PROTEINS: Structure, Function, and Genetics 51:245–257 (2003)

Kinetics of Cytochrome C Folding: Atomically Detailed Simulations

Alfredo E. Cárdenas and Ron Elber*
Department of Computer Science, Cornell University, Ithaca, New York

ABSTRACT **The vast range of time scales (from nanoseconds to seconds) during protein folding is a challenge for experiments and computations. To make concrete predictions on folding mechanisms, atomically detailed simulations of protein folding, using potentials derived from chemical physics principles, are desired. However, due to their computational complexity, straightforward molecular dynamics simulations of protein folding are impossible today. An alternative algorithm is used that makes it possible to compute approximate atomically detailed long time trajectories (the Stochastic Difference Equation in Length). This algorithm is used to compute 26 atomically detailed folding trajectories of cytochrome c (a millisecond process). The early collapse of the protein chain (with marginal formation of secondary structure), and the earlier formation of the N and C helices (compare to the 60's helix) are consistent with the experiment. The existence of an energy barrier upon entry to the molten globule is examined as well. In addition to (favorable) comparison to experiments, we show that non-native contacts drive the formation of the molten globule. In contrast to popular folding models, the non-native contacts do not form off-pathway kinetic traps in cytochrome c. Proteins 2003;51:245–257.**

Key words: stochastic difference equation; molecular dynamics; protein folding; computer simulations

INTRODUCTION

The mechanism by which a protein chain finds its correct three-dimensional structure from the enormous number of alternative conformations available to it has attracted the interest and imagination of many biophysicists.[1] Plausible mechanisms, such as hydrophobic collapse[2] (forming first a compact state with little or no secondary structure), and the framework model[3] (early formation of stable helices) have been proposed. Atomically detailed computer simulations, in conjunction with experiments, hold the promise to a comprehensive spatial and temporal view of the process[4–6] and to a test of the above hypotheses on concrete examples.

Cytochrome c is a small protein (104 amino acids) that was studied extensively in the past. Its known high-resolution structure[7] has three well-defined secondary structure elements: the N, C, and the so-called 60's helix. The helices are packed around a heme group that is covalently linked to the protein chain at residues Cys 14 and 17 and provides a useful spectroscopic probe.

A great wealth of experimental information is available for cytochrome c on different folding steps and time scales.[8] Experiments probe the initial collapse of the protein chain,[9] the early formation of secondary structure,[10] and intermediates induced by heme misligation.[11]

Hydrogen exchange experiments indicate that the N and C helices are the first to form.[12] Trapping of folding intermediates with cobalt substitutions of the heme iron also indicates that the N/C helices fold early.[11] This is one of the observables that is computed in this study and compared to the experimental data.

The radius of gyration of the molten globule was estimated as 18Å[13] based on X-ray scattering experiments. Hence, the radius of gyration of the molten globule phase is only 30% larger than the radius of gyration of the native structure.[13] A complementing two-dimensional view of the folding landscape was suggested by simultaneous analysis of X-ray scattering and circular dichroism (CD) measurements.[14,15] A mechanism was proposed in which chain collapse occurs before significant formation of secondary structure. Our simulations suggest an alternative (kinetic) definition of the molten globule that is consistent with the above experimental estimate. We examine the two distinct kinetic phases that are observed in the calculations: one phase of compactness with no formation of secondary structure, and a second phase of significantly slower compression but with secondary structure formation. The onset of the second phase can be used to define the molten globule state.

Intermediates along the folding pathways were suggested to be protein conformations with misligated heme (instead of the native link to Met 80, a bond to one of the histidines 26 or 33 is formed).[16–18] This is a slow process as suggested by comparing the overall collapse time (submilliseconds) with the rate of folding misligated proteins (milliseconds and more).[19] Interestingly, Hagen et al.[20] estimated the ratio of misligation to histidine 26 and histidine 33 to be about 1. However, equilibrium measure-

Grant sponsor: NIH; Grant sponsor: NSF.

*Correspondence to: Ron Elber, Department of Computer Science, 4130 Upson Hall, Cornell University, Ithaca, NY 14853.
E-mail: ron@cs.cornell.edu

Received 26 September 2002; Accepted 15 November 2002

ments[11] of misligation at the unfolded state estimated a ratio of 1:3. Here, we present a crude calculation of the ratio, which is based on collision frequency, and is more consistent with the estimate of Hagen.[20]

Finally, it was further suggested for cytochrome c that an energy barrier exists at the entrance to the collapsed state, based on the temperature dependence of the rate.[9,21] We have examined this observation as well; however, the results are inconclusive. The numerical noise of the sparse data makes it difficult to reach quantitative conclusions on the barrier height.

Mechanisms of protein folding were simulated extensively in the past using reduced or simplified models.[2,22–24] These studies help develop a general understanding of protein folding and outline many important principles. However, the simplifications make it difficult to propose concrete predictions to be tested on a specific molecule. For example, weighting alternate folding mechanisms (secondary structure formation vs. hydrophobic collapse) depends on the strength of detailed interactions. Atomically detailed simulations of the type reported here can supplement general models by adding a specific and detailed view of folding mechanisms.

Unfortunately, the extended temporal scale of cytochrome c folding (milliseconds) is too long compared to the temporal scales of molecular dynamics simulations that are limited to nanoseconds (six orders of magnitude shorter than the folding time). This limitation, which is general for protein folding and is not specific to cytochrome c, motivated the use of approximate molecular dynamics techniques.

Past atomically detailed simulations of protein folding with broadly tested chemical-physics potentials either probe only early events[5] or were based on one of the following approximations: (1) the use of high temperature *unfolding* trajectories[25] and (2) the construction of free energy profiles along pre-specified reaction coordinates.[26] The comparison to experimental data (when available) was encouraging.[4,6] Examples of the high-temperature trajectories suggest that the effects on the pathways are small.[25] However, these approaches are not simulations of the *folding dynamics* that can proceed along unknown temperature-dependent pathways. It is, therefore, desired to have alternative simulation techniques that do not use the above assumptions. One interesting approach that probes folding dynamics directly is using massively distributed computing.[57]

Here, we discuss such an alternative approach that computes room-temperature atomically detailed folding trajectories without assuming a reaction coordinate or an equilibrium state. The new method was used already to investigate a number of biological systems: C peptide[27] folding, the R to T transition in hemoglobin,[28] an enzymatic reaction,[29] ion permeation through the gramicidin channel,[30] and the folding of a small helical protein (protein A).[31]

Our approach (as applied to the folding of cytochrome c) is also approximate with respect to a most detailed and accurate classical trajectory: (1) we modeled the water solvation by a continuum (the Generalized-Born model [GB] of Hawkins et al.[32]; the GB code that we implemented into our program MOIL is of Tsui and Case[33]) and (2) we filter out from the trajectories motions that are rapid on the scale of the step.[34,35] The rapid motions are modeled as noise. These approximations (that are very different from those that were used in the past) are likely to fail when: (1) structured water molecules are making significant contribution to the kinetics of folding, a contribution that cannot be captured with a continuum model; and (2) the (filtered) rapid vibrations strongly influence slow modes; the filtered trajectory clearly misses their contributions to the dynamics.

Simulations were also carried out with explicit water solvation (i.e., not using the first approximation). In our studies of C peptide folding,[27] hemoglobin R to T transition,[28] and the permeation of ions through the gramicidin channel,[30] we have used *explicit* water molecules to obtain average solvation forces for intermediate structures. In that procedure, the solute coordinates were moving adiabatically in the solvent field in the spirit of the Carr and Parrinello procedure.[36] The calculations with explicit water molecules are, however, significantly more expensive (by a factor of about 50 compared to calculations with implicit solvation; Ron Elber, unpublished data). For protein folding, the computational saving of the implicit solvation model is extremely helpful, and made the current study feasible.

The present implementation of this algorithm was briefly discussed in Ghosh et al.[31] These authors described the methodology and an application to the folding of protein A. Here, we provide a more detailed discussion of the algorithm and describe an application to a larger protein (cytochrome c). In contrast to protein A, the folding kinetics of cytochrome c was studied extensively by experiments. This makes cytochrome c a better candidate for testing our model against experimental data.

THE ALGORITHM

Recently, we have introduced an algorithm to calculate approximate classical trajectories at extended time scales. The method is called SDE (Stochastic Difference Equation) and evolved through a series of publications.[31,35,37] Here, we present the general ideas and emphasize the modifications that brought us to the current implementation that we call SDEL (Stochastic Difference Equation in Length).

