Supporting Information

a Beccara et al. 10.1073/pnas.1111796109

SI Text

Computational Procedures and Simulation Details. The atomistic Dominant Reaction Pathways (DRP) calculations were performed using DOLOMIT. The code calls a librarized version of GROMACS 4.5.2 (1) to calculate the molecular potential energy and its gradient.

We defined the native state as the set of conformations with a rmsd to the crystal structure of the C_α in the hairpins smaller than 3.5 Å. A configuration was considered denatured if the rmsd to native of both hairpins was larger than 6 Å. The stability of the native state within the present force field was checked by running 12 unbiased 2 ns-long MD simulations at the room temperature (300 K). In all such trajectories the protein remained in the native state.

We then generated 44 independent initial conditions, by running a 50 ps MD at 1,600 K, starting from the energy minimized native state, followed by a 100 ps relaxation at 300 K. The time step employed in all the simulations was 1 fs. From 24 starting configurations we computed 96 trial trajectories each consisting of 50,000 rMD (ratchet-and-pawl MD) steps. For each of the remaining 20 initial conditions, the number of trial paths was limited to 48. In such rMD simulations, the ratchet spring constant was set to $k_R = 0.02$ kcal/mol. Using this value, the modulus of the biasing force was always found to be at least one order of magnitude smaller than the norm of internal forces.

We observed that 5 of the 44 initial conditions did not correspond to denatured states, hence they were rejected. In addition, in 13 of the remaining 39 sets, more than 80% of the trial trajectories did not reach the native state within the simulation time. In these cases the exploration of the path space was limited to very few trial trajectories, so the corresponding dominant paths were discarded. For the remaining 26 sets of paths, we identified the most probable by computing the Onsager-Machlup (OM) action. The information about the set of parameters used in the simulation and the number of trajectories considered is summarized in Table S1.

The simulations of the equilibrium properties of the WW-domain in the coarse-grained models were performed using a Monte Carlo algorithm based on a combination of Cartesian, crankshaft (2) and pivot (3) moves. The free energy as a function of an arbitrary set of reaction coordinates (potential of mean force) was obtained from the frequency histogram calculated from long MC trajectories.

Dependence of the Dominant Pathways on the Strength of the Biasing Constant k_R. The rMD simulations depend on an external parameter k_R, which sets the strength of the biasing force. At very low values of k_R, the rMD trajectories are minimally biased, and the system performs the first folding transition on typical time intervals comparable with the folding time $t \sim 1/k_f$—where k_f is the folding rate. On the other hand, the DRP method is based on the transition probability given in Eq. 2, where the time interval t is of the order of the transition path time $t_{\mathrm{TPT}} \ll 1/k_f$. Hence, DRP simulations based on ratchet simulations in which k_R is chosen too small are computationally very expensive, because most paths do not reach the native state within the time interval considered, hence do not contribute to the transition probability.

In the opposite high k_R limit, the bias force becomes comparable with the physical internal forces acting on the atoms and the dynamics are affected by a significant bias. In this regime, if the ratcheting coordinate is not optimal, the system is driven into relatively large free energy regions. The unbiased statistical weight given by the exponent of the OM action is expected to penalize these trajectories. The evolution of the biasing coordinate, in two typical folding trial rMD trajectories is shown in Fig. S1.

It is important to study to what extent DRP results obtained this way depend on the value of the bias constant k_R adopted in the rMD simulations. In Fig. S2 we plot the dominant reaction pathway obtained starting from the same initial condition, using different values of k_R which span over almost two orders of magnitude. We see that in most simulations the folding occurs through the same qualitative pathway, in which the first hairpin forms before the second begins to fold. Only in one case —for a low value of the coupling constant k_R— we find that the protein travels across the denatured state before taking a different pathway to the native state, in which the order of formation of the hairpins is reversed. Such a trajectory spends a much longer fraction of rMD steps in exploring the denatured state and initiates the transition from a very different configuration.

