PROTEINS: Structure, Function, and Genetics 39:252–260 (2000)

Computer Simulations of Protein Folding by Targeted Molecular Dynamics

Philippe Ferrara, Joannis Apostolakis, and Amedeo Caflisch*
Department of Biochemistry, University of Zürich, Zürich, Switzerland

ABSTRACT **We have performed 128 folding and 45 unfolding molecular dynamics runs of chymotrypsin inhibitor 2 (CI2) with an implicit solvation model for a total simulation time of 0.4 microseconds. Folding requires that the three-dimensional structure of the native state is known. It was simulated at 300 K by supplementing the force field with a harmonic restraint which acts on the root-mean-square deviation and allows to decrease the distance to the target conformation. High temperature and/or the harmonic restraint were used to induce unfolding. Of the 62 folding simulations started from random conformations, 31 reached the native structure, while the success rate was 83% for the 66 trajectories which began from conformations unfolded by high-temperature dynamics. A funnel-like energy landscape is observed for unfolding at 475 K, while the unfolding runs at 300 K and 375 K as well as most of the folding trajectories have an almost flat energy landscape for conformations with less than about 50% of native contacts formed. The sequence of events, i.e., secondary and tertiary structure formation, is similar in all folding and unfolding simulations, despite the diversity of the pathways. Previous unfolding simulations of CI2 performed with different force fields showed a similar sequence of events. These results suggest that the topology of the native state plays an important role in the folding process. Proteins 2000;39:252–260.**
© 2000 Wiley-Liss, Inc.

Key words: protein folding; chymotrypsin inhibitor 2; molecular dynamics simulation; implicit solvation model

INTRODUCTION

Large theoretical and experimental research efforts are devoted to try to understand how a protein folds into its native structure.[1–5] Two kinds of approaches can be used on a computer to address this question. The first uses a simplified description of the protein where each amino acid is represented by one[6,7] or two beads[8] on a three-dimensional lattice. Such extremely simplified models have the advantage that the conformational space is small enough to allow the search for the native state on a reasonable timescale. The second approach is based on molecular dynamics (MD) simulation, in which most or all of the atoms of the solute and the solvent are treated explicitly. In MD simulations one is generally confronted with the problem of the conformational sampling efficiency. This is particularly critical in the case of protein folding, because it is experimentally known that most proteins need from milliseconds to minutes to reach their functional state. The common approach to study folding with MD simulation is to unfold starting from the native state because it is easier to speed up the events under denaturing conditions than under folding conditions. The energy landscapes under folding and unfolding conditions are not necessarily similar and therefore the issue of whether unfolding simulations are representative for the folding process is still open.[9] An elegant but time-consuming way of addressing this problem is to construct the free-energy surface of the folding/unfolding process at 300 K starting from conformations obtained during high temperature unfolding simulations.[10,11] Recently, the computational power of massively parallel computers has made MD simulations on the microsecond timescale possible to study the early stage of folding of the villin headpiece subdomain, a 36-residue protein.[12]

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.[13] Recently, it has been used to predict reaction paths for the conformational changes in *ras* p21.[14,15] For the alanine dipeptide it was shown that most of the TMD paths follow the bottom of the free energy valleys.[16] 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 different trajectories. 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[17,18] and theoretical[19,20] work in the past. We have chosen this protein as a model system because of the abundance of experimental and theoretical results which allow a precise comparison. Here, we show that TMD can be used to generate folding pathways for CI2 with the native state as

Grant sponsor: Swiss National Science Foundation; Grant number: 31-53604.98.

Joannis Apostolakis's present address is GMD-SCAI, Schloss Birlinghoven, D-53754 St. Augustin, Germany.

*Correspondence to: Amedeo Caflisch, Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland. E-mail: caflisch@bioc.unizh.ch

Received 29 October 1999; Accepted 28 January 2000

© 2000 WILEY-LISS, INC.

final structure. Despite the diversity of the pathways, the folding simulations have similar sequence of events, and the unfolding runs show the inverse sequence.

MODEL AND METHODS

The CHARMM force field[21] was used with a united-atom description[22] of the protein. The ionizable amino acids were neutralized[20,23] and a distance-dependent dielectric function, $\epsilon(r) = 2r$, was used to approximate the screening effects of the electrostatic interactions. The CHARMM PARAM19 default cutoffs for long range interactions were used, i.e., a shift function[21] was employed 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).[24] In this frame, the mean solvation term is given by:

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

for a protein having M atoms with Cartesian coordinates $\mathbf{r} = (\mathbf{r}_1, \ldots, \mathbf{r}_M)$. $A_i(\mathbf{r})$ is the solvent-accessible surface area of atom i, computed by an approximate analytical expression[25] and using a 1.4 Å probe radius. 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.[26] The atomic solvation parameters were determined by performing MD simulations on six proteins: crambin (1crn, 46 residues), trypsin inhibitor (1bpi, 58 residues), CI2 (2ci2, 64 residues), ubiquitin (1ubq, 76 residues), SH3 domain of the p85α subunit of bovine phosphatidylinositol 3-kinase (1pht, 83 residues), and histidine-containing phosphocarrier protein (1hdn, 85 residues). Optimization of the atomic solvation parameters yielded the same values as in a previous work[26] (0.012 kcal/mol Å^2 for carbon and sulfur atoms, −0.060 kcal/mol Å^2 for nitrogen and oxygen atoms and zero for the remaining atoms). With these parameters, in a 1 ns MD simulation at 300 K, the RMSD from the native structure for the C_α 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 equilibration dynamics at 300 K started from the minimized (by 400 steps of conjugate gradient) crystal structure has a C_α RMSD of 1.3 Å and was taken as the initial (final) point for the unfolding (folding) 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_α 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 Apostolakis et al.[16] Folding and unfolding simulations are performed with an additional time-dependent harmonic restraint:

