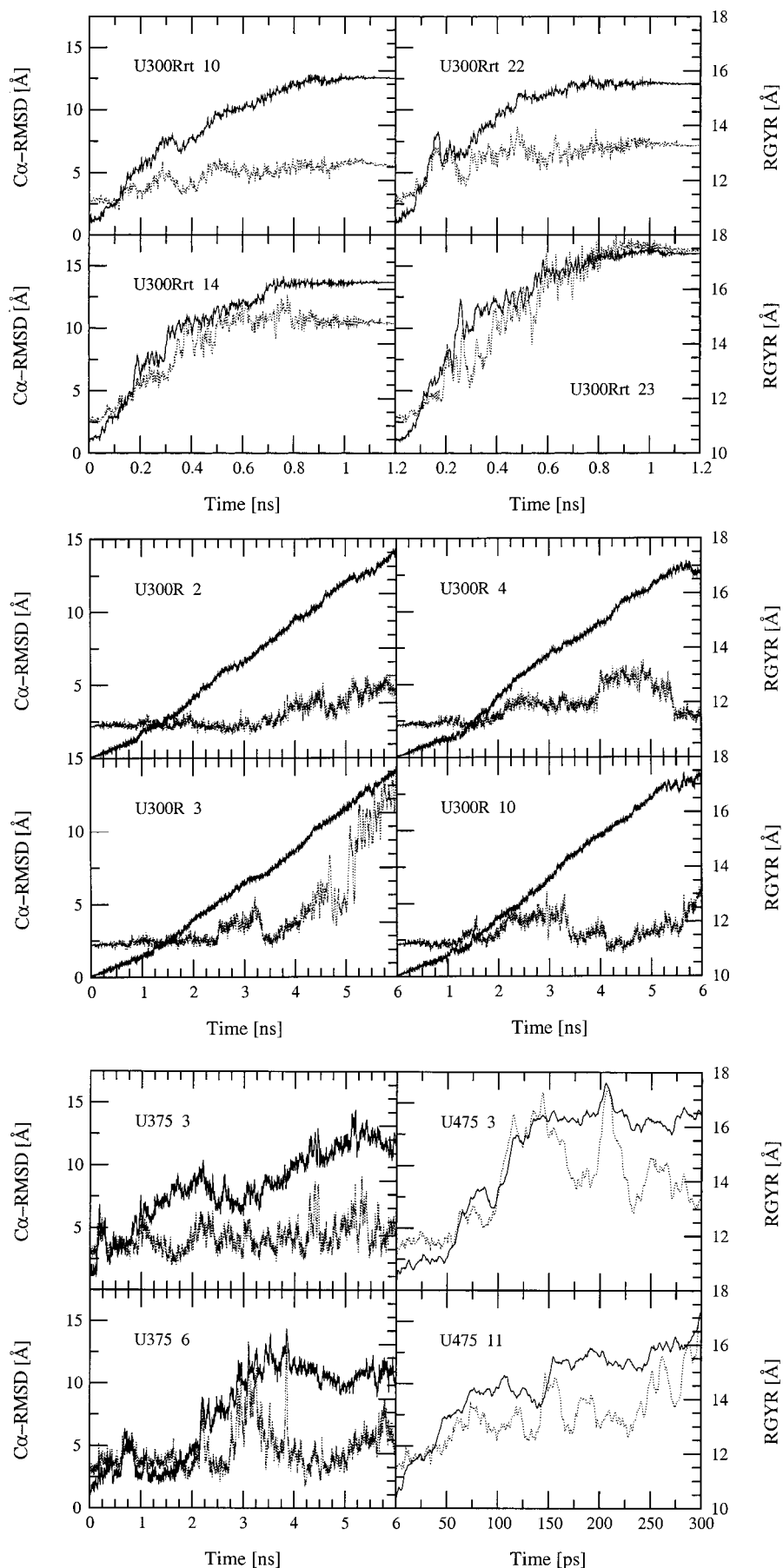
**Figure 2.** Snapshots from one of the U300R unfolding trajectories at, from left to right and top to bottom, 0, 0.75, 1.5, 2.25, 3.0, 3.75, 4.5, 5.25, and 6.0 ns. The secondary structure was assigned with the DSSP program<sup>29</sup> and plotted with MOLSCRIPT.<sup>30</sup> The following elements were defined for the native structure (top left): strand  $\beta 1$ , residues 3–5;  $\alpha$ -helix, 12–24; strand  $\beta 3$ , 27–33; loop, 34–45; strand  $\beta 4$ , 46–52; strand  $\beta 6$ , 59–63. In all frames, segments of the polypeptide chain which correspond to the  $\alpha$ -helix ( $\beta$ -sheet) in the native fold are colored in blue (red).

