

The Trp Cage: Folding Kinetics and Unfolded State Topology via Molecular Dynamics Simulations

Christopher D. Snow, Bojan Zagrovic, and Vijay S. Pande*

Department of Chemistry and Biophysics Program, Stanford University, Stanford, California 94305-5080

Received September 18, 2002

A number of rapidly folding proteins have been characterized in recent years.¹ These small proteins can provide the first direct comparisons between simulated and experimental protein folding kinetics and pathways.² Proteins have been characterized through thermodynamic sampling methods,³ unfolding simulations,⁴ and folding simulations using simple potentials.⁵ Here, as described recently,² we use several thousand stochastic dynamics simulations in a generalized-Born implicit solvent (in atomic detail) to simulate the folding dynamics of the Trp cage mini-protein under experimental conditions (27 °C with full solvent viscosity, $\gamma = 91 \text{ ps}^{-1}$). The Folding@Home distributed computing project was used to generate an aggregate simulation time of $\sim 100 \mu\text{s}$ (~ 250 CPU years). First we capture the rapid relaxation from an extended starting condition to a relaxed unfolded state ensemble of thousands of conformations. With continued simulation, a small fraction of these simulations reach the folded state. Furthermore, the topology of the collapsed unfolded state closely resembles the native state.

The 20-residue Trp cage (NLYIQWLKDG GPSSGRPPPS) contains a short α -helix from residues 2 through 8, a 3_{10} -helix from residues 11 through 14, and a C-terminal poly-proline II helix to pack against the central tryptophan.⁶ Hagen and co-workers present direct experimental evidence for cooperative two-state folding with a room-temperature folding time of $4 \mu\text{s}$.⁷ With the assumption of two-state folding kinetics, short simulations from the unfolded state under native conditions have a certain small probability of folding. The length of the simulations must, of course, exceed the barrier-crossing time to fold. Furthermore, each simulation must include the time necessary to relax from a denatured state (here modeled as a completely extended form) to the unfolded state under native conditions. The Trp cage simulations collapsed in only nanoseconds. In fact, the entire extended state ensemble relaxed to a compact unfolded state ensemble in less than 10 ns.

We continued the folding simulations well past the initial relaxation (over 1000 simulations extend to 30 ns) and observed folding events consistent with a two-state folding pathway. For a two-state protein with a folding constant $\tau = 4 \mu\text{s}$, we would expect ~ 7.5 out of 1000 molecules to fold within the first 30 ns assuming rapid relaxation to the unfolded state and rapid two-state barrier crossing. Our simulations were generated through distributed computing techniques⁸ and the full set of data includes simulations of many different lengths (i.e. ~ 3000 simulations run to 1 ns, ~ 2000 simulations extend to 5 ns, ~ 1000 to 30 ns, ~ 650 to 40 ns, ~ 160 to 80 ns).

Given that the relaxed unfolded ensemble has fairly low RMSD_{C α} (the distribution is centered about ~ 4.8 after 30 ns) and that the relaxed folded ensemble is centered about ~ 2.4 after 10 ns, the choice of a good RMSD_{C α} cutoff to define "folded" conformations from the unfolded ensemble is not trivial. In Figure 1 we present

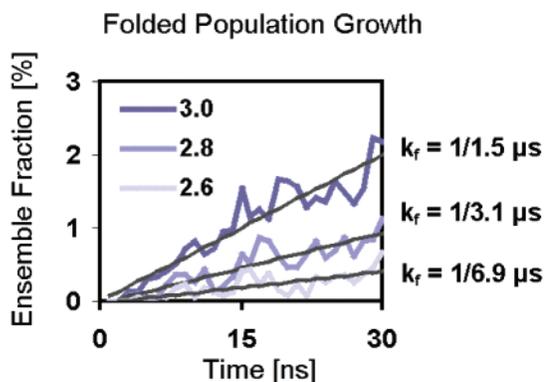


Figure 1. Estimated Trp cage folding kinetics. (A) The time-resolved population of folded molecules in the ensemble of over a thousand folding simulations for a variety of RMSD_{C α} cutoff values. We fit each of these population growth curves to the initial stage of a single-exponential folding reaction to determine rate constants assuming the folding reaction will reach completion at long times. With 3.0 Å $\tau = 1.5 \mu\text{s}$, 2.9 Å $\tau = 2.0 \mu\text{s}$, 2.8 Å $\tau = 3.1 \mu\text{s}$, 2.7 Å $\tau = 5.5 \mu\text{s}$, 2.6 Å $\tau = 6.9 \mu\text{s}$, 2.5 Å $\tau = 8.7 \mu\text{s}$.

the estimated folding rate given a range of reasonable RMSD_{C α} cutoffs at and below 3.0 Å. The estimated rate varies more from the arbitrary RMSD_{C α} cutoff than any fitting errors. Thus we present a bounded folding time estimate ranging from $\tau = 1.5 \mu\text{s}$ using a 3.0 Å cutoff to $8.7 \mu\text{s}$ for a 2.5 Å cutoff. We emphasize that the estimated folding rate assumes two-state folding to completion.