$$U_{res}(\mathbf{r}) = KM(\mathrm{RMSD}_N(t) - \rho(t))^2,$$

$$\text{and } \mathrm{RMSD}_N(t) = \frac{|\mathbf{r}(t) - \mathbf{r}_N|}{\sqrt{M}}, \quad (2)$$

where K is the force constant (a value of 25 kcal/mol Å^2 was used here), RMSD_N is the all-atom root mean square deviation from the native state $\mathbf{r}_N$, and M is the number of constrained atoms, which were chosen to be all the atoms of CI2. $\mathbf{r}(t)$ is the $3M$ dimensional vector of the protein coordinates at time t. Hence, the zero of the harmonic restraint is not given by the native state, but by the difference between $\mathrm{RMSD}_N(t)$ and $\rho(t)$. At the beginning of a folding simulation, $\rho(t)$ is set to the value of the RMSD between the initial conformation and the folded state and is then discontinuously decreased to zero in three hundred intervals. Therefore, during most of the simulation the zero of the harmonic restraint is very far away from the native state conformation. It should be stressed that this does not guarantee that the native structure will be reached since the simulation can be trapped before as discussed below.

To investigate the relevance of unfolding simulations, we also generated 45 MD unfolding trajectories, starting from the equilibrated crystal structure of CI2, at high temperature (375 K and 475 K) and/or by means of the harmonic restraint. For unfolding, the same scheme is applied, but in the opposite way: the starting value of $\rho(t)$ is set to 0 Å and is then increased in three hundred intervals to 15 Å, which corresponds nearly to the average RMSD_N of the random conformations used as starting points for folding. Three interval lengths were tested: 1 ps, 4 ps, and 20 ps, resulting in a total simulation time of 300 ps, 1.2 ns, and 6 ns, respectively. In all simulations, the temperature was kept almost constant by coupling to an external bath. The Eckart conditions with the native state as reference were applied to avoid rigid body motions when the harmonic restraint was used.[27] The integration time step was 1 fs (2 fs for the high temperature unfolding simulations without the harmonic restraint) and the nonbonded interactions were 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 ns and 6 ns simulations.

The method applied here is essentially the same as the one used by Diaz et al.,[14] and Ma and Karplus,[15] to simulate large conformational changes in *ras* p21. The main difference is that a harmonic restraint is employed here, whereas the following holonomic constraint was used in the *ras* p21 studies to control the sampling,

$$\Phi(\mathbf{r}) = |\mathbf{r}(t) - \mathbf{r}_N|^2 - \rho(t)^2 = 0. \quad (3)$$

The difference between Eq. 2 and 3 is that $|\mathbf{r}(t) - \mathbf{r}_N|$ has to be equal to $\rho(t)$ in Eq. 3, whereas in Eq. 2 $|\mathbf{r}(t) - \mathbf{r}_N|$ can oscillate around $\rho(t)$ with a harmonic potential. A half-harmonic restraining potential has been recently used to study the forced unfolding of titin[28] and the unfolding of lysozyme in vacuo with the radius of gyration as reaction coordinate.[29]

TABLE I. Simulations Performed

	Folding				Unfolding			
Simulation	FR1[a]	FR2[b]	FS1[c]	FS2[d]	U475	U475R	U375	U300R
Temperature (K)	300	300	300	300	475	475	375	300
Length of each run (ns)	1.2	6	6	1.2	0.3	0.3	6	6
Number of simulations	50 (25[e])	12 (6[e])	16 (11[e])	50 (44[e])	15	10	10	10
Use of the restraint	Yes	Yes	Yes	Yes	No	Yes	No	Yes

[a]Different random starting conformations. Five hundred structures were generated by randomizing the dihedral angles of the rotatable bonds, followed by thousand steps of energy minimization. The fifty structures with the most favorable energies were retained as starting conformations. Their average RMS deviation from the native state is 14.4 Å.
[b]Different random starting conformations randomly picked out from the 25 initial FR1 structures which led to a successful folding simulation. Their average RMS deviation from the native state is 14.0 Å.
[c]Different initial velocities and same starting conformation obtained at the end of one of the U475 simulations. This structure has none of the fifty-two native contacts and its RMS deviation from the native state is 11.2 Å. The only common features with the native state found by the DSSP program[39] were an H-bonded turn at residues 34–35 and a bend at residue 55. The latter is involved in an H-bonded turn in the native structure which is not present in this conformation. Therefore, this structure can be considered as having little or no memory of the native state.
[d]As in footnote c, with a starting conformation obtained at the end of another U475 simulation. This structure has none of the fifty-two native contacts and its RMS deviation from the native state is 15.0 Å. No common features with the native state were found by DSSP.[39]
[e]Number of simulations that reached the native conformation within 0.2 Å RMS deviation.

RESULTS AND DISCUSSION

General Behavior

Four kinds of folding simulations for a total time of 288 ns were carried out (Table I). First, fifty simulations of 1.2 ns each were started from fifty random conformations (FR1). These folding simulations were successful in 50% of the cases leading to a final all atom root mean square deviation from the native state ($RMSD_N$) below 0.2 Å. An additional set of simulations (FR2) were initiated from twelve random structures randomly picked out from the twenty-five initial random conformations that had led to successful folding in FR1. This did not improve the success rate, which was 50% also for the FR2 runs. The final $RMSD_N$ was about 1.5 Å for the 31 trajectories (25 FR1 and 6 FR2) which got trapped mainly because of a very localized backbone entanglement. Two series of folding runs (FS1, FS2) were started from conformations obtained by the high temperature unfolding simulations. In the FS1 runs, the native structure was reached in 11 of 16 runs. In the five unsuccessful FS1 simulations, the final $RMSD_N$ ranges between 0.8 Å and 1.1 Å. The only difference with the native state is that the turn between the β4 strand and β6 strand is not correctly formed. The FS2 simulations were able to find the native state in nearly 90% of the cases. The final $RMSD_N$ for the unsuccessful simulations was 1.4 Å on average. The small final values of the $RMSD_N$ for the unsuccessful folding simulations do not necessarily imply that the final structure is close to the native state. Indeed, in most of the unsuccessful simulations, the final conformations are characterized by an entanglement of the backbone and the deviation from the native state topology is therefore significant. To assess the stability of the structures obtained at the end of the successful folding trajectories ($RMSD_N < 0.2$ Å), seven 100 ps unrestrained MD simulations were performed starting from the final conformation of some of the FR1 and FR2 folding trajectories. The conformation was stable ($RMSD_N$ for the C_α atoms less than 2 Å) during all unrestrained trajectories. These results and the aforementioned final values of the $RMSD_N$ show that the native state was reached.