The SDE is different from the usual molecular dynamics approach in a number of ways. Instead of solving an initial value differential equation (The Newton's law) we optimize an action, which is a functional of the trajectory. A stationary solution of the functional (a trajectory) provides the desired classical path. The solution of the Newton's equations of motion also implies a solution of the trajectory as a function of time, $X(t)$. In the last few years, we computed long-time trajectories with the SDE where the classical path was parameterized as a function of time[34] similarly to the Newton's equation.

Here we deviate from some of our earlier studies[27,28,30] and parameterize the trajectories as a function of the path

length. There are a number of reasons given below why the length formulation is preferred. We first describe the alternative parameterization.

We start with the principle of minimal action

$$S = \int_{X_u}^{X_f} \sqrt{2(E - U)}dl \tag{1}$$

The functional S is the classical action, X_u and X_f are the (mass weighted) coordinates of the unfolded and the folded states respectively, E is the total energy, U is the potential energy of the system, and dl is a mass-weighted length element of the path. We seek a trajectory that will make the action S stationary, given the two (fixed) end points (X_u and X_f), and the total energy E. We consider a discrete version of the above action $S \equiv \bar{S} \equiv \sum\sqrt{2(E-U_i)}\Delta l_{i,i+1}$. For a sufficiently small step $\Delta l_{i,i+1}$ ($\Delta l_{i,i+1} = |X_i - X_{i+1}|$) we will obtain a sound approximation to the classical action. A sequence of structures $\{X_1^{\min}, \ldots, X_N^{\min}\}$ that makes $\bar{S}$ stationary is determined by optimization. The optimized set of coordinates represents different slices of the trajectory as a function of the length index. The (fixed) end points are the unfolded and folded structures, respectively ($X_0 \equiv X_u$ and $X_{N+1} \equiv X_f$).

Since the action is not necessarily a minimum (the principle of least action applies only for the first order variation, $\delta S/\delta X = 0$) we use the norm of the action gradient as a target for optimization:

$$T = \sum(\partial S/\partial X_i)^2 + \lambda(\Delta l_{i,i+1} - \langle\Delta l\rangle)^2 \tag{2}$$

The second term on the right-hand side is a penalty function introduced for numerical convenience (λ is an empirical constant). It forces the points to be equally distributed along the trajectory $\left(\langle\Delta l\rangle = \frac{1}{N+1}\sum\Delta l_{i,i+1}\right)$. There is a total of $N + 1$ distances since X_u and X_f also contribute to equation (2).

A quick look at the boundary value formulation of classical mechanics suggests that it is considerably more expensive to compute than the initial value approach. Instead of considering one or two structures at a time, as in the initial value formulation, the whole trajectory is kept and manipulated during the optimization of the target function. Hence, a much larger system compared to the initial value formulation (larger by a factor of the number of steps) is considered in the boundary value approach. It is, therefore, not surprising that this approach is significantly less popular than the direct solution of the Newton's equation.

While exact trajectories in a boundary value formulation were computed in the past,[38,39] the implementations were not aimed at large systems. Instead, the focus was on cases in which fixed boundary conditions were crucial for the problem at hand. For example, semi-classical approximations of the scattering matrix[40] require a state-to-state classical trajectory. State-to-state trajectories are more easily obtained using a boundary value formulation since it is difficult to ensure that a trajectory will terminate at a desired state if only initial conditions are specified.

The promise of the functional formulation for complex systems is not in producing trajectories of the same quality as the initial value approach, but in providing an easy framework for introducing systematic approximations. Such systematic approximations (variable step sizes) were difficult to find in the initial value approach, though significant research was done along that direction.[41–44]

Large time steps and their consequences in SDE were discussed in earlier reports.[27,34] It was shown that their effect is to filter high-frequency modes from the trajectories (motions with frequencies ω larger than $\pi/\Delta t$, Δt – the time step, are removed adiabatically). A similar filtering of high-frequency modes exists also for the length formulation[35,37] that we use here. To account for some of the effects of the high-frequency motions that we miss after the filtering, we modeled them statistically as Gaussian noise.[35]

Based on numerical experimentations and the above modeling of the high-frequency motions as Gaussian noise, we developed a statistical theory for the trajectories. The theory identifies most probable trajectories as optimized solutions of the target function T with arbitrary step size. When all or most of the frequencies are filtered out (the step is very large), the difference between the optimized (most probable) path and the exact trajectory is dominated by the "noise" term. When using large steps, it is, therefore, necessary to consider an ensemble of plausible trajectories sampled with "noise levels" appropriate for the approximation at hand. A single optimal trajectory is insufficient in the time formulation.

In contrast (which is one of the advantages of the length formulation), the optimized trajectory with a large length step has a simple physical interpretation. It is the steepest descent path (Elber, forthcoming; and Ref.[35]), a path that is widely used in chemical physics as a reaction coordinate of molecular processes.[45] At intermediate step lengths, approximate trajectories that interpolate between the steepest descent path and an exact classical trajectory are computed. Since both limits have clear physical interpretation, it is possible to use the optimized trajectory directly, something that is harder to justify for the time formulation. We follow this approach here. Twenty-six optimized trajectories of cytochrome c folding, using 900 length slices, are computed and analyzed.

A second advantage of the length formulation is of energy conservation. The SDE formulation in the *time domain* does not conserve energy. The problem is the lack of an accurate estimate of the time derivatives of the coordinates. In the time formulation, we used a finite difference formula to calculate derivatives. As a result, a large time step provides a poor estimate of the velocities. The velocities are not required in the calculations but are necessary if we attempt to compute the kinetic energy, K, and the total energy, E. Others[46] formulate a variant of the SDE approach in which energy conservation was implemented as an external constraint in the time domain. The computational cost of the energy estimate is, however, high. In the length representation, the total energy is conserved by definition and does not require an external

constraint. The kinetic energy is computed without time derivatives (it depends only on one coordinate set ($K = E - U(X_i)$) and is not based on (inaccurate) numerical derivatives.

A third advantage of the length formulation is the homogenous distribution of points in *space*. In the time formulation, it is possible to miss a rapid transition between alternative minima even if the structural change is large. This is when the transition time scale is small compared to the time step. In the length parameterization, the spatial distance between sequential structures along the trajectory is kept constant. Hence, transitions that include large spatial changes (even if they are rapid in time) are presented consistently throughout the progress of the trajectory.

Returning back from advantages to technical considerations, we note that it is important to factor out from individual structures along the path their overall translations and rotations. This is since the distances $\Delta l_{i,i+1}$ are computed in Cartesian space. The minimization is, therefore, subject to the following linear constraints: $\sum_j x_{ij} = 0$ and $\sum_j x_{ij} \times x_{ij}^0 = 0$, x_{ij} are the (mass weighted) Cartesian coordinate vectors of atom j in structure i, and x_{ij}^0 are the coordinates of a reference structure that is used to determine the "laboratory" coordinate frame. The coordinates x_{ij}^0 are the initial coordinate set.

A final comment is concerned with the calculations of the total time. Our trajectories are parameterized as a function of the trajectory length. This parameterization fixes the total energy of the system, to be contrasted with the time formulation in which the total time of the trajectory is fixed. The energy is a property that can be estimated from equilibrium considerations and does not require prior knowledge on the timescale of the process. It, therefore, enables independent verification of the calculations by, for example, the computation of the reaction time scale.

In principle there is one-to-one correspondence between the length of the trajectory and its time. The following integral over the classical path relates the time and length $t = \int_0^L \frac{dl}{\sqrt{2(E-U)}} \approx \sum_i \frac{\Delta l_{i,i+1}}{\sqrt{2(E-U_i)}}$, where L is the total length of the trajectory. We note some expected differences between the two parameterizations: Processes that include transitions over barriers tend to be slow in time but are more rapid in length since the system is "sitting" at one position in space. On the other hand, downhill motions on an energy surface can occur rapidly in time but the spatial changes (length) may occur on a much longer scale. For a concrete example, consider chain collapse that is rapid in time; however, the length of the trajectory from the extended configuration to the collapsed state is long.

There is a subtle difficulty in numerical computations of the total time of a trajectory using the above formula. The filtering of high-frequency oscillations as a function of length makes the path considerably shorter compared to the exact trajectory.[37] An analogous problem can be found in the estimate of the length of the coast of England using yardsticks of different sizes.[47] In our case, the estimate of the total time depends on the size of the yardstick ($\Delta l_{i,i+1}$). The larger the length-step, the shorter is the observed time of the trajectory. The addition of high-frequency components (that we miss in the optimal trajectory) will make the trajectory longer and more similar to the exact classical path. We performed a fractal analysis for a trajectory of a short peptide in order to obtain an estimate of the time scale.[35] The calculations, however, were difficult to converge even for a short blocked peptide.