Typical examples of the evolution of the RMSD from the native state and radius of gyration in the unrestrained high-temperature simulations are depicted in Figure 3c. The RMSD from the native state shows periods of increase, which are more drastic at 475 K than 375 K, intercalated with plateau regions. As expected, the final radius of gyration is significantly larger at 475 K than at 375 and 300 K. This points out a drawback of high-temperature simulations, which favor extended conformations having more favorable conformational entropy.

**3.2. Evolution of Native Contacts.** Fifty-two contacts have been chosen to describe the native state of CI2.<sup>12</sup> The fraction of native contacts ( $Q$ ) decreases faster in the U300Rt and U300Rrt runs than in the U300R runs because in the first half of the simulation the radius of the restraining sphere is larger in the former than in the latter. Furthermore, the time scale of the U300Rt and U300Rrt runs (1.2 ns) is 5 times shorter than the one of the U300R simulations (6 ns). Fifty percent of the native contacts are broken at around 0.18, 0.23, and 3.0 ns in the U300Rt, U300Rrt, and U300R simulations. It has been shown by mutagenesis studies that the interactions between Ala16 and residues of the hydrophobic core, mainly Leu49 and Ile57, are already present in the transition state.<sup>10</sup> Val13 was also found to be almost as folded in the transition state as in the native state. Table 2 lists the average value of  $Q$  at the first disappearance and last appearance of four contacts which involve the aforementioned residues. The first disappearance

occurs generally at medium values of  $Q$ , while the last appearance at low-medium  $Q$  values. This is in agreement with the results of Li and Daggett (Table 2 in ref 11). The first disappearance and last appearance of these contacts, which have been suggested to be critical for folding,<sup>10</sup> are almost simultaneous in the U300Rt and U300Rrt simulations, which indicates that the contacts do not re-form after having been broken (see below).

**3.3. Sequence of Events.** The unfolding trajectories show significant diversity (Figure 3), yet a statistically preferred path emerges from the simulation results. The diversity is also reflected in the RMSD averaged over the conformations with  $Q = 0.25$  obtained during unfolding which has a value of 9.2 Å. It is 6.5 Å at  $Q = 0.50$ , and 3.4 Å at  $Q = 0.75$ . The evolution of the native contacts as a function of the  $C_{\alpha}$  RMSD from the native state is shown in Figure 4. The striking similarity between the results provided by the TMD and unrestrained trajectories suggests that the harmonic restraint has no major impact on the sequence of events during unfolding. Interestingly, the sequence of events is not only similar in all simulations but also comparable to previous unfolding simulations of CI2 performed with different force fields.<sup>11,12</sup> The disappearance of tertiary interactions between the  $\alpha$ -helix and the first strand of the  $\beta$ -sheet is the first unfolding event, and the disruption of the  $\alpha$ -helix is the last unfolding event (see also Figure 2). The U300Rt simulations have somewhat larger standard deviations,