Nine snapshots along one of the 1.2 ns FR1 folding trajectories are shown in Figure 1. The sequence of secondary and tertiary structure formation is typical of all the folding simulations, apart from the majority of the FS2 runs where the α-helix forms after the β3–β4 sheet. There is transient formation of small non-native regular elements of secondary structure in the first half of the simulation (Fig. 1). The formation of the native α–helix is the first folding event, followed by formation of the β3–β4 contacts. Evidence from protein-engineering experiments indicates that the transition state is the same for folding and unfolding and that the α–helix is the only relatively well formed secondary structure in the transition state.[18] This is therefore in agreement with most of the folding and unfolding trajectories, for the latter the disruption of the contacts in the α–helix being the last unfolding event. The contact between Ala16 and Leu49, which is known experimentally to be crucial,[18] is formed for the first time in the sixth snapshot of Figure 1 (after 1 ns).

Time Evolution of Native Contacts

Fifty-two contacts have been chosen to describe the native state of CI2 and the analysis presented here is similar to previous simulation studies of CI2.[20] The evolution of the fraction of the native contacts (Q) as a function of the simulation time is depicted in Figure 2 for four FR2 folding trajectories. A similar behavior is found for all folding simulations. Fifty percent of the native contacts are formed at around 5 ns (FR2, FS1) and 1 ns (FR1, FS2). Abrupt transitions in the evolution of Q can be used to define transition states. Conformational changes occur fast at this state because it is a maximum of the free energy. In the simulation shown in the top left panel of Figure 2, such a transition occurs at $Q \approx 0.6$, whereas it happens at $Q \approx 0.3$ in the trajectory depicted in the bottom right panel. Two abrupt transitions can be seen in

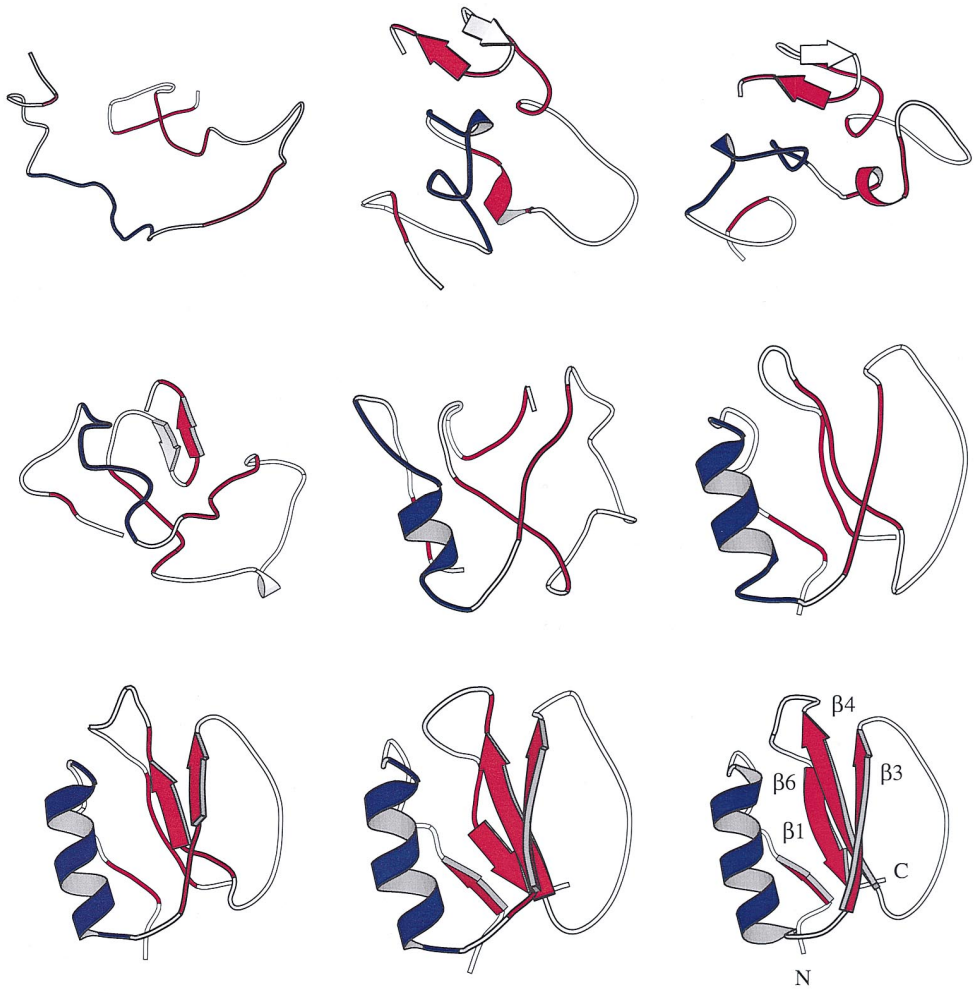

Fig. 1. Snapshots from one of the FR1 folding trajectories at, from left to right and top to bottom: 0, 200, 400, 550, 700, 1000, 1050, 1100, and 1200 ps. The secondary structure was assigned with the DSSP program[39] and plotted with MOLSCRIPT.[40] The following elements were defined for the native structure (bottom, right): strand β1, residues 3 to 5; α–helix, 12 to 24; strand β3, 27 to 33; loop, 34 to 45; strand β4, 46 to 52; strand β6, 59 to 63. In all frames, segments of the polypeptide chain which correspond to the α–helix (β–sheet) in the native fold are colored in blue (red).