For the present study of a significantly more complex system than the blocked peptide, we consider only the length and discuss *the order of events* rather than the absolute time scale. We rely on the monotonic (but not linear) relationship between time and length given by the above formula. Hence, we anticipate correct temporal ordering of folding events, but we are uncertain about the absolute time scale.

COMPUTATIONAL PROTOCOL

In all the calculations, the molecular dynamics package MOIL was used.[48] The code is in the public domain and is available from http://cbsu.tc.cornell.edu/software/proteins_dynamics.htm

In addition to the usual energy, minimization and straightforward molecular dynamics algorithms, MOIL includes a parallel implementation (using MPI) of the Stochastic Difference Equation. The energy function in MOIL is the combination of AMBER[49] and OPLS[50] force fields.

The calculations of the SDE path require initial and final coordinate sets. In the present study, they are (1) an unfolded configuration, X_u, from a set of unfolded structures and (2) the folded structure, X_f. The preparation of both is discussed below.

The native structure of cytochrome c was taken from the protein data bank (1HRC). Crystallographic water molecules were removed and the energy of the protein was minimized using a conjugate gradient algorithm for 5,000 steps with the Generalized Born model mentioned earlier.[31–33] The model includes the covalent bonding of the heme group to Cys 14 and Cys 17. It does not include the link of the heme iron to Met 80 consistent with the unfolded state. Hence, we do not model the heme-Met bond formation in the folding process. We justify the simplification by the difference in time scales of bond formation and chain organization (the first is much longer[19]).

To generate unfolded configurations, ten straightforward Molecular Dynamics trajectories were run at 1,500 K. The simulation lengths were 4 ns each, and the structures were picked every 100 psec after 2 nsec of simulation. The run produced 200 unfolded structures that were minimized using a conjugated gradient algorithm. Of the 200 structures, 26 unfolded configurations with RMS distances of (at least) 12Å from each other were used to represent unfolded configurations.

Given the minimized unfolded and folded structures, minimum energy paths were calculated and used as initial guesses for the trajectories. First, we used the SPW (Self-Penalty Walk) functional[51] to compute minimum

energy paths with 300 grid points. Then, intermediate structures between existing length slices were added to increase the path representation to 900 structures. The paths were optimized for 10,000 steps. At the end of the calculations of the minimum energy paths, the RMS distance between sequential structures was of the order of 0.6–0.8Å.

The total energy of the system is required in the SDEL formulation and it was estimated from equilibrium runs at 300 K. It was the average energy of 20-psec equilibration runs at both the folded and unfolded states (E = −6,400 kcal/mol). A simulated annealing protocol was then used to minimize the target function T for the 900-structure paths. The simulated annealing was done in cycles of 200 steps and a total of 2,000 steps was used for each path. The typical gradient of the target function by the end of the run was about 10 *amu·kcal/mol/*Å. Further conjugate gradient optimization using an additional 1,000 steps did not change the coordinates in a significant way.

In the SDEL computations, a parallel code was used. The parallelization of the SDE algorithm is quite simple[28] since the coupling of different length slices (and the level of message passing between parallel processors) is small. From the perspective of communication overhead of a parallel computer, the worst-case scenario is having a single length slice on one processor, p_k. In this case, and independent of the total number of length slices, p_k communicates with four processors, ($p_{k\pm1}$, $p_{k\pm2}$). If the communication is made in parallel, then the communication is independent of the number of processors. Serially, it is proportional to the number of computing nodes. The last is the correct model for our typical computing environment: a cluster of PC-s.

We further note that the overall execution time is dominated by the calculation of the molecular forces and not by the communication. Even in a relatively slow network (like one finds in a typical cluster), the calculations were done efficiently, since the communication time is only a few percents of the total execution time.[28]

The calculations were done on both: (1) the Dell-Edge Windows-2000 cluster of the Cornell Theory Center (36 CPUs were used for a single path calculation), and (2) the authors' LINUX cluster with 45 CPUs for a single path calculation. It took about 48 h to optimize a single folding trajectory. The calculation of a single path on each of the above systems required about the same time since the Windows-2000 nodes (Intel P3 1 GHz) are faster than the LINUX CPUs, (Intel P3 600 MHz).

RESULTS

We present the results of the 26 folding trajectories that we computed. We describe individual trajectories but also present averages, and probability densities.

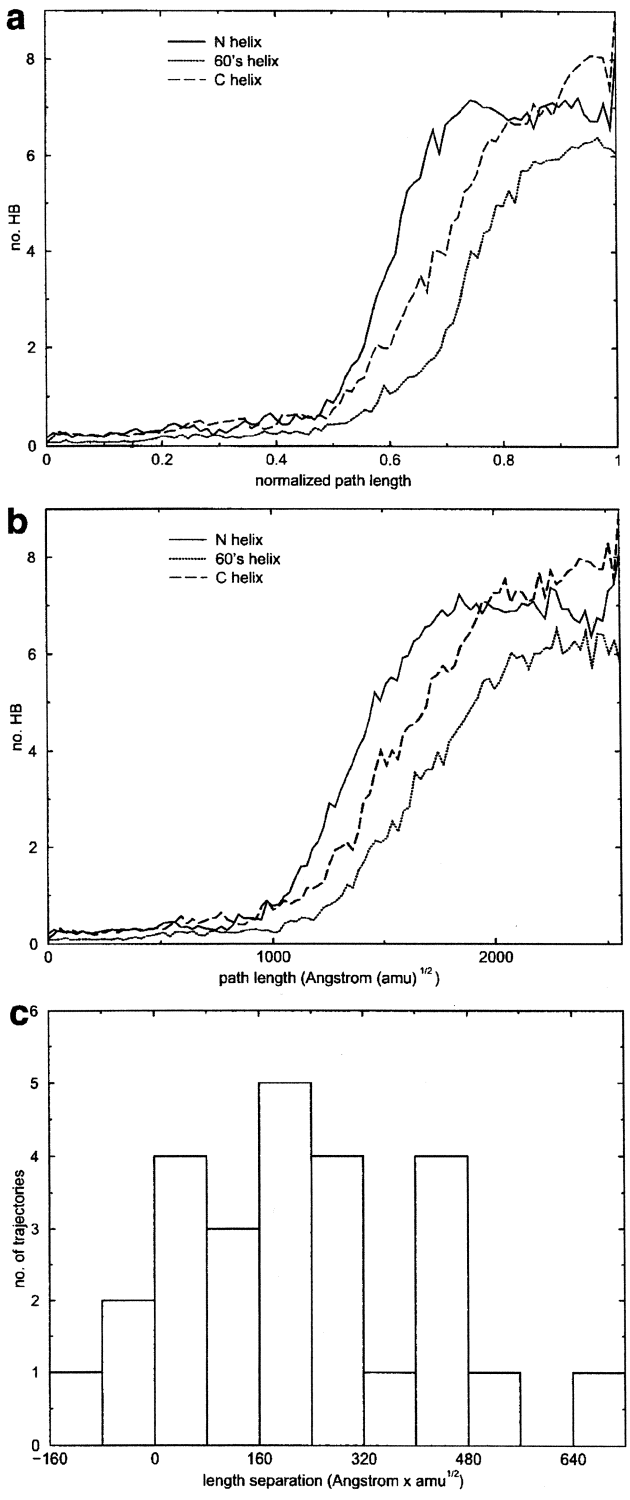

Fig. 1. The number of hydrogen bonds (HB) of a given helix (N, C, or the 60's helix) as a function of the path length. (A carbonyl oxygen and amide hydrogen form a hydrogen bond if their separation is 3Å or less.) The average of the 26 trajectories is reported. Two plots are shown. **a:** The path length is normalized to one and then the averages are performed. **b:** The absolute lengths of the paths are used. It is not obvious which one is a better measure of the properties of the average and, therefore, both are shown. Also shown **(c)** is the distribution of length intervals between the formation of the N and C helices and the formation of the 60's helix. The distribution is determined from the complete set of trajectories. Most of the length differences are positive indicating that the N and C helices form first.