By capturing the folding ensemble very early into folding (tens of ns), our methodology gives us a detailed view of the unfolded state under native conditions. Recently, we have postulated that in certain cases the ensemble averaged mean structure of the unfolded state may have natively-like topology.⁹ The sampling of the unfolded state of Trp cage in our simulations has allowed us to test this mean structure hypothesis. At each time point along the folding trajectory we calculated the C β -C β distance matrix. By averaging these matrices, we obtained mean matrices representative of the ensemble at each time point. Since these matrices do not correspond to any individual structure, we further asked which real, physical conformer in the simulated ensemble at each time point was closest to these mean matrices, such that their distance root-mean square deviation (dRMS) from the mean C β -C β matrix is smallest (dRMS is a standard metric for comparing two structures, i.e., distance matrices, defined as $\text{dRMS} = [2\sum_{i>j} [D_{ij}(1) - D_{ij}(2)]^2 / n(n-1)]^{1/2}$, where $D_{ij}(x)$ refers to the distance between atoms i and j in structure x and n is the total number of atoms included within each structure).

We find that the mean structure is similar to the NMR structure of the native state, with dRMS and RMSD_{C α} values within the variation of the native state. For all time points after 10 ns (after collapse), the mean structure is, on average, closer (RMSD_{C α}) to the NMR structure than 87% of the other ensemble members. In the most favorable example (at 33 ns) RMSD_{C α} = 2.1 Å. Figure 2

* Corresponding author. E-mail: pande@stanford.edu.