the top right panel, whereas the curve in the bottom left panel shows a smoother behavior. This points out the difficulty of identifying a transition state from MD simulations. Itzhaki et al. have found that Ala16 is a nucleation site in the folding of CI2 and that the transition state is stabilized by interactions between the side-chain of Ala16 and residues of the hydrophobic core, mainly Leu49 and Ile57.[18] The first appearance of the contacts between these three residues is indicated in Figure 2 by small vertical lines (labeled 1 to 3). The first appearance of the Val13-Val51 contact is also shown (label 4). Experimental results from site-directed mutagenesis studies indicate that the structure of the transition state in the neighborhood of Val13, Ala16, and Leu49 is almost as folded as in the native state.[18] In three of the four folding simulations shown in Figure 2, all of the four contacts appear at low Q. The value of Q averaged over all the folding simulations is $Q = 0.32 \pm 0.27$ (first appearance of contact 1), $Q = 0.45 \pm 0.28$ (contact 2), $Q = 0.46 \pm 0.29$ (contact 3) and $Q = 0.42 \pm 0.33$ (contact 4). For unfolding, the average value of Q is $Q = 0.24 \pm 0.14$ (last disappearance of contact 1), $Q = 0.35 \pm 0.16$ (contact 2), $Q = 0.20 \pm 0.15$ (contact 3) and $Q = 0.28 \pm 0.15$ (contact 4). This is also

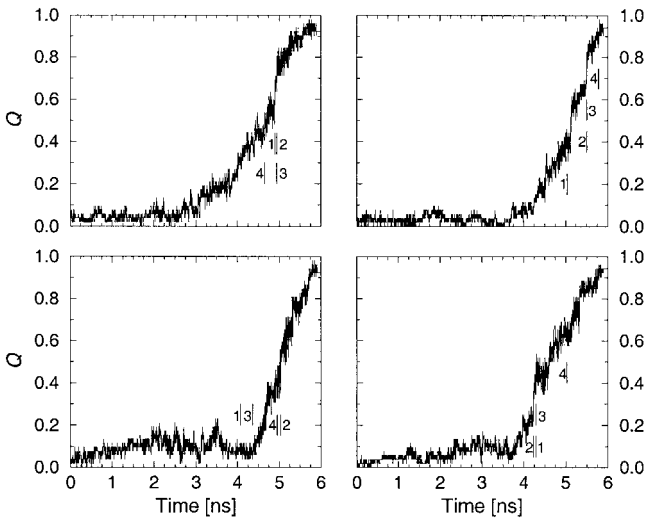

Fig. 2. Evolution of the fraction of the native contacts (Q) as a function of the simulation time for four FR2 folding trajectories. The first appearance of four important contacts is shown in the four panels by small vertical lines. Contact 1 is between Ala16 and Leu49, contact 2 between Ala16 and Ile57, contact 3 between Leu49 and Ile57, and contact 4 between Val13 and Val51.

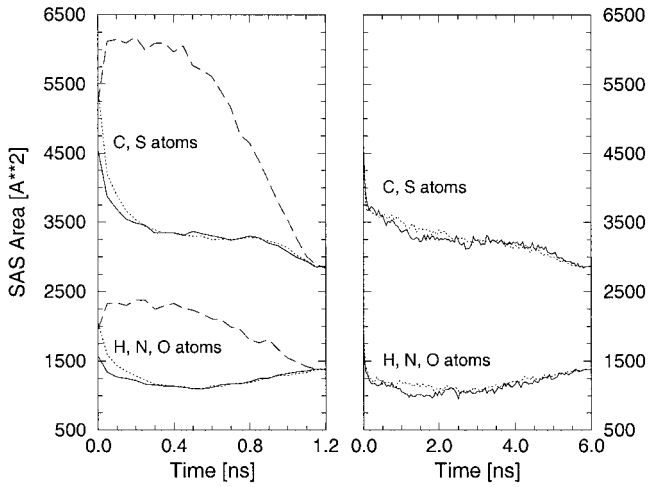

Fig. 3. Hydrophobic (upper curves) and hydrophilic (lower curves) solvent-accessible surface area as a function of time averaged over the FR1 (left, solid lines), FS2 (left, dotted lines), FR2 (right, solid lines), FS1 (right, dotted lines), and soft-repulsive-sphere model (left, dashed lines) simulations which converged. The standard deviations are comparable in all the trajectories and are around ± 150 Å^2.

consistent with the experimental data which indicate that most of the secondary and tertiary structure of the native state is not yet formed in the transition state.

Initial Collapse

In the folding run shown in Figure 1, the solvent-accessible surface decreases from 7,200 Å^2 to 4,570 Å^2 (this value is around 4,200 Å^2 for the native state) during the first 200 ps. This is due mainly to the reduction of the hydrophobic (C and S atoms) surface area from 5,390 Å^2 to 3,320 Å^2. Figure 3 shows the evolution of the hydrophobic and hydrophilic (H, N, and O atoms) solvent-accessible surface area averaged over all the folding trajectories which converged. It is evident that the burial of hydrophobic surface is the main driving force in protein folding, as suggested by Kauzmann[30] and observed recently in a 200-nanosecond molecular dynamics simulation of the early phase of folding of villin headpiece subdomain.[31]

To check whether this collapse originates from the interactions between the atoms or from the usage of the harmonic restraint, six additional folding simulations of 1.2 ns each were performed with a modified potential and the harmonic restraint. These trajectories were initiated from the six conformations which had led to the successful FR2 folding simulations. The electrostatic, attractive van der Waals, and solvation energy terms were turned off (soft-repulsive-sphere model). Practically, all the charges were set to zero and the van der Waals interactions were multiplied by $(1 - (r/r_c)^2)^2$ with $r_c = 3$ Å (see Paci and Karplus[28]). For the five trajectories which reached the native conformation, the evolution of the hydrophobic and hydrophilic solvent-accessible surface area is depicted in Figure 3 (dashed lines in left panel). The behavior is significantly different from the one with the full potential. This clearly shows that the initial collapse observed in the simulations with the full potential is not a result of the harmonic restraint, it rather originates from the interatomic interactions.