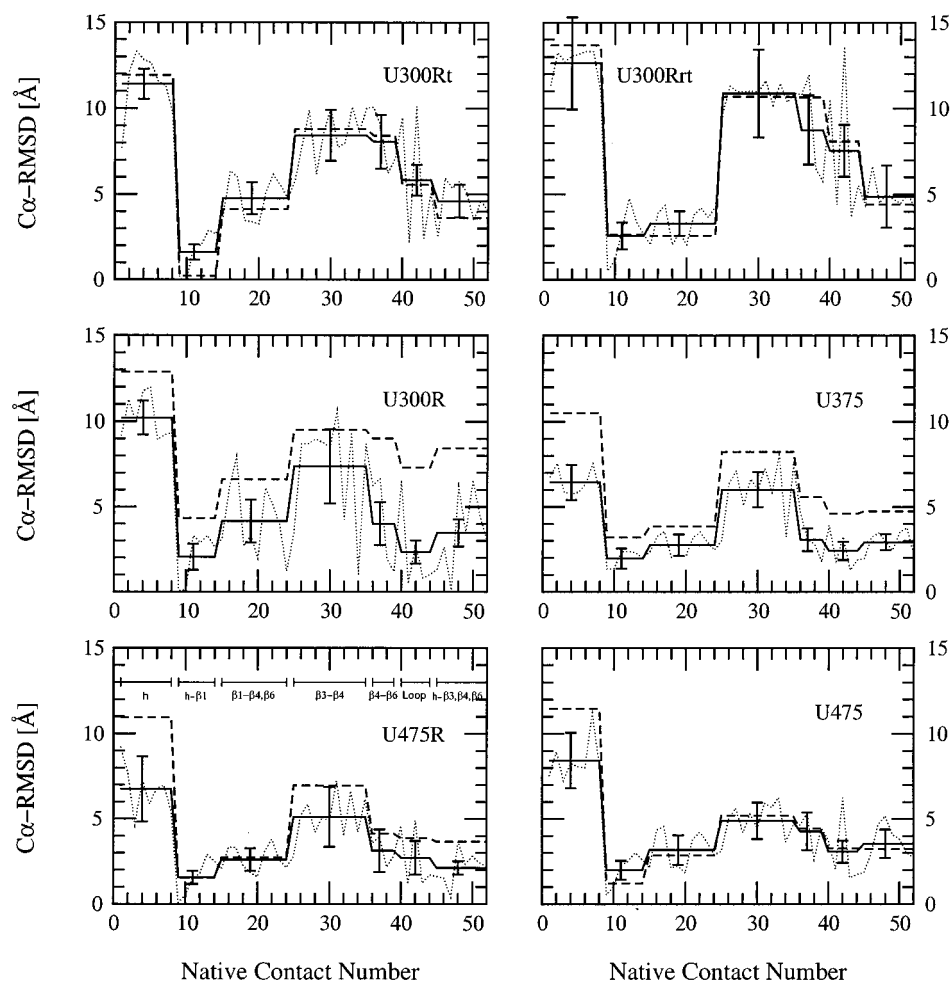


**Figure 3.** C $\alpha$  RMSD from the CI2 native structure (solid lines with scale on the left) and radius of gyration (dotted lines with scale on the right) as a function of simulation time. Data are shown for (a, top) four successful U300Rrt, (b, middle) four U300R, and (c, bottom) (left) two U375 and (right) two U475 runs.