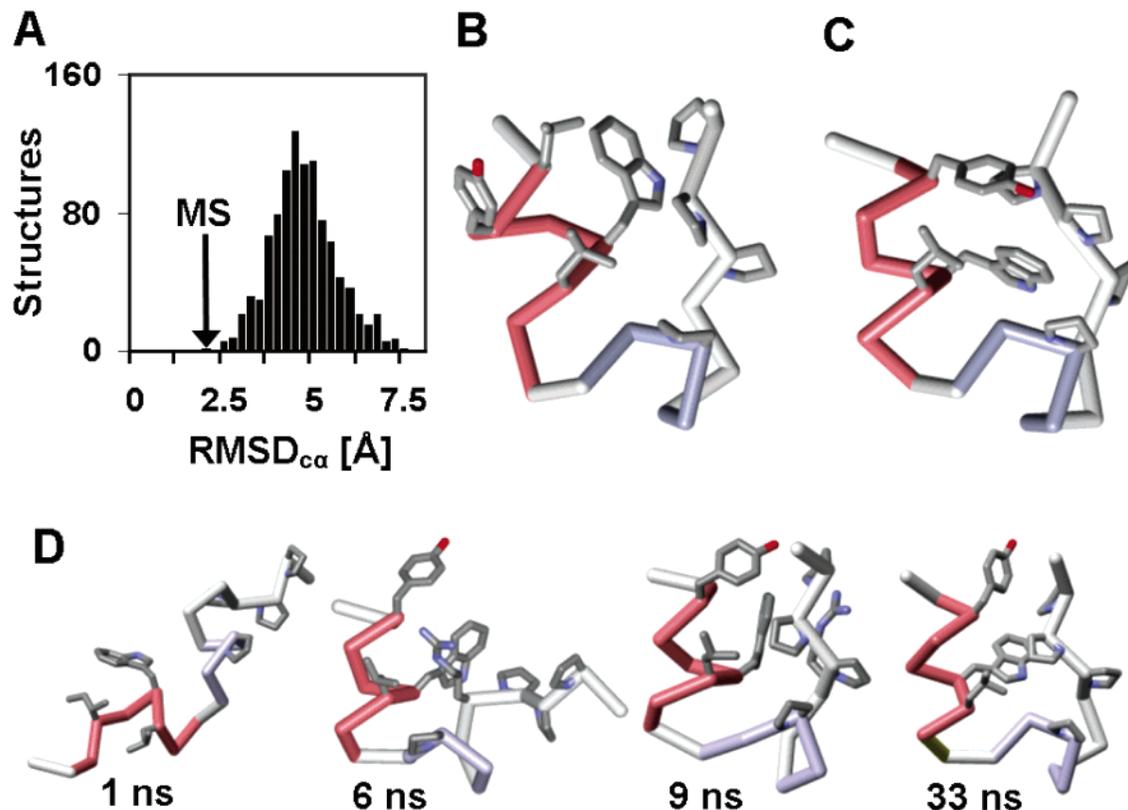


Figure 2. Trp cage unfolded state topology and folding results. For the unfolded ensemble we show (A) the $\text{RMSD}_{\text{C}\alpha}$ distribution at 33 ns, (B) the mean structure at 33 ns, and for comparison (C) the Trp cage NMR structure (1L2Y). We pick the member of the folding simulation ensemble that has a $\text{C}_\beta\text{--C}_\beta$ distance matrix closest to the ensemble averaged $\text{C}_\beta\text{--C}_\beta$ distance matrix (the mean structure, MS). The structure is more similar ($\text{RMSD}_{\text{C}\alpha} = 2.1 \text{ \AA}$) to the NMR structure than 99.6% of the unfolded ensemble at 33 ns, $\langle \text{RMSD}_{\text{C}\alpha} \rangle = 4.8 \pm 0.9 \text{ \AA}$. We also show (D) an example folding trajectory, with snapshots at 1, 6, 9, and 33 ns. This trajectory contains the lowest $\text{RMSD}_{\text{C}\alpha}$ conformation reached in a folding simulation (1.4 \AA from the NMR structure at 33 ns, eight trajectories pass below 2 \AA). The helix (2–9) is shown in red and the 3_{10} turn (11–14) in blue. Note the rapid formation of fluctuating nativelylike secondary structure, the initial packing of Arg16 between Trp6 and Leu7 at 6 ns, and the repacking of Arg16 behind Trp6 at 9 ns. At 29 ns the Arg16 formed its native interaction with Asp9 allowing the trajectory to reach very low $\text{RMSD}_{\text{C}\alpha}$.

shows the $\text{RMSD}_{\text{C}\alpha}$ distribution at 33 ns (from 996 conformations) where the mean structure is more similar to the NMR structure than 99.6% of the unfolded ensemble at that time point. We note that this example reproduces the native state topology but contains a π -helix, rather than an α -helix, and an altered core packing. In general, if we consider the mean-structure hypothesis as a prediction method, we can select nativelylike states from among the vast majority of unfolded Trp cage structures; this fact holds for all time points after 10 ns, when the unfolded ensemble has collapsed to a compact globule.

The Trp cage has lived up to its promise as a meeting point for simulation and experiment. Not only can molecular dynamics generate folding trajectories, but it appears to have quantitative predictive power for both folding rates and structure. We look forward to extending the grounds for comparison to folding mechanism, unfolded state structural detail, and temperature or viscosity rate dependence.

Acknowledgment. We thank the worldwide Folding@Home volunteers who made this work possible (<http://folding.stanford.edu>). C.S. and B.Z. are supported by predoctoral fellowships from HHMI. The computation was supported by the NIH (R01GM62868-01A2), ACS-PRF (36028-AC4), and NSF MRSEC CPIMA (DMR-9808677) and a gift from Intel.

References

- (1) (a) Mayor, U.; Johnson, C. M.; Daggett, V.; Fersht, A. R. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 13518–13522. (b) Ferguson, N.; Johnson, C. M.; Macias, M.; Oschkinat, H.; Fersht, A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 13002–13007. (c) Burton, R. E.; Huang, G. S.; Daugherty, M. A.; Calderone, T. L.; Oas, T. G. *Nat. Struct. Biol.* **1997**, *4*, 305–310. (d) Wittung-Stafshede, P.; Lee, J. C.; Winkler, J. R.; Gray, H. B. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 6587–6590. (e) Myers, J. K.; Oas, T. G. *Nat. Struct. Biol.* **2001**, *8*, 552–558. (f) Munoz, V.; Thompson, P. A.; Hofrichter, J.; Eaton, W. A. *Nature* **1997**, *390*, 196.
- (2) (a) Duan, Y.; Kollman, P. A. *Science* **1998**, *282*, 740–744. (b) Zagrovic, B.; Sorin, E. J.; Pande, V. *J. Mol. Biol.* **2001**, *313*, 151–169. (c) Simmerling, C. L.; Strockbine, B.; Roitberg, A. E.; *J. Am. Chem. Soc.* **2002**, *124*, 11258–11259 (d) Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, advance online publication, October 20 (doi: 10.1038/nature01160).
- (3) Shea, J. E.; Brooks, C. L. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499–535.
- (4) Daggett, V.; *Acc. Chem. Res.* **2002**, *35*, 422–429.
- (5) Ferrara, P.; Caflisch, A.; *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 10780–10785.
- (6) (a) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425–430. (b) Gellman, S. H.; Woolfson, D. N. *Nat. Struct. Biol.* **2002**, *9*, 408–410.
- (7) Qiu, L.; Pabit, S. A.; Roitberg, A. E.; Hagen, S. J. *J. Am. Chem. Soc.* **2002**, *124*, 12952–12953.
- (8) (a) Shirts, M. R.; Pande, V. S.; *Science* **2000**, *290*, 1903–1904. (b) Shirts, M. R.; Pande, V. S. *Phys. Rev. Lett.* **2001**, *86*, 4983–4987.
- (9) Zagrovic, B.; Snow, C. D.; Khaliq, S.; Shirts, M. R.; Pande, V. S. *J. Mol. Biol.* **2002**, *323*, 153–164.

JA028604L