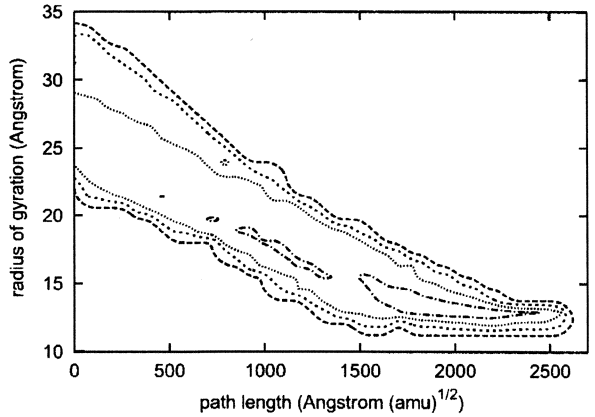

Fig. 2. A contour plot summarizing the folding progress along two "reaction coordinates": the radius of gyration and the length of the trajectory. The contours are the maximal values of the logarithm of the probability as a function of the above two variables. Note that the parameterization of a reaction coordinate as a function of length is perhaps the least biased. The radius of gyration is roughly parallel to the length for most of the length values. However, towards the end of the process, folding is independent of the radius of gyration, which is no longer a good reaction coordinate.

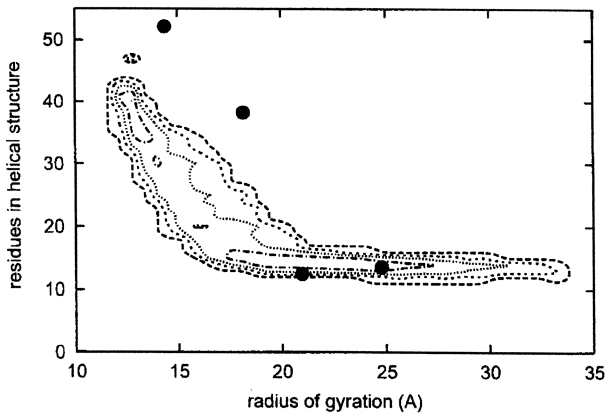

Fig. 3. A contour plot summarizing the folding progress along two reaction coordinates: The radius of gyration and the number of residues in helical conformations. The calculations of the contours (that are the logarithm of the probability as a function of these coordinates) use averages over the 26 trajectories and over all the 900 length slices of each trajectory. Also shown are the experimental measurements of Akiyama et al.[14] The first two experimental points show changes primarily in the radius of gyration (little change in the secondary structure content) that are followed by the parallel formation of secondary structure and further reduction of the radius of gyration. These qualitative features are in agreement with the simulation. Quantitatively, the secondary structure forms in the experiment at a somewhat larger radius of gyration than the theoretical calculations.

We measure the progress of the folding process by the path length. In Figure 1(a,b) we show the number of hydrogen bonds for each of the three helices as a function of the path length. It is clear from the plot that after a significant incubation period (at 40% of the trajectory length no substantial new secondary structure was formed), the N helix forms first and is followed by the C helix. The 60's helix is closing last.

Figure 1a,b presents averages. To appreciate the deviations of individual trajectories from those averages, we also consider a histogram of length differences in Figure 1c. The length difference between the formations of the 60's helix and the N and C helix are binned and plotted. A positive length difference indicates that the pair N/C forms first; a negative length interval suggests that the 60's helix forms first. The individual trajectories show a broad range of length differences with only a small fraction (three trajectories of a total of 26) with negative length differences.

It is also of interest to examine the radius of gyration as a function of path length (Fig. 2). In Figure 2, we plot the logarithm of the two-dimensional probability, $P(R_{gyr}, l)$, as a function of the radius of gyration, R_{gyr}, and trajectory length, l. It is not surprising that the distribution of the radius of gyration is broader at the beginning and is more focused at the end of the trajectories (the folded state). Perhaps more interesting is the flattening out of the monotonic (almost linear) decrease in the value of the radius of gyration at the end of the process suggesting that the last folding step is done at a relatively constant volume that cannot be accounted for using R_{gyr} as a reaction coordinate.

Another useful projection of the progress of cytochrome c folding is presented in Figure 3. The logarithm of the two-dimensional probability, $P(R_{gyr}, H)$, as a function of the radius of gyration, R_{gyr}, and the number of residues in a helical conformation, H, is shown. A "helical" conformation was defined as negative (φ, ψ) values, the backbone dihedral angles. This plot is timely; since the recent experimental analysis of Akiyama et al.[14] examined simultaneously these two coordinates and their results are presented as four black circles in Figure 3. The computational results are presented using a double average: (1) an average over the 26 folding trajectories, and (2) an average over all the length slices.

The plot is striking in its separation into two kinetic phases. Phase one includes an early collapse in which the radius of gyration decreases without significant formation of new secondary structure. In the second phase, further collapse of the chain is observed with the parallel formation of secondary structure. The experimental points of Akiyama et al.[14] are qualitatively similar to the simulations.

The two distinct processes are even clearer in Figure 4 in which we follow the sequence of events as a function of the trajectory length. In Figure 4(a), we show the logarithm of the probability for the first fifth of the trajectory, in Figure 4(b) for the second fifth and so on. In the first three length slices, the trajectory progresses only along the radius of gyration with no significant increment in the marginal content of secondary structure. Only in the last two fifths is secondary structure formed in parallel to further reduction in the radius of gyration. Sequential snapshots of the trajectory progress (compared to the average in Fig. 3) make it possible to separate in an unambiguous way equilibrium and dynamics.

Another interesting projection of the multi-dimensional folding process onto two dimensions is in Figure 5(a–e), in which we follow the formation of hydrogen bonds and

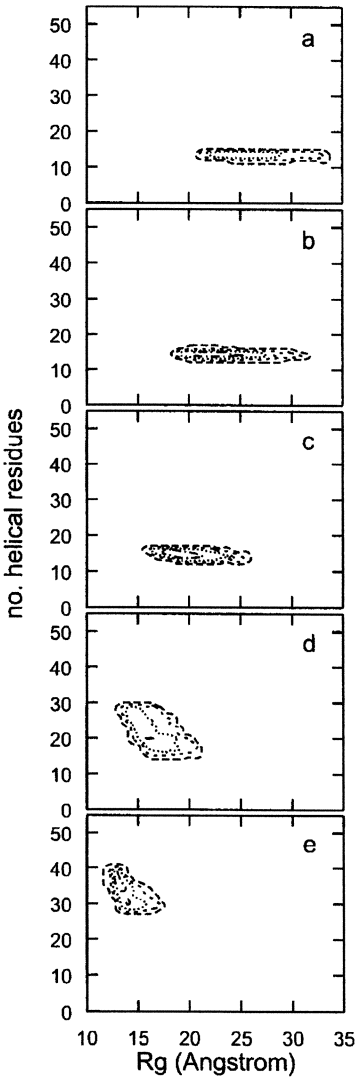

Fig. 4. A series of contour plots as a function of the trajectory length. Every slice is an average over one fifth of the trajectory. **a:** An average over the first fifth. **b:** An average over the second fifth, and so on. The contour plots are of the logarithm of the probability to observe the trajectory at a given radius of gyration R_{gyr} and secondary structure *H*. Note that in the first few panels, the progress of folding is primarily along the radius of gyration.

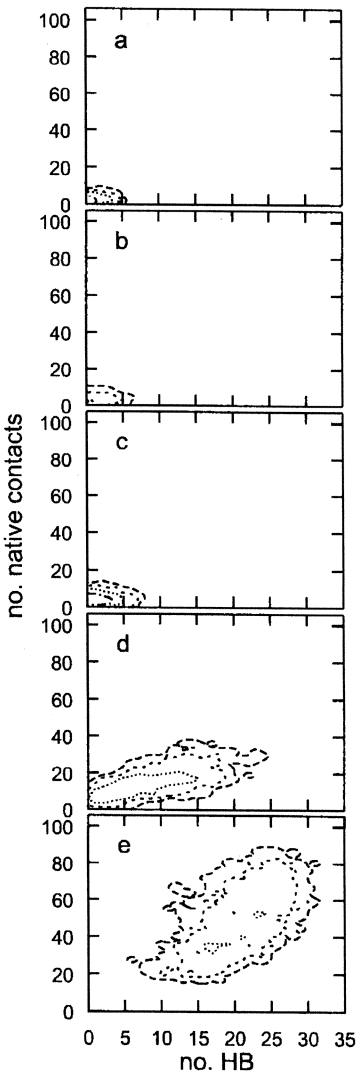

Fig. 5. Similar to Figure 4, this time for the coordinates: helix content (measured as the number of hydrogen bonds) and number of native contacts. Averages as before are done on fifths of the total lengths of the 26 trajectories. Note that both the native contacts and the secondary structure form late during the folding process with the formation of the secondary structure somewhat earlier than the formation of native contacts.

native contacts as a function of the trajectory length. The sequence of events suggests a late formation of both, with earlier formation of the secondary structure. The emerging picture, which is not inconsistent with available experimental data (see Discussion), is of a chain collapse with little formation of secondary structure to a radius of gyration below 18Å. Then substantial secondary structure starts to form in parallel to the slower accumulation of native contacts and further (slower) compactness of the protein chain.