It is interesting to note that the most probable path turns out not to be the one with the lowest ratchet constant. In particular, the trajectory taking the second pathway (labeled with K_1 in Fig. S2) is among those with the lowest statistical weight, because the OM action tends to penalize paths which travel long Euclidean distances in configuration space. This result illustrates that choosing a low ratchet constant k_R produces an enhanced exploration of the denatured state, but does not necessarily lead to a more efficient identification of the most probable trajectories connecting given initial denatured configuration with the native state, in a fixed short time.

1. Hess B, Kutzner C, van derSpoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput* 4:435–447.
2. Kremer K, Binder K (1988) Monte Carlo simulation of lattice models for macromolecules. *Computer Physics Reports* 7:259–310.
3. Madras N, Sokal AD (1988) The pivot algorithm: a highly efficient Monte Carlo method for self-avoiding walk. *Journal of Statistical Physics* 12:109–186.

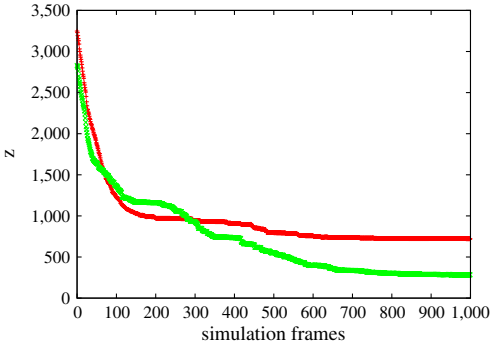

Fig. S1. The evolution of the biasing coordinate, in two typical folding trial rMD trajectories.

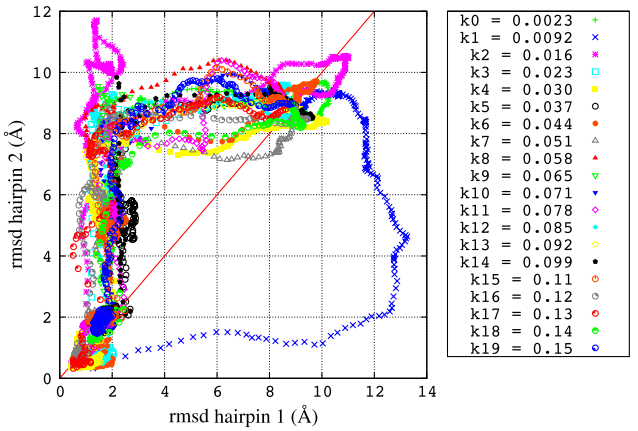

Fig. S2. Dominant pathways obtained with different values of the ratchet force, starting from the same initial condition. The values of k_R are given in units of kcal/mol.

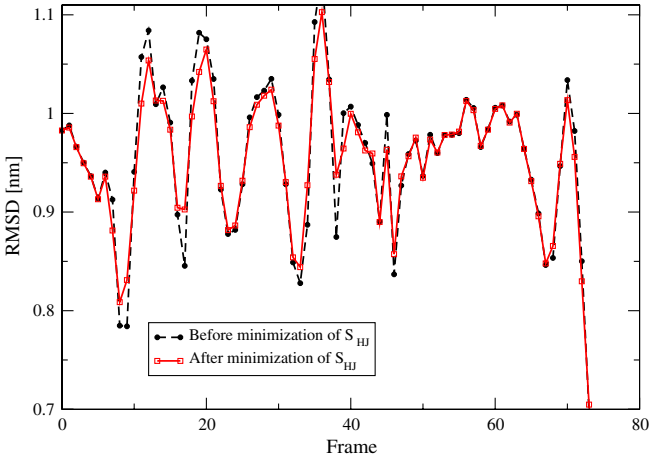

Fig. S3. The effect of the minimization of the HJ functional on a trial path obtained from the rMD algorithm on the rmsd to native of the C_α atoms in the backbone. This second refinement step does not change the qualitative structure of the dynamics.

Table S1. Summary of the simulation details

dt	k_R	Unfold. MD steps ($T = 1{,}600$ K)	Therm. MD steps ($T = 300$ K)	Fold. rMD steps ($T = 300$ K)	Denatured initial conditions	Rejected
1 fs	0.02 kcal/mol	50×10^3	100×10^3	50×10^3	44	18