Supporting Information
Folding intermediate in the villin headpiece domain arises from disruption of a N-terminal hydrogen bonded network

Jana Khandogin[1], Daniel Raleigh and Charles L. Brooks, III[1,*]

1. Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037
2. Department of Chemistry, State University of New York at Stony Brook, Stony Brook, NY 11795.
* *brooks@scripps.edu*

1 Simulation details

To obtain the pK_a value of H41 for state **1** and **2**, we carried out the 1-ns REX-CPHMD simulations using the molecular dynamics program CHARMM (version c33a1) (1) and MMTSB tool set (http://mmtsb.scripps.edu (2)) starting from an X-ray structure (PDB ID: 1YU5) at 1, 2, 3, 4, 5 and 7, and a minimized average solution NMR structure (PDB ID: 1QQV) at pH 3, 5 and 7. We used 16 replicas in an exponentially spaced temperature range between 298 and 450 K. Adjacent replicas were allowed to attempt exchange every 2 ps, at which time the atomic and titration coordinates were collected. The pK_a value was then determined by fitting the unprotonated fractions computed from the titration coordinates at different pH to the generalized form of the Henderson-Hasselbach equation

$$S^{\text{unprot}} = \frac{1}{1+10^{n(\text{p}K_a-\text{pH})}} \tag{1}$$

where n is the Hill coefficient. Other details of the REX-CPHMD titration simulations as well as the CPHMD method have been previously described (3; 4; 5).

To obtain the structural properties for state **1** and **2**, we extended the above simulations at pH 7 to 4 ns. The simulations show good convergence in terms of the total energy and main-chain heavy atom RMS deviation with respect to the starting structure of intact HP67 (Figure S1). The last 2-ns of data from the 298 K window were used for analysis, including calculations of radius of gyration, solvent accessible surface area and inter-residue contact maps as well as conformational clustering. As for the latter, an RMS binning method with a cluster RMS cutoff of 1 Å, as implemented in the GROMACS molecular dynamics software package (6), was employed.

2 Calculation of NMR relaxation parameters

To obtain NMR relaxation parameters, we carried out 5-ns CPHMD simulations at 298 K and pH 7 starting from the X-ray crystal and NMR solution structures. In the latter case, the total energy decreases rapidly in the first 1 ns. The convergence of the trajectories is demonstrated by the time evolution of the total energy and main-chain heavy atom RMS deviations with respect to the starting structures (Figure S2). Coordinates were collected every 1 ps from the last 2 ns of the trajectories for evaluation of the time correlation functions of the backbone N-H and side chain methyl C-H vectors. The overall rotation was removed by RMS fitting each snapshot onto the average structure from the last 2 ns of the trajectories. The calculation of

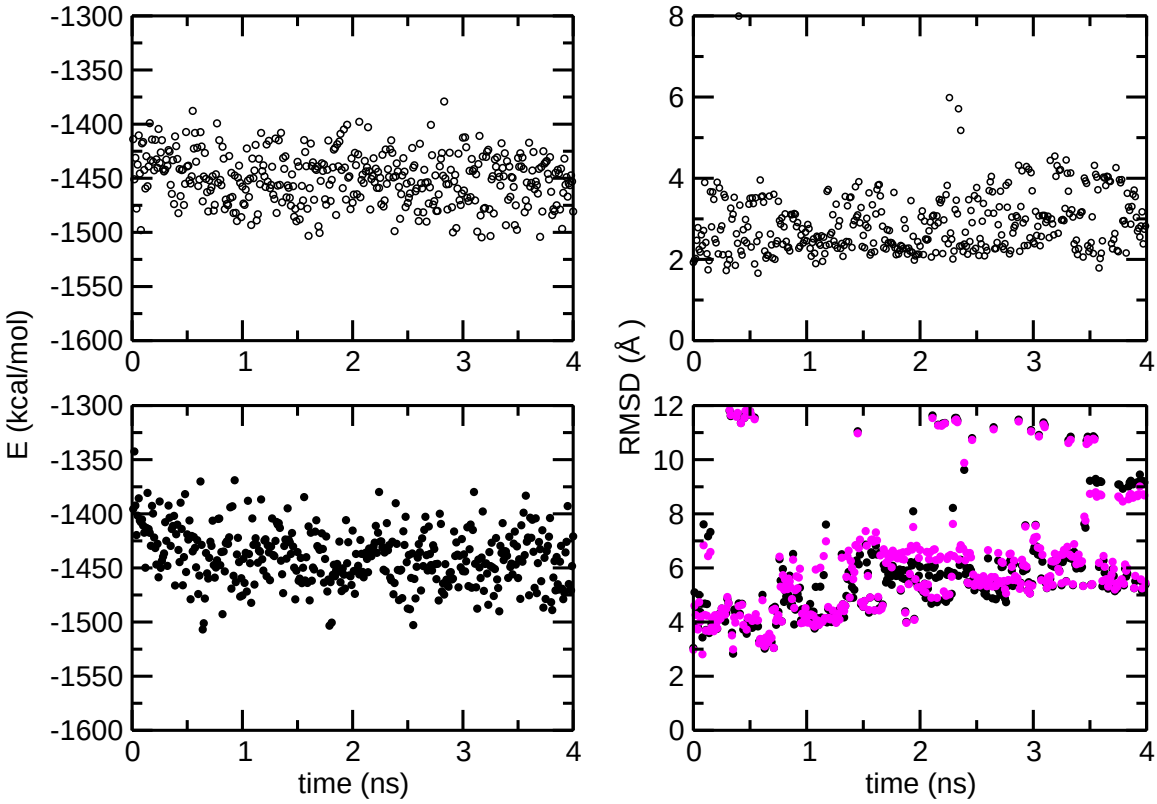

Figure S1: Convergence of the REX-CPHMD simulations initiated from the X-ray crystal (top, open symbols) and NMR solution (bottom, filled symbols) structures. The total energy and main-chain heavy atom RMS deviations with respect to the starting structure of HP67 sampled every 10 ps are plotted as a function of simulation time. For the simulation initiated from the NMR structure, RMS deviations with respect to the X-ray structure (in magenta) follow closely those with respect to the NMR structure (black), suggesting that state **1** and **2** are not closely related.

the NMR order parameters was performed using the NMR module in CHARMM (1) assuming a magnetic field strength of 11.74 T and an isotropic tumbling of 5000 ps. The full length of the correlation function is one fourth of the total trajectory length ($t_{\max} = 500$ ps). To minimize statistical uncertainty, simulations were repeated 3 times starting with a different initial velocity seed. The computed NMR parameters were averaged to give the reported values.