This is further illustrated in Figure 6, which shows snapshots of structures along a folding trajectory. Figure 6a shows an unfolded configuration. The chain is far from compact and no secondary structure is observed. A structure from the molten globule with a radius of gyration of 16Å is in Figure 6b. The N and C helices are clearly formed while the 60's helix is still struggling to form in accord with the experiment.[12,52,53] Note also that the overall packing is incorrect and significant adjustment of the structure is required to reach the native fold (Fig. 6c).

It was proposed (based on experimental measurements[9,21]), that an energy barrier exists on entry to the compact state. In Figure 7, we plot the average energy as a function of the radius of gyration. We, indeed, observe a small energy barrier around a radius of gyration of 18Å, but the error bars are very significant. Note that this barrier is observed without the formation of a chemical bond between the heme iron and the histidines.

Another interesting twist to the folding of cytochrome c, is the possible formation of nonnative heme iron-histidine (26 or 33) bonds. In Figure 6c, we show the native structure of cytochrome c and mark groups that make the above nonnative bonds or contacts.

In Figure 8, we plot the probability of a contact between the heme iron and the nitrogen of the histidine rings as a function of the trajectory length. A contact is defined when the distance between the two atoms is less than 6.5Å. Also indicated on the plot are the accumulated probabilities of having a contact sometime during the trajectory. These contact probabilities can be used to obtain an upper bound for the rate of heme misligation. Note that these specific nonnative contacts are considerably more probable at the early stages of the folding process and diminish rapidly at the last fifth of the process.

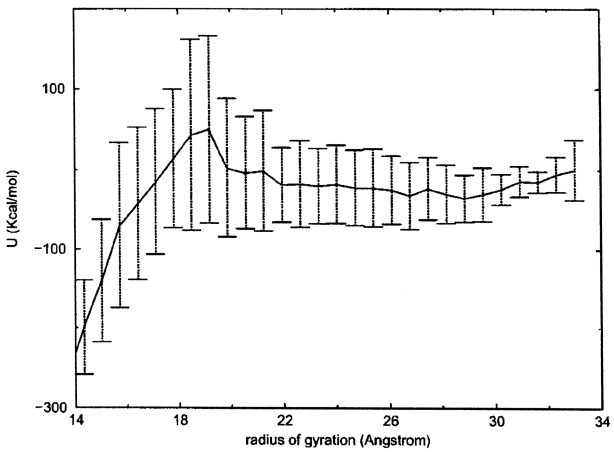

Fig. 7. The average energy (an average over the 26 trajectories) as a function of the radius of gyration. The error bars are the standard deviation. Note the barrier at a radius of gyration of (around) 18Å. This barrier can be used to denote the boundary between the molten globule and the extended unfolded configurations.

The nonnative contacts with the heme is a special phenomenon for cytochrome c with the advantage that they can be probed experimentally.[16–18] However, the presence of nonnative contacts along the folding pathways of proteins is not unique to cytochrome c. Since the overall agreement of other observables and experimental data is encouraging, we use the simulations to probe the dynamics of other nonnative contacts.

In Figure 9a, we show the logarithm of the joint probability for the radius of gyration and ***native*** contacts. In Figure 9b, the corresponding logarithm of the distribution of ***nonnative*** contacts is shown. Note that the two probabilities have significant spatial overlaps. According to

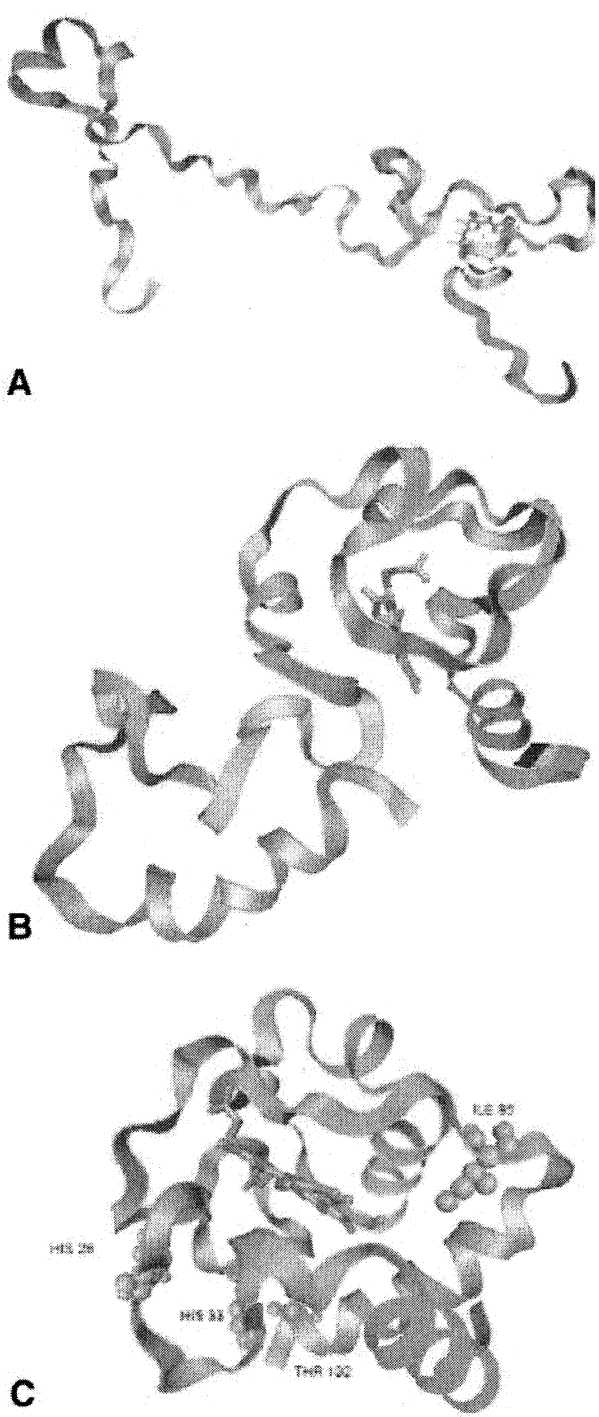

Fig. 6. Ribbon view of cytochrome c at three different positions along a folding trajectory, which is one of the 26 trajectories studied. The amino acids of the N, C, and 60's helices are in red, purple, and yellow, respectively, the heme group is in pink, and the rest of the backbone chain is in cyan. **a:** Typical extended unfolded configuration that is used to initiate a folding trajectory. **b:** Compact unfolded structure (form the molten globule state) with a radius of gyration of 16Å. The terminal helices are formed but the 60's helix is still shaping. **c:** The native crystal structure with a radius of gyration of 12.5Å. Marked on this structure are the two histidines (26 and 33, in orange) that are likely to form a bond with the heme group and create misligated folding intermediates. Also marked are residues 85 and 102 (in green) that form a nonnative contact. [A contact is on if the distance between the geometric center of the side chains is less than 6.5Å.] See text and Figures 8 and 10 for more details. Prepared using VMD.[56]

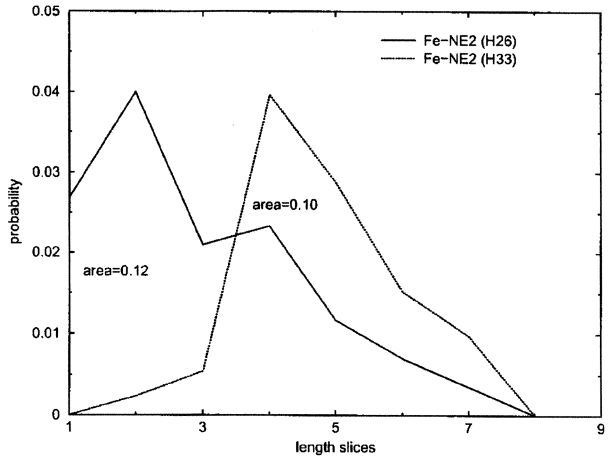

Fig. 8. The probability of a contact between the heme iron and the nitrogen atoms of histidines 26 and 33 as a function of the trajectory length. Note also the accumulated contact probabilities over all lengths (averaged over all trajectories) that are 0.12 for His 26 and 0.1 for His 33. These probabilities present an upper bound for the formation of bonded histidine-iron intermediates. They provide the reaction probability if every encounter was reactive. Information on the reaction probability per collision can lead to a better estimate. At present, the molecular dynamics simulations suggest that the formation of such intermediates is feasible.