Energetics

The average value of the harmonic restraint (1.6 ± 0.3 kcal/mol and 1.4 ± 0.1 kcal/mol for folding and unfolding, respectively) is negligible compared to the energy terms of the force field. This shows that the harmonic restraint does not bias significantly the energetics of the folding/unfolding process. The effective energy (intraprotein en-

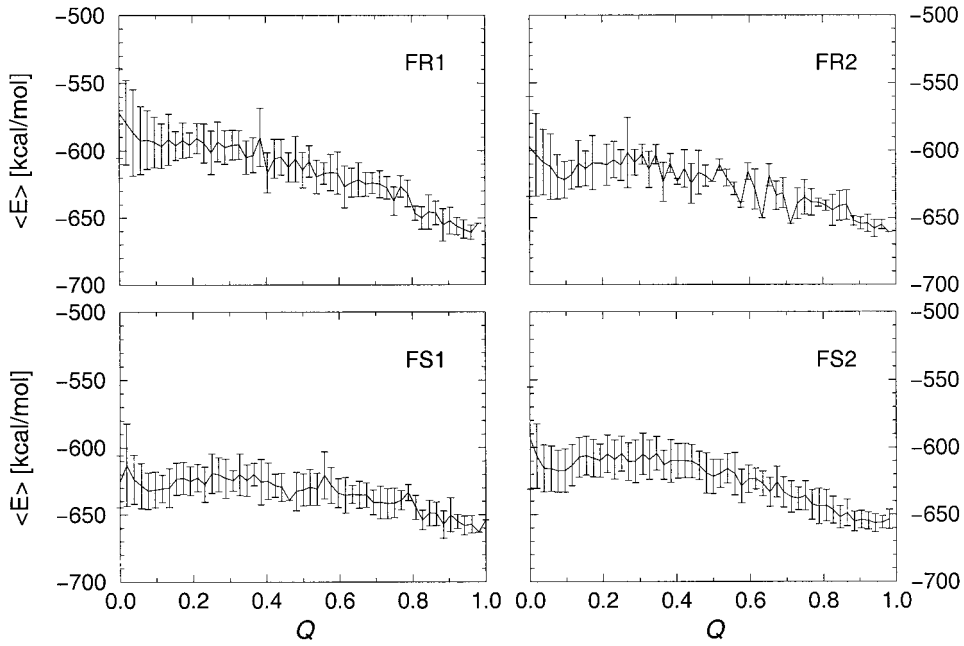

Fig. 4. Average effective energy, <E>, as a function of the fraction of the native contacts Q for the folding simulations which converged. Sixty (120) conformations were selected from each 0.3 ns or 1.2 ns (6 ns) trajectory. They were submitted to a 10 ps unrestrained MD run at 300 K, followed by 300 steps of energy minimization, before evaluation of the energy. Bars = 2 SD.

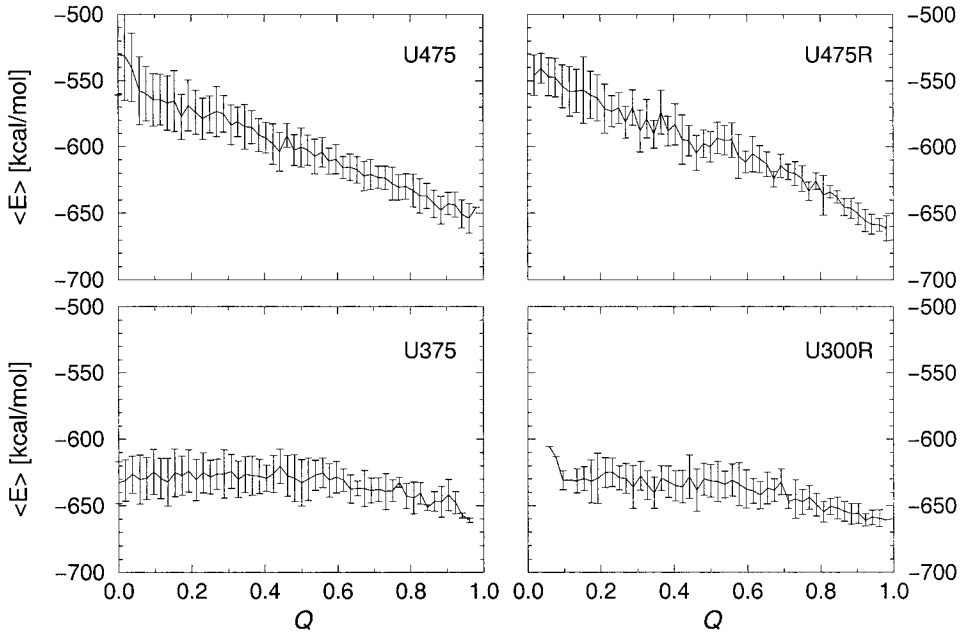

Fig. 5. Average effective energy, <E>, as a function of the fraction of the native contacts Q for the unfolding simulations. See Figure 4 for the description of the procedure used.

ergy and mean solvation energy, averaged over the MD runs) as a function of Q is shown in Figure 4 for the folding simulations which converged and in Fig. 5 for all unfolding simulations. The general shape for the FR2, FS1, and FS2 trajectories is comparable to the one of the unfolding simulations at 300 K (U300R in Fig. 5). On the contrary, the energy surface of the FR1 simulations shows a more funnel-like shape, especially at low Q. This is probably a consequence of the shorter simulation times (1.2 ns for FR1 instead of 6 ns for FR2 and FS1). In fact, the FR2 simulations have a flatter energy landscape for $Q < 0.4$ and their starting conformations are not lower in energy on average than the initial FR1 structures of which they are a subset. Although the FS1 simulations were initiated from a U475 conformation, they have lower energies, because they had enough time to equilibrate.