**TABLE 2: Values of  $Q$  at Rupture of Core Contacts<sup>a</sup>**

	U300Rt	U300Rrt	U300R	U375	U475R	U475
Ala16 C <sup>β</sup> –Leu49 C <sup>δ1</sup>						
$Q$ at first disappearance	0.44 ± 0.12	0.54 ± 0.12	0.61 ± 0.14	0.50 ± 0.07	0.50 ± 0.07	0.53 ± 0.10
$Q$ at last appearance	0.48 ± 0.15	0.58 ± 0.11	0.44 ± 0.19	0.31 ± 0.19	0.50 ± 0.24	0.37 ± 0.15
Ala16 C <sup>β</sup> –Ile57 C <sup>γ1</sup>						
$Q$ at first disappearance	0.63 ± 0.09	0.61 ± 0.09	0.46 ± 0.13	0.51 ± 0.06	0.44 ± 0.16	0.47 ± 0.09
$Q$ at last appearance	0.71 ± 0.07	0.67 ± 0.12	0.45 ± 0.11	0.40 ± 0.19	0.44 ± 0.15	0.41 ± 0.14
Leu49 C <sup>δ2</sup> –Ile57 C <sup>γ1</sup>						
$Q$ at first disappearance	0.30 ± 0.06	0.43 ± 0.11	0.61 ± 0.15	0.56 ± 0.11	0.42 ± 0.17	0.44 ± 0.17
$Q$ at last appearance	0.34 ± 0.15	0.42 ± 0.19	0.40 ± 0.15	0.51 ± 0.23	0.37 ± 0.23	0.30 ± 0.18
Val13 C <sup>γ2</sup> –Val51 C <sup>γ1</sup>						
$Q$ at first disappearance	0.55 ± 0.12	0.56 ± 0.11	0.58 ± 0.13	0.58 ± 0.11	0.44 ± 0.15	0.59 ± 0.20
$Q$ at last appearance	0.56 ± 0.11	0.65 ± 0.18	0.44 ± 0.10	0.46 ± 0.18	0.34 ± 0.20	0.45 ± 0.19

<sup>a</sup> Average values and standard deviations of the fraction of native contacts  $Q$  at the first disappearance and last appearance of the contacts which have been shown experimentally to be formed in the transition state. See the caption of Figure 4 for the definition of first disappearance and last appearance.