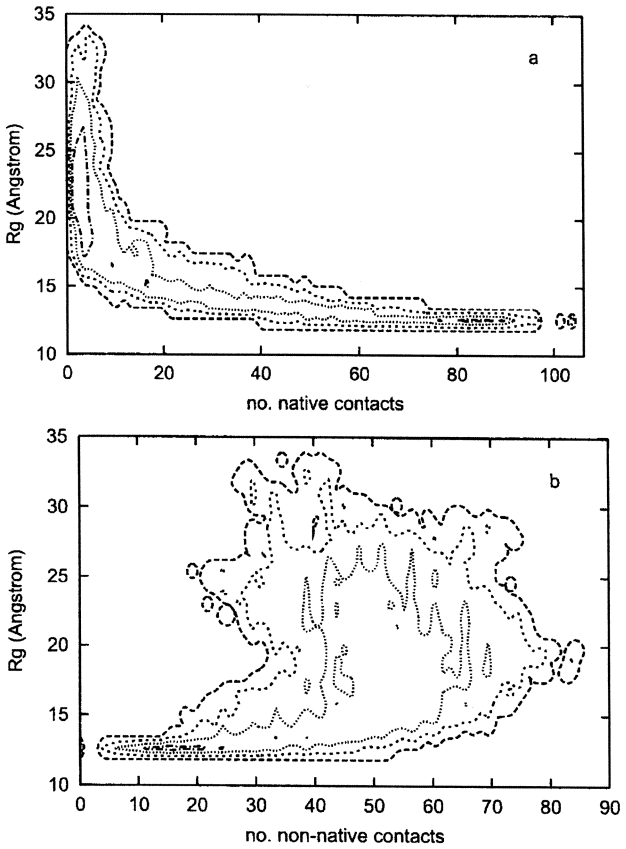

Fig. 9. A study of the native and nonnative contacts as a function of the radius of gyration. **a:** Contours of native contacts. **b:** Nonnative contacts. If we use 18Å for the size of the molten globule (as suggested by the energy barrier in Fig. 7 and the start of secondary structure formation in Fig. 3), we find a large number of nonnative contacts that coexists with native contacts at the molten globule state.

Figures 3 and 7, the onset of the second kinetic phase (that we use to define the molten globule) is around 17–18Å, to the left of the energy barrier. At a radius of gyration lower than 17Å, we still find significant concentration of nonnative contacts at the compact state [Fig. 9b]. These nonnative contacts co-exist with a smaller number of native contacts. One possible "solution" to the high concentration of nonnative contacts is recrossing of the molten globule boundary, easily fixing the wrong contacts at the extended chain configuration and reentering the "molten-globule" state. This is similar in spirit to the annealing mechanism proposed by Thirumalai and Lorimer for the action of chaperones.[54] It is of interest to check if a similar behavior is found in the folding of a protein chain in the absence of chaperon.

A sample trajectory of a non-native contact (of leucine 85 and threonine 102) is shown in Figure 10 [their positions in the native conformation are shown in Fig. 6c]. The contact was formed very early in the process and broke for the first time near the energy barrier. Then it associates in the "molten-globule" phase. The second dissociation of the nonnative contact does not include a transition over the barrier. Hence, there is sufficient structural plasticity even in the molten globule state to enable adjustments of contacts without the necessity to return to extended chain configurations.

DISCUSSIONS

In this study, we report straightforward atomically detailed simulations of the complete folding process of a protein (cytochrome c). Experimentally, the process takes longer than a millisecond and we are able to study it using a novel algorithm (SDEL) to compute classical trajectories at highly extended timescales. We divide the discussion of our results into two parts: (1) comparison to experimental data and (2) analysis of the dynamics of nonnative contacts.

Many of the results presented in the last section can be compared directly to experimental data. In Figure 1, we focus on one main experimental observable: the early formation of the N and C helices compared to the 60's helix. Numerous experiments such as hydrogen exchange[10,12,53] clearly indicated that this pair of helices forms a folding nucleus. Figure 1, in which we used hydrogen bonding to indicate helical structure in accord with the hydrogen exchange experiments, is consistent with the experimental observation. It is also interesting to note the significant "incubation period" before the start of helix formation. However, as we emphasized in The Algorithm, it is not obvious how to correlate the length with absolute time. The time is monotonic but not necessarily linear with the length. The incubation may be significantly shorter in time than what is suggested by the length parameterization.

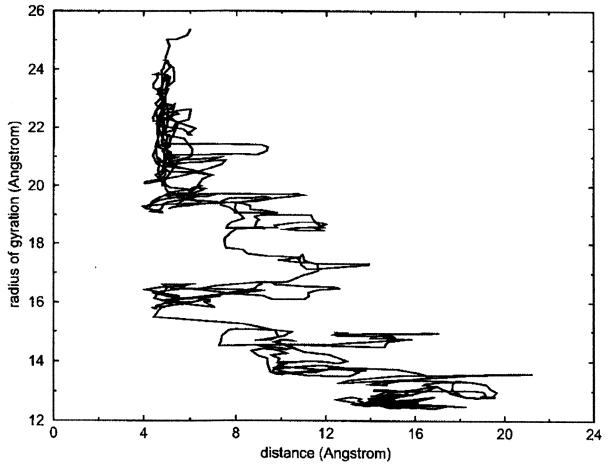

Fig. 10. An example of a length history of a nonnative contact. We consider the contact between leucine 85 and threonine 102 (shown in Fig. 6c). A contact is formed and dissociates within the boundary of the molten globule (18Å). Note the (approximate) monotonic decrease of the radius of gyration as a function of trajectory length.

The results in Figure 1(a,b) are an average over 26 trajectories. To appreciate the deviation from this average, we consider the relative ordering (in length) of the formation of the helices in individual trajectories. In Figure 1c, we show a histogram plot of the length differences between the formation of the N/C pair and the 60's helix. If a length difference is positive, then the N/C pair forms first. If it is negative, the first to form is the 60's helix. The distribution is clearly dominated by positive length differences; only 3 of the 26 trajectories have a negative length difference. A sample structure from a trajectory demonstrating the co-existence of the N/C pair before the formation of the 60's helix is shown in Figure 6b.

Another useful comparison to the experiment is provided in Figure 3. The experiment in reference 14 examined the correlated changes of the radius of gyration and the secondary structure. The four experimental points (extracted from fig. 5 of ref. 14) provide two-dimensional information on folding pathways that in principle can differentiate between alternative folding mechanisms.

Two extreme models exist for the initiation of the folding process: (1) The framework model[3] in which secondary structure elements form first and then are assembled as relatively rigid bodies into the correct tertiary structure, and (2) Hydrophobic collapse[2] in which the chain collapses first to a compact shape with little secondary structure, and following the collapse the helices "grow" in relatively compact conformations. One can imagine also intermediate cases between these extreme models.

The experimental data of Akiyama et al.[14] support a model that is more similar to the hydrophobic collapse. The first two experimental points show a very significant reduction in the radius of gyration without a parallel increase in the secondary structure. Only the last two points show growth in helical content, a growth that is done jointly with further reduction in the radius of gyration. Our study is in general agreement with the above observation. Our average plot, and detailed length slices (Fig. 4) clearly indicate an early collapse with no significant formation of new secondary structure at the early phases of the folding process. Akiyama et al.[15] observed, "This suggests that the initial collapse is not coupled with formation of secondary structure," in agreement with the calculation and Figures 3 and 4. The overall agreement is, however, not quantitative. The radius of gyration of the native state, as measured experimentally by low angle X-ray scattering (14Å)[13,14] is larger than the radius of gyration of the crystal structure (12.5Å). We used the crystal structure to model the folded state. As a result, the radii of gyration at the *folded* configuration do not agree.