In the unfolding simulations, the average effective energy depends strongly on the temperature. This has been found previously in lattice simulations.[3,32] The funnel-like behavior at 475 K (Fig. 5) is in agreement with previous results obtained by Lazaridis and Karplus, who performed twenty-four unfolding MD simulations of CI2 at 500 K with an implicit solvation model different from the one used here. Their solvation energy is proportional to the volume of groups of atoms,[20,23] whereas ours is related to the surface area of atoms. The plot of the effective energy as a function of Q for the U475R simulations is similar to the one of the U475 trajectories. This is a further support for the negligible role of the harmonic restraint on the unfolding energetics.

The average effective energy obtained from the unfolding simulations at 300 K and 375 K differs significantly from the one at 475 K. The effective energy is nearly flat, except for the basin at $Q > 0.7$ which corresponds essentially to the native state. The funnel-like shape at 475 K is due to the high conformational entropy (not included in the effective energy) which allows the system to explore regions with high effective energy. This is not the case at 300 K and 375 K (flat profile of the effective energy). 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 $Q > 0$ the effective energy is lower at 300 K than at 475 K. Even at 375 K, where the final conformations are as unfolded in term of Q as at 475 K, the energies are significantly lower than at 475 K. The structures obtained upon unfolding show smaller radii of gyration at 300 K and 375 K than at 475 K (not shown). The rather compact unfolded state up to 375 K is in agreement with thermal denaturation experiments which usually produce only collapsed conformations that can be further unfolded by strong denaturants.[33] As a result, the accessible conformational space is greatly reduced at 300 K, as demonstrated by lattice statistical mechanics[34] and lattice simulations.[6] Nevertheless, the form of the energy landscape at 300 K cannot explain fast folding (see also Dobson et al.[3]). The effective energies at 375 K and 300 K are similar. Therefore, this model suggests that the difference in the conformational entropy between 300 K and 375 K is small.

Sequence of Events

The folding and unfolding trajectories show significant diversity, yet a statistically preferred path emerges from the simulation results. The diversity is reflected in the RMSD averaged over the conformations with $Q = 0.25$ obtained during unfolding (folding) which has a value of 9.1 Å (4.9 Å). It is 6.2 Å (3.3 Å) at $Q = 0.50$, and 3.3 Å (1.9 Å) at $Q = 0.75$. The evolution of the native contacts as a function of the C_α RMSD from the native state is depicted in Figures 6 and 7 for the folding and unfolding simulations, respectively. The similarity between the results provided by the U475 and U475R trajectories suggests

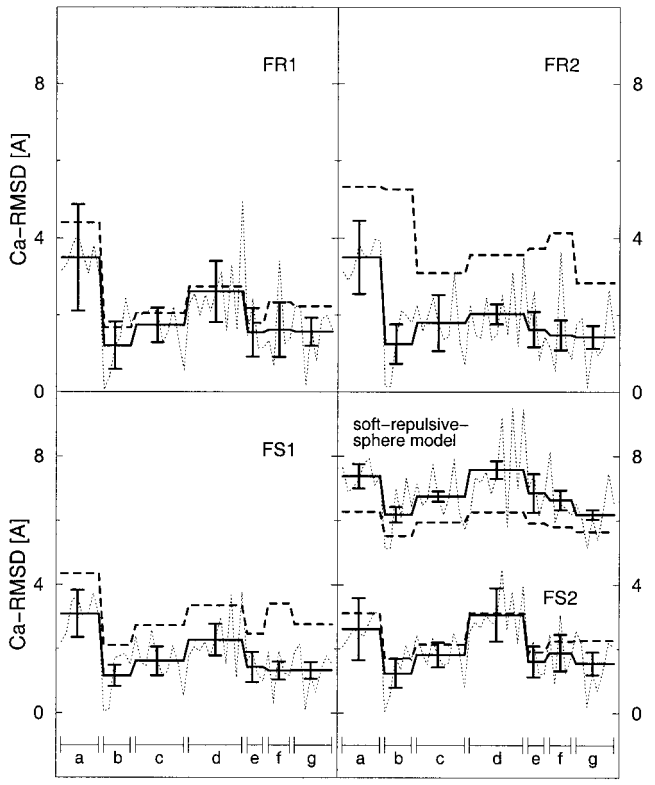

Fig. 6. The 52 native contacts were classified in seven groups according to their location in elements of secondary structure:[20] a, α-helix; b, α-helix–β1; c, β1–β4 and β1–β6; d, β3–β4; e, β4–β6; f, loop; g, α-helix–β3, α-helix–β4 and α-helix–β6. The thin dotted lines represent the C_α–RMSD (Å) from the native state at the last disappearance of the native contacts. Averages over each group of contacts for the last disappearance are in thick lines, and for the first appearance in thick dashed lines. The results using the soft-repulsive-sphere model are shown in the bottom-right panel with an offset of 5 Å. Only folding simulations which converged were used. The first appearance of a contact is defined as the first 20 ps interval during which that contact is present, and the last disappearance as the last 20 ps interval during which it is disrupted. 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 last disappearance; bars = 2 SD.