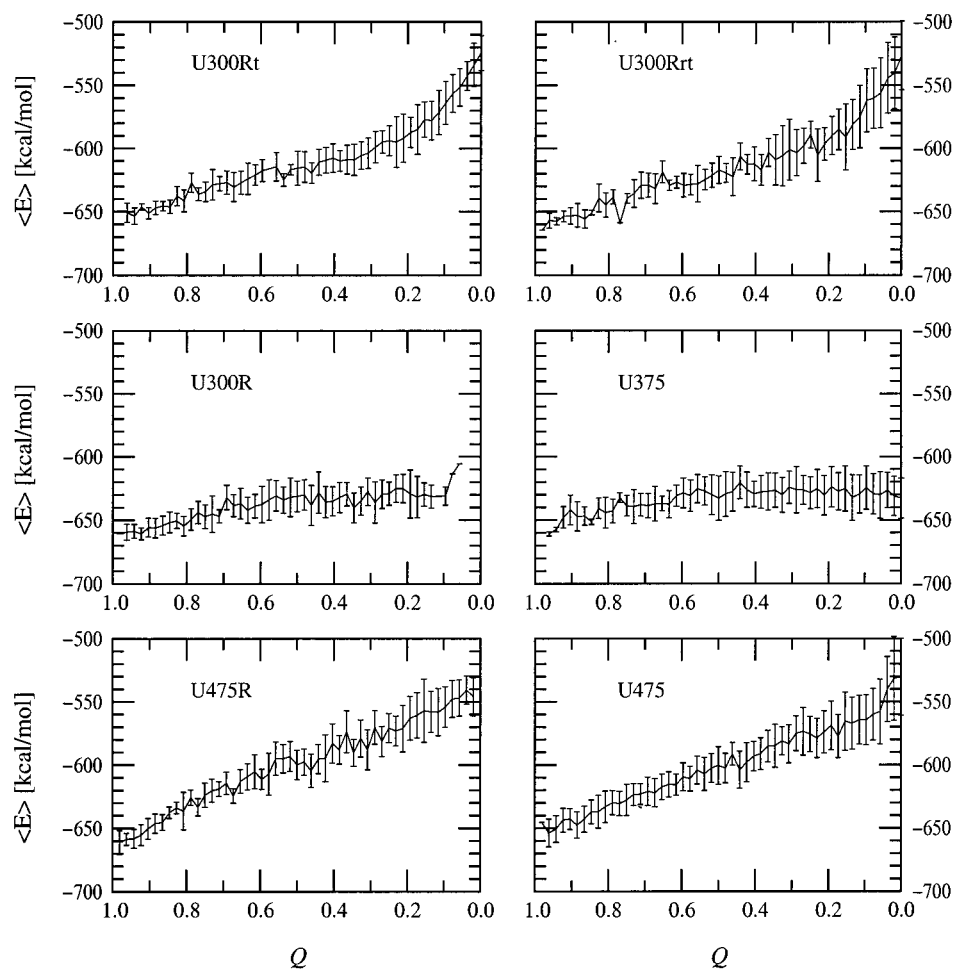


**Figure 4.** The native contacts were classified in seven groups according to their location in elements of secondary structure.<sup>12</sup> The groups are shown in the bottom-left panel. The thin dotted lines represent the  $C_{\alpha}$  RMSD ( $\text{\AA}$ ) from the native state at the first disappearance of the native contacts. Averages over each group of contacts for the first disappearance are in thick lines, and for the last appearance in thick dashed lines. The first disappearance of a contact is defined as the first 20 ps interval during which that contact is absent, and the last appearance as the last 20 ps interval during which it is present. A 5 ps interval was used for the 0.3 ns simulations. A contact is said to be present if the distance between the two atoms defining the contact is less than that in the crystal structure times 1.5. The standard deviations for each group of contacts are shown for the first disappearance; bars = 2 SD. For the U300 Rrt plot only the 12 successful runs were used.

especially for the contacts which are broken late during unfolding, because the target conformations used have a large structural variability. This does not seem to affect the average values.

The difference between the first disappearance and the last appearance indicates that the native contacts can re-form after having been broken. This difference is very small in the U300Rt

and U300Rrt simulations probably because the final conformation is specified. The last appearance can occur before the first disappearance because of their definition (see the caption of Figure 4). A significant shift of the last appearance with respect to the first disappearance is found for the  $\alpha$ -helix in the simulations at 475 K and for most of the native contacts in the U375 and U300R trajectories. In this respect and with respect



**Figure 5.** Average effective energy,  $\langle E \rangle$ , as a function of the fraction of the native contacts  $Q$  for the unfolding simulations. Sixty (120) conformations were selected from each 0.3 or 1.2 ns (6 ns) trajectory. They were submitted to a 10 ps unrestrained MD run at 300 K, followed by 300 steps of energy minimization, before evaluation of the energy. Bars = 2 SD. In the U300R simulations there is no structure at  $Q \approx 0$  probably because the runs were not long enough to obtain such conformations. For the U300Rrt plot only the 12 successful runs were used.

to the effective energy (see the next section), the results provided by the U375 simulations agree better with the U300R than with the U475 results. This indicates that the restraint on the increasing distance from the native structure produces a bias which is not stronger than the one of the entropic effects at 475 K. The small differences between the first appearance and last disappearance at 475 K may originate from the low stability of secondary and tertiary structure at very high temperatures.

The sequence of events displayed in Figure 4 is in agreement with the results of Li and Daggett, who performed unfolding simulations of CI2 at 498 K using explicit water molecules.<sup>11</sup> The comparison with the results of Lazaridis and Karplus<sup>12</sup> shows as the only notable difference that the rupture of the  $\beta 3-\beta 4$  contacts is concomitant with the unfolding of the helix in most of their 24 unfolding trajectories at 500 K. This is not the case in the present simulations apart from the U375 runs.