Furthermore, the secondary structure content estimated from CD measurements[14] is higher than the native secondary structure computed from the structure 1HRC[7] of the protein data bank. Our results are consistent with the X-ray structure, making the slopes of the second phase (Fig. 4d,e) somewhat different.

Nevertheless, our results still agree on the existence of two separate phases and on the approximate location of the molten globule.

The question of when a compact state starts and when the extended chain configurations end (or the boundaries of the molten globule) is also of interest. In the calculations, we can use two alternative definitions. The first option is the onset of secondary structure formation. That is, the radius of gyration in which the "pure" hydrophobic collapse stops. At that point, the "rate" of the collapse is significantly smaller and it is accompanied by helix build-up. The second option is the position of the small energy barrier between extended chain configurations and the collapsed state. The kinetic-geometric criterion of the first option and the energetic measure of the second option lead to a similar radius of the collapsed state. This suggests that both measures are consistent.

The initial formation of secondary structure at the onset of the compact phase is associated with a small energy barrier.[9,21] For both of these options, the estimate of the radius of gyration of the molten globule/"compact state" (more compact than extended chain configurations) is around 18Å. This value is in accord with the estimate of Pollack et al.,[13] suggesting that the radius of gyration of the molten globule state is about 30% larger than the native.

The existence of an energy barrier on entry to the collapsed state is also consistent with the experimental measurements[9,21] (Hagen and Eaton[9] estimated it as $9kT$). We are unable to make a quantitative estimate of the barrier height (and comparison to the experimental value) due to a lack of statistics and an uncertainty regarding the reaction coordinate.

An intriguing discussion[16–18] is concerned with heme misligation. Instead of forming the iron bond to Met 80, it may form initially to one of the histidines (His 33 or His 26). This is an example of a long-lived nonnative contact that can be observed and analyzed experimentally. While our model does not include the formation of a chemical

bond between the iron and the nitrogen of the histidine, we can still perform a feasibility study for the probability that the histidine is sufficiently near the heme iron. This probability is clearly an upper bound for the probability of bond formation and is extracted from the simulation in Figure 8. Our conclusion is that relevant encounters are possible.

Is a bond to His 33 or His 26 preferred? Hagen et al.[20] estimated the ratio of misligation to His 33 or His 26 to be roughly equal. On the other hand, equilibrium studies of misligation to a heme with an iron atom substituted by a cobalt atom suggest that the ratio is 3:1.[11] Our calculations of misligation are simplistic especially since the explicit chemistry of bond formation is not included. Keeping these difficulties in mind, we note that our results are consistent with the measurements in Hagen et al.[20] This may suggest that using an equilibrium assumption to describe the folding process of cytochrome c is invalid.

It is also interesting that Shastry et al.[19] noted a weak dependence of the time scale of the initial collapse on pH. Since bond formation to the heme iron is influenced by pH, it is likely that the barrier for entry (that is ***not*** influenced by pH) is induced by conformational constraints. The energy barrier that we observe in our calculations can arise only from conformational bottlenecks since we do not model computationally chemical bond formation (between the heme iron and the histidine nitrogen). This observation supports the suggestion made by Shastry et al.[19]

The above analysis indicated that we are able to reproduce, at least qualitatively, a number of experimental observations. We are, therefore, extending our analysis below to features not easily accessible to experiment that are nevertheless of significant interest. How important are nonnative contacts during the folding process? Do they slow down the process and form kinetic traps (off-pathway intermediates)? Are they found at all? Or perhaps the requirements from a fast folder also imply that the number of nonnative contacts are small in number and can be ignored. These are sample questions that we address below in our discussion on nonnative contacts.

The Go model[55] that artificially biased a contact or a local conformation to the native state diminishes the importance of nonnative contacts in the overall dynamics. Is this approximation consistent with the atomically detailed model of cytochrome c folding? Another mechanistic question of interest is the elimination of nonnative contacts. Is the molten globule state sufficiently flexible to allow for structural adjustments and reformation of correct native contacts while maintaining a compact state? Perhaps it is not necessary to leave the "compact phase" (which is still significantly larger than the native) in order to remove most types of nonnative contacts.

The simulations of cytochrome c that we performed indicate first that nonnative contacts play a major role in the folding process. A typical maximal number of nonnative contacts in a folding trajectory is sixty, more than half the number of contacts in the native state (104). The simulation also indicates a large number of nonnative contacts in the compact state and a plastic molten globule

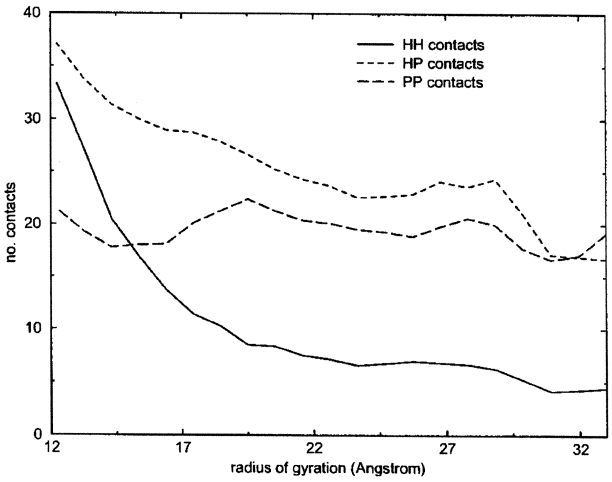

Fig. 11. A plot that shows the number of hydrophobic-hydrophobic (HH), hydrophobic-polar (HP), and polar-polar (PP) contacts as a function of the radius of gyration. The plot is obtained as an average over the 26 folding trajectories studied here.

that can correct many nonnative contacts without a transition to extended configurations (Figs. 9 and 10).

Consider that in Figure 9b, with an estimate of the molten globule radius of gyration of 17–18Å, a large number of nonnative contacts at the compact state are found. Hence, it is not necessary to form the majority of native contacts before overcoming the energy barrier for the entry to the collapsed phase. In fact, there are considerably more nonnative contacts than native contacts immediately following the entry.

In Figure 10, we show one typical example for a nonnative contact between leucine 85 and threonine 102. Many other trajectories and nonnative contacts that we have examined behave similarly. The length history of the contact does not show a significant increase in the radius of gyration at any length slice as the folding moves forward. Rather, the radius of gyration decreases almost monotonically as the reaction progresses.

To summarize, the mechanism for cytochrome c folding that we extracted from the simulations agrees with the overall experimental suggestions of (1) a rapid collapse to a compact phase without significant formation of secondary structure and (2) further "compression" of the molten globule state with parallel formation of helices; the N and C helices form first. We have also looked at a folding event that so far was not probed in detail experimentally and suggested the existence (and correction) of a large number of nonnative contacts in the molten globule state. These numerous nonnative contacts are easily "fixable" within the framework of the molten globule state without the need for transitions between the molten globule and extended conformations. Since the formation of molten globule is driven by a sharp increase in hydrophobic contacts (Fig. 11), and since the majority of contacts at the formation of the molten globule are nonnative, we conclude that nonnative contacts drive the formation of the

molten globule and, as such, play a positive role in the folding process.

ACKNOWLEDGMENTS

This research was supported by an NIH grant to R.E. and by an equipment NSF grant to the Cornell Theory Center (Keshav Pingali, PI). We thank Dave Thirumalai for many useful discussions and for suggesting the problem of nonnative contacts.