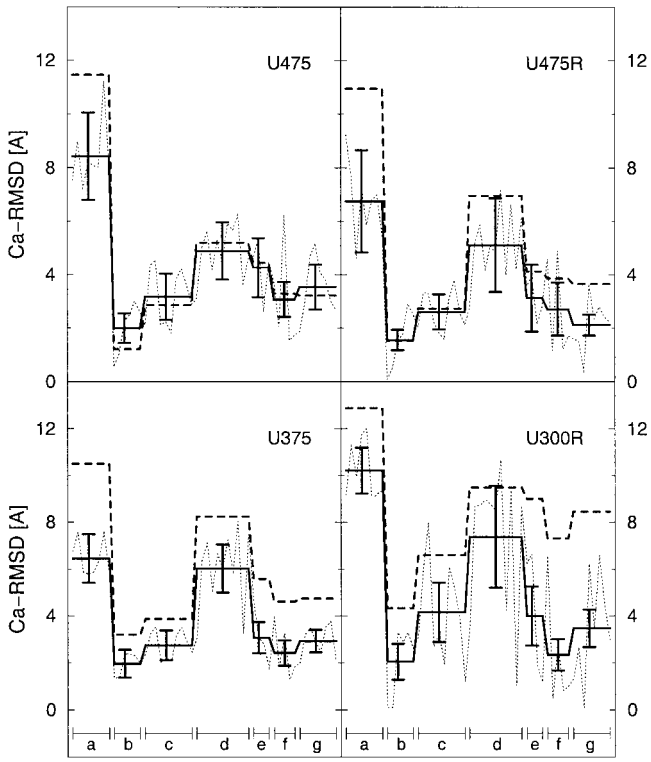

Fig. 7. The thin dotted lines represent the C_α–RMSD (Å) from the native state at the first disappearance of the native contacts. Averages over each group of contacts for the first disappearance are in thick lines, and for the last appearance in thick dashed lines. The standard deviations for each group of contacts are shown for the first disappearance; bars = 2 SD. See Figure 6 for the definition of the contact groups, first disappearance, and last appearance.

that the harmonic restraint has no major impact on the sequence of events upon unfolding. Interestingly, the sequence of events is similar in all simulations and is comparable to previous unfolding simulations of CI2 performed with different force fields.[19,20] The formation of the α–helix is the primary folding event (see also Fig. 1), and the appearance of tertiary interactions between the α–helix and the first strand of the β–sheet is the last folding event. A preferred pathway seems to emerge from the folding simulations (Fig. 6), although folding events are found to be closer to the native state than unfolding events. This is probably due to the non-adiabatic acceleration of the events by TMD, which introduces hysteresis in the system, but is not expected to affect the sequence of events (see also below).

The comparison with the results of Lazaridis and Karplus[20] shows that the β3–β4 sheet unfolds earlier in the present work, in agreement with the results of Li and Daggett, who performed unfolding simulations of CI2 at 498 K using explicit water molecules.[19] The sequence of events obtained during folding with the soft-repulsive-sphere model is somewhat similar to the one with the full potential. This suggests that the folding process is mainly determined by the native state topology. The main difference between simulations performed with the full force field and the soft-repulsive-sphere model is that in the latter the α–helix and β3–β4 strands fold simultaneously and are the first folding events as in FS2 and previous simulations.[20] The difference between the first disappearance and the last appearance indicates that the native contacts can reform after having been broken. Whereas this is only seen for the α–helix in the U475 simulations, it is observed for most of the native contacts in the U375 and U300R trajectories. In this respect and with respect to the effective energy, the results provided by the U375 simulations agree better with the U300R than with the U475 results, which further supports that the bias originating from the harmonic restraint is weak.

CONCLUSIONS

TMD found a folding path in 50% of the simulations started from random conformations of CI2 and in 83% of the runs which were initiated from conformations un-

folded by high temperature MD. Critical tests were performed to understand the effect of the harmonic restraint. They consisted of unfolding simulations at elevated temperatures (375 K and 475 K) without the harmonic restraint, and folding simulations with the harmonic restraint and a soft-repulsive-sphere interaction potential. The shape of the average effective energy depends strongly on the simulation temperature and partly on the simulation time. The funnel-like shape observed to some extent in the 1.2 ns folding simulations started from random conformations might be an equilibration effect. At 300 K, the surface of the effective energy is almost flat for conformations with less than about 50% of the native contacts formed, and therefore cannot explain fast folding. However, lattice-based simulations have shown that at low temperatures a rather flat energy landscape leads to a finite folding time which is much shorter than expected from a random search in the conformational space.[3] This mainly originates from a restriction of the accessible conformational space even for structures with few native contacts.[3,32]

The simulation results for CI2 indicate that the folding and unfolding pathways agree in terms of energetics and sequence of events. Despite a significant diversity in the pathways, a statistically preferred sequence of events emerges when the simulations are analyzed in terms of native-like contacts, in agreement with previous unfolding simulations.[20] The first event in the folding of CI2 (last in unfolding) is the formation of the α–helix and the last event (first in unfolding) is the formation of contacts between the α–helix and the β1 strand. The agreement between the sequence of folding and unfolding events, and the fact that three different force fields[19,20] and an oversimplified potential (soft-repulsive-sphere model) led to similar unfolding/folding pathways suggest that, at least for CI2, the folding process is tied to some extent to the topology of the folded state. This is consistent with experimental results on the src and α-spectrin SH3 domains,[35,36] (see also Alm and Baker[5]), as well as with MD unfolding simulations of the src SH3 domain,[37] and with a theoretical study of CI2 and barnase.[38] Previously, the shape of the free energy landscape as a function of the radius of gyration and number of native contacts was found by MD simulations to depend strongly on the native topology.[10,11] Our results go beyond previous findings and indicate that not only the energy landscape but also the sequence of events is determined by the native topology. If the topology plays a dominant role in the folding mechanism, it should be possible to understand protein folding with the help of simple physical principles.