**3.4. Energetics.** The average value of the harmonic restraint was  $3.8 \pm 0.1$ ,  $1.0 \pm 0.5$ ,  $2.0 \pm 0.1$ , and  $4.6 \pm 0.1$  kcal/mol for the U300Rt, U300Rrt, U300R, and U475R simulations, respectively. This is negligible compared to the energy terms of the force field and provides further support that the harmonic restraint does not bias significantly the energetics of the unfolding process. The effective energy (intraprotein energy and solvation energy) averaged over the MD runs is shown as a function of  $Q$  in Figure 5. There are two distinct behaviors. An almost monotonous increase of the effective energy is seen for the simulations restrained to a target unfolded conformation

(U300Rt and U300Rrt) and at very high temperature (U475 and U475R). A steep increase for  $Q < 0.2$  is present in the U300Rt and U300Rrt trajectories, which are restrained to reach a target conformation whose energy is very high. On the contrary, the 375 K trajectories and the 300 K simulations with a restraint on the increasing distance from the folded state (U300R) show a nearly flat behavior, except for the basin at  $Q > 0.7$ , which corresponds essentially to the native state. Although all simulations show the same average sequence of events (Figure 4), the U300R simulations describe the energetics of the unfolding process at ambient temperature better than the U300Rt and U300Rrt, because the latter were forced to reach target conformations generated at 475 K or randomly. The plot of the effective energy as a function of  $Q$  for the U475R simulations is similar to the one of the U475 trajectories. The same holds for the U300R and U375 simulations. This lends further support to the relatively weak effect of the harmonic restraint on the unfolding energetics and shows that the system is not forced over unnaturally high barriers. However, trapping of trajectories can occur as mentioned above.

That the effective energy depends strongly on the simulation temperature has been found previously in lattice simulations.<sup>24,25</sup> The funnel-like behavior at 475 K (Figure 5) is in agreement with previous results obtained by Lazaridis and Karplus, who performed 24 unfolding MD simulations of CI2 at 500 K with an implicit solvation model different from the one used here. Their solvation energy consists of a sum of atom contributions

and is proportional to the volume of atoms,<sup>12,14</sup> whereas ours is based on the surface area of atoms.

At intermediate values of  $Q$ , i.e.,  $Q$  larger than about 0.2 and smaller than about 0.8, the effective energy is lower at 300 K than at 475 K. Even at 375 K, where the final conformations are as unfolded in terms of  $Q$  as at 475 K, the energies are significantly lower than at 475 K. The accessible conformational space is greatly reduced at 300 K, as demonstrated by lattice statistical mechanics<sup>26</sup> and lattice simulations.<sup>27</sup> Nevertheless, the effective energy behavior of U300R and U375 cannot explain fast folding (see also ref 24).

#### 4. Conclusions

The objective of this work was to evaluate the capabilities of TMD to simulate large conformational transitions in proteins by assessing the relevance of the TMD pathways. For this purpose, TMD simulations of the unfolding of CI2 were compared with unrestrained high-temperature trajectories and previous simulations performed by others with different force fields and solvation models. A comparison based on the unfolding sequence of events and energetics indicates that the bias originating from the harmonic restraint is weak and does not affect the overall unfolding mechanism. Due to the size of conformational space explored by different trajectories the comparison has to be based on qualitative features and average quantities (sequence of events and effective energy behavior). This comparison is relevant because it corresponds to the kind of structural and dynamic information which is usually derived from molecular dynamics simulations. Therefore, TMD seems to be a useful approach to simulate large and topologically nontrivial conformational changes in proteins. Although it is more meaningful to unfold a protein by increasing the RMSD from the native state (U300R runs, Figure 1b), pathways obtained by decreasing the RMSD to a target conformation (U300Rt and U300Rrt runs, Figure 1a) show a comparable sequence of events. Using a set of diverse conformations (randomly generated) as target structures increased the variance of the sequence of events and resulted in premature trapping in about 50% of the runs.

The choice of the reaction coordinate and the implementation of the bias are critical. In the present work some equilibration was allowed after every update of the restraining potential, which is not possible if either the restraint is modified continuously or a constant force is applied. Hünenberger *et al.* studied the unfolding of lysozyme with a constant radial force and compared with a high-temperature (500 K) simulation.<sup>28</sup> They found a significant bias on the unfolding sequence of events due to the center of mass dependence of their constant force and concluded that the two methods, radial force and high temperature, lead to different unfolding processes. In their approach the unfolding of a regular element of secondary structure depends on its location with respect to the center of mass and not only on its inherent stability.<sup>28</sup> With the present

TMD implementation this artifact was not observed and is not expected.