REFERENCES

1. Creighton TE, editor. Protein folding. New York: Freeman; 1992.
2. Dill KA. Theory for the folding stability of globular proteins. Biochemistry 1985;24:1501–1509.
3. Kim PS, Baldwin RL. Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding. Annu Rev Biochem 1982;51:459–489.
4. Boczko EM, Brooks CL III. First-principles calculation of the folding free-energy of a 3-helix bundle protein. Science 1995;269: 393–396.
5. Duan Y, Kollman PA. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. Science 1998;282:740–744.
6. Mayor U, Johnson CM, Daggett V, Fersht AR. Protein folding and unfolding in microseconds to nanoseconds by experiment and simulation. Proc Natl Acad Sci USA 2000;97:13518–13522.
7. Bushnell GW, Louie GV, Brayer, GD. High-resolution three-dimensional structure of horse heart cytochrome c. J Mol Biol 1990;214:585–595.
8. Englander SW, Sosnick TR, Mayne LC, Shtilerman M, Qi PX, Bai Y. Fast and slow folding in cytochrome c. Acc Chem Res 1998;31: 737–744.
9. Hagen SJ, Eaton WA. Two-state expansion and collapse of a polypeptide. J Mol Biol 2000;301:1019–1027.
10. Elöve GA, Chaffotte AF, Roder H, Goldberg ME. Early steps in cytochrome c folding probed by time-resolved circular dichroism and fluorescence spectroscopy. Biochemistry 1992;31:6876–6883.
11. Tezcan FA, Findley WM, Crane BR, Ross SA, Lyubovitsky JG, Gray HB, Winkler JR. Using deeply trapped intermediates to map the cytochrome c folding landscape. Proc Natl Acad Sci USA 2002;99:8626–8630.
12. Roder H, Elöve GA, Englander SW. Structural characterization of folding intermediates in cytochrome c by H-exchange labeling and proton NMR. Nature 1988;335:700–704.
13. Pollack L, Tate MW, Darnton NC, Knight JB, Gruner SM, Eaton WA, Austin RH. Compactness of the denatured state of a fast-folding protein measured by submillisecond small-angle x-ray scattering. Proc Natl Acad Sci USA 1999;96:10115–10117.
14. Akiyama S, Takahashi S, Kimura T, Ishimori K, Morishima I, Nishikawa Y, Fujisawa T. Conformational landscape of cytochrome c folding studied by microsecond-resolved small-angle x-ray scattering. Proc Natl Acad Sci USA 2002;99:1329–1334.
15. Akiyama S, Takahashi S, Ishimori K, Morishima I. Stepwise formation of α-helices during cytochrome c folding. Nat Struct Biol 2000;7:514–520.
16. Elöve GA, Bhuyan AK, Roder H. Kinetic mechanism of cytochrome c folding: involvement of the heme and its ligands. Biochemistry 1994;33:6925–6935.
17. Sosnick TR, Mayne L, Hiller R, Englander SW. The barriers in protein folding. Nat Struct Biol 1994;1:149–156.
18. Takahashi S, Yeh S-R, Das TK, Chan C-K, Gottfried D, Rousseau DL. Folding of cytochrome c initiated by submillisecond mixing. Nat Struct Biol 1997;4:44–50.
19. Shastry MCR, Sauder JM, Roder H. Kinetic and structural analysis of submillisecond folding events in cytochrome c. Acc Chem Res 1998;31:717–725.
20. Hagen SJ, Latypov RM, Dolgikh DA, Roder H. Rapid intrachain binding of histidine-26 and histidine-33 to heme in unfolded ferrocytochrome c. Biochemistry 2002;41:1372–1380.
21. Shastry MCR, Roder H. Evidence for barrier-limited protein folding kinetics on the microsecond time scale. Nat Struct Biol 1998;5:385–392.
22. Shaknovich EI. Theoretical studies of protein folding thermodynamics and kinetics. Curr Opin Struct Biol 1997;7:29–40.
23. Camacho CJ, Thirumalai D. Kinetics and thermodynamics of folding in model proteins. Proc Natl Acad Sci USA 1993;90:6369–6372.
24. Onuchic JN, Luthey-Shulten Z, Wolynes PG. Theory of protein folding: the energy landscape perspective. Annu Rev Phys Chem 1997;48:545–600.
25. Daggett V. Molecular dynamics simulations of the protein unfolding/folding reaction. Acc Chem Res 2002;35:422–429.
26. Brooks III CL. Protein and peptide folding explored with molecular simulations. Acc Chem Res 2002;35:447–454.
27. Elber R, Meller J, Olender R. Stochastic path approach to compute atomically detailed trajectories: application to the folding of *C* peptide. J Phys Chem 1999;103:899–911.
28. Zalov V, Elber R. Parallel computations of molecular dynamics trajectories using stochastic path approach. Comput Phys Commun 2000;128:118–127.
29. Uitdehaag JCM, van der Veen BA, Dijkhuizen L, Elber R, Dijkstra BW. The enzymatic circularization of a malto-octaose linear chain studied by stochastic reaction path calculations on cyclodextrin glycosyltransferase. Proteins 2001;43:327–335.
30. Siva K, Elber R. Ion permeation through the gramicidin channel: atomically detailed modeling by the stochastic difference equation. Proteins 2003;50:63–80.
31. Ghosh A, Elber R, Scheraga HA. An atomically detailed study of the folding pathways of protein A with the stochastic difference equation. Proc Natl Acad Sci USA 2002;99:10394–10398.
32. Hawkins GD, Cramer CJ, Truhlar DG. Pairwise solute descreening of solute charges from a dielectric medium. Chem Phys Lett 1995;246:122–129.
33. Tsui V, Case DA. Theory and applications of the generalized Born solvation model in macromolecular simulations. Biopolymers 2000; 56:275–291.
34. Olender R, Elber R. Calculation of classical trajectories with a very large time step: Formalism and numerical examples. J Chem Phys 1996;105:9299–9315.
35. Elber R, Ghosh A, Cárdenas A, Stern H. Bridging the gap between reaction pathways, long time dynamics and calculation of rates. Adv Chem Phys (in press).
36. Carr R, Parrinello M. Unified approach for density functional theory and molecular dynamics. Phys Rev Lett 1985;55:2471–2474.
37. Elber R, Ghosh A, Cárdenas A. Long time dynamics of complex systems. Acc Chem Res 2002;35:396–403.
38. Gillilan RE, Wilson KR. Shadowing, rare events, and rubber bands: a variational Verlet algorithm for molecular dynamics. J Chem Phys 1992;97:1757–1772.
39. Cho AE, Doll JD, Freeman DL. The construction of double-ended classical trajectories. Chem Phys Lett 1994;229:218–224.
40. Miller WH. Classical-limit quantum mechanics and the theory of molecular collisions. Adv Chem Phys 1974;25:69–177.
41. Grubmüller H, Heller H, Windemuth A, Shulten K. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. Mol Sim 1991;6:121–142.
42. Humphreys DD, Friesner RA, Berne BJ. A multiple time step molecular dynamics algorithm for macromolecules. J Chem Phys 1994;98:6885–6892.
43. Batcho PF, Case DA, Schlick T. Optimized particle-mesh Ewald/multiple-time step integration for molecular dynamics simulations. J Chem Phys 2001;115:4003–4018.
44. Schlick T, Skeel RD, Brunger AT, Kale LV, Board JA, Hermans J, Schulten K. Algorithmic challenges in computational molecular biophysics. J Comp Phys 1999;151:9–48.
45. Elber R. Reaction path studies of biomolecules. In: Elber R, editor. Recent developments in theoretical studies of proteins. Singapore: World Scientific; 1996. p 65–136.
46. Passerone D, Parrinello M. Action-derived molecular dynamics in the study of rare events. Phys Rev Lett 2001;87:108–302.
47. Mandelbort BM. The fractal geometry of nature. San Francisco: Freeman; 1977.
48. Elber R, Roitberg A, Simmerling C, Goldstein R, Li H, Verkhivker G, Keasar C, Zhang J, Ulitsky A. Moil: a program for simulations of macromolecules. Comp Phys Commun 1995;91:159–189.
49. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta Jr. S, Weiner, P. A new force field for molecular mechanical simulation of nucleic acids and proteins. J Am Chem Soc 1984;106:765–784.

50. Jorgensen WL, Tirado-Rives J. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. J Am Chem Soc 1988;110:1657–1666.
51. Czerminski R, Elber R. Self avoiding walk between two fixed points as a tool to calculate reaction paths in large molecular systems. Int J Quantum Chem 1990;24:167–186.
52. Bai Y. Kinetic evidence for an on-pathway intermediate in the folding of cytochrome c. Proc Nat Acad Sci USA 1999;96:477–480.
53. Hoang L, Bédard S, Krishna MMG, Lin Y, Englander SW. Cytochrome c folding pathway: kinetic native-state hydrogen exchange. Proc Nat Acad Sci USA 2002;99:12173–12178.
54. Thirumalai D, Lorimer GH. Chaperonin-mediated protein folding. Annu Rev Biophys Biomol Struct 2001;30:245–269.
55. Taketomi H, Ueda Y, Go N. Studies on protein folding, unfolding and fluctuations by computer-simulation. 1. Effect of specific amino-acid sequence represented by specific inter-unit interactions. Int J Pept Prot Res 1975;7:445–459.
56. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. J Mol Graph 1996;14:33–38.
57. Snow CD, Nguyen N, Pande VS, Gruebele M. Absolute comparison of simulated and experimental protein folding dynamics. Nature 2002;420:102–106.