ACKNOWLEDGMENTS

We thank Prof. M. Karplus, Dr. E. Paci, and Prof. E. Shakhnovich for interesting discussions and helpful comments. We also thank Prof. T. Lazaridis for providing us with the partial charges of the ionizable side chains 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

1. Onuchic JN, Luthey-Schulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 1997;48:545–600.
2. Dill K, Chan H. From Levinthal to pathways to funnels. Nat Struct Biol 1997;4:10–19.
3. Dobson CM, Sali A, Karplus M. Protein folding: A perspective from theory and experiment. Angew Chem Int Ed 1998;37:869–893.
4. Dobson CM, Karplus M. The fundamentals of protein folding: bringing together theory and experiment. Curr Opin Struct Biol 1999;9:92–101.
5. Alm E, Baker D. Matching theory and experiment in protein folding. Curr Opin Struct Biol 1999;9:189–196.
6. Šali A, Shakhnovich E, Karplus M. How does a protein fold? Nature 1994;369:248–251.
7. Dill KA, Bromberg S, Yue K, et al. Principles of protein folding—a perspective from simple exact models. Protein Sci 1995;4:561–601.
8. Skolnick J, Kolinski A. Simulations of the folding of a globular protein. Science 1990;250:1121–1125.
9. Finkelstein AV. Can protein unfolding simulate protein folding? Protein Eng 1997;10:843–845.
10. Boczko EM, Brooks III CL. First-principles calculation of the folding free energy of a three-helix bundle protein. Science 1995; 269:393–396.
11. Sheinermann FB, Brooks III CL. Calculations on folding of segment B1 of streptococcal protein G. J Mol Biol 1998;278:439–456.
12. Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 1998;282:740–744.
13. Schlitter J, Engels M, Krüger P, Jacoby E, Wollmer A. Targeted molecular dynamics simulation of conformational change: Application to the T ↔ R transition in insulin. Mol Simul 1993;10:291–309.
14. Diaz J, Wroblowski B, Schlitter J, Engelborghs Y. Calculation of pathways for the conformational transition between the GTP- and GDP-bound states of the Ha-*ras*-p21 protein: calculations with explicit solvent simulations and comparison with calculations in vacuum. Proteins 1997;28:434–451.
15. Ma J, Karplus M. Molecular switch in signal transduction: reaction paths of the conformational changes in *ras* p21. Proc Natl Acad Sci USA 1997;94:11905–11910.
16. Apostolakis J, Ferrara P, Caflisch A. Calculation of conformational transitions and barriers in solvated systems: application to the alanine dipeptide in water. J Chem Phys 1999;110:2099–2108.
17. Jackson SE, Fersht AR. Folding of chymotrypsin inhibitor 2. 1. evidence for a two-state transition. Biochemistry 1991;30:10248–10435.
18. Itzhaki LS, Otzen DE, Fersht AR. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: evidence for a nucleation-condensation mechanism for protein folding. J Mol Biol 1995;254:260–288.
19. Li A, Daggett V. Identification and characterization of the unfolding transition state of chymotrypsin inhibitor 2 by molecular dynamics simulations. J Mol Biol 1996;257:412–429.
20. Lazaridis T, Karplus M. "New view" of protein folding reconciled with the old through multiple unfolding simulations. Science 1997;278:1928–1931.
21. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. J Comput Chem 1983;4: 187–217.
22. Neria E, Fischer S, Karplus M. Simulation of activation free energies in molecular systems. J Chem Phys 1996;105:1902–1921.
23. Lazaridis T, Karplus M. Effective energy function for proteins in solution. Proteins 1999;35:133–152.
24. Eisenberg D, McLachlan AD. Solvation energy in protein folding and binding. Nature 1986;319:199–203.
25. Hasel W, Hendrickson TF, Still WC. A rapid approximation to the solvent accessible surface areas of atoms. Tetrahedron Comput Methodol 1988;1:103–116.

26. Fraternali F, van Gunsteren WF. An efficient mean solvation force model for use molecular dynamics simulations of proteins in aqueous solution. J Mol Biol 1996;256:939–948.
27. Eckart C. Some studies concerning rotating axes and polyatomic molecules. Phys Rev 1935;47:552–558.
28. Paci E, Karplus M. Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. J Mol Biol 1999;288:441–459.
29. Marchi M, Ballone P. Adiabatic bias molecular dynamics: a method to navigate the conformational space of complex molecular systems. J Chem Phys 1999;110:3697–3702.
30. Kauzmann W. Some factors in the interpretation of protein denaturation. Adv Protein Chem 1959;14:1–64.
31. Duan Y, Wang L, Kollman PA. The early stage of folding of villin headpiece subdomain observed in a 200-nanosecond fully solvated molecular dynamics simulation. Proc Natl Acad Sci USA 1998;95: 9897–9902.
32. Karplus M. The Levinthal paradox: yesterday and today. Folding and Design 1997;2:S69–S75.
33. Ptitsyn OB. Molten globule and protein folding. Advan Protein Chem 1995;47:83–230.
34. Dill KA. Theory for the folding and stability of globular proteins. Biochemistry 1985;24:1501–1509.
35. Grantcharova VP, Riddle DS, Santiago JV, Baker D. Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. Nat Struct Biol 1998;5:714–720.
36. Martinez JC, Pisabarro MT, Serrano L. Obligatory steps in protein folding and the conformational diversity of the transition state. Nat Struct Biol 1998;5:721–729.
37. Tsai J, Levitt M, Baker D. Hierarchy of structure loss in MD simulations of src SH3 domain unfolding. J Mol Biol 1999;291:215–225.
38. Micheletti C, Banavar JR, Maritan A, Seno F. Protein structures and optimal folding from a geometrical variational principle. Phys Rev Lett 1999;82:3372–3375.
39. Kabsch W, Sander C. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 1983;22:2577–2637.
40. Kraulis P. Molscript, a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 1991;24: 946–950.