In summary, the present results indicate that TMD is a useful approach that allows qualitative predictions on the overall behavior of a macromolecule during a large conformational transition. TMD generates reasonable CI2 unfolding pathways and relevant information on the unfolding process even without knowledge of the final conformation.

**Acknowledgment.** We thank Nicolas Budin for computer support. We also thank Dr. T. Lazaridis for providing us the partial charges and the list of native contacts of CI2. Ph.F. is a Fellow of the Roche Research Foundation. This work was supported in part by the Swiss National Science Foundation (Grant No. 31-53604.98 to A.C.).

#### References and Notes

- (1) Fersht, A. *Structure and Mechanism in Protein Science*; W. H. Freeman and Co.: New York, 1999.
- (2) Karplus, M.; Petsko, G. A. *Nature* **1990**, *347*, 631–639.
- (3) Schlitter, J.; Engels, M.; Krüger, P.; Jacoby, E.; Wollmer, A. *Mol. Simul.* **1993**, *10*, 291–309.
- (4) Diaz, J. F.; Wroblowski, B.; Schlitter, J.; Engelborghs, Y. *Proteins: Struct., Funct. Genet.* **1997**, *28*, 434–451.
- (5) Ma, J.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 11905–11910.
- (6) Apostolakis, J.; Ferrara, Ph.; Caflisch, A. *J. Chem. Phys.* **1999**, *110*, 2099–2108.
- (7) Paci, E.; Karplus, M. *J. Mol. Biol.* **1999**, *288*, 441–459.
- (8) Marchi, M.; Ballone, P. *J. Chem. Phys.* **1999**, *110*, 3697–3702.
- (9) Jackson, S. E.; Fersht, A. R. *Biochemistry* **1991**, *30*, 10248–10435.
- (10) Itzhaki, L. S.; Otzen, D. E.; Fersht, A. R. *J. Mol. Biol.* **1995**, *254*, 260–288.
- (11) Li, A.; Daggett, V. *J. Mol. Biol.* **1996**, *257*, 412–429.
- (12) Lazaridis, T.; Karplus, M. *Science* **1997**, *278*, 1928–1931.
- (13) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (14) Lazaridis, T.; Karplus, M. *Proteins: Struct., Funct. Genet.* **1999**, *35*, 133–152.
- (15) Eisenberg, D.; McLachlan, A. D. *Nature* **1986**, *319*, 199–203.
- (16) Fraternali, F.; van Gunsteren, W. F. *J. Mol. Biol.* **1996**, *256*, 939–948.
- (17) Hasel, W.; Hendrickson, T. F.; Still, W. C. *Tetrahedron Comput. Methodol.* **1988**, *1*, 103–116.
- (18) Eckart, C. *Phys. Rev.* **1935**, *47*, 552–558.
- (19) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (20) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (21) Caflisch, A.; Karplus, M. *J. Mol. Biol.* **1995**, *252*, 672–708.
- (22) Caflisch, A.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 1746–1750.
- (23) Dinner, A.R.; Karplus, M. *J. Mol. Biol.* **1999**, *292*, 403–419.
- (24) Dobson, C. M.; Sali, A.; Karplus, M. *Angew. Chem., Int. Ed.* **1998**, *37*, 869–893.
- (25) Karplus, M. *Folding Design* **1997**, *2*, S69–S75.
- (26) Dill, K. A. *Biochemistry* **1985**, *24*, 1501–1509.
- (27) Šali, A.; Shakhnovich, E.; Karplus, M. *Nature* **1994**, *369*, 248–251.
- (28) Hünenberger, P. H.; Mark, A. E.; van Gunsteren, W. F. *Proteins: Struct., Funct. Genet.* **1995**, *21*, 196–213.
- (29) Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577–2637.
- (30) Kraulis, P. *J. Appl. Crystallogr.* **1991**, *24*, 946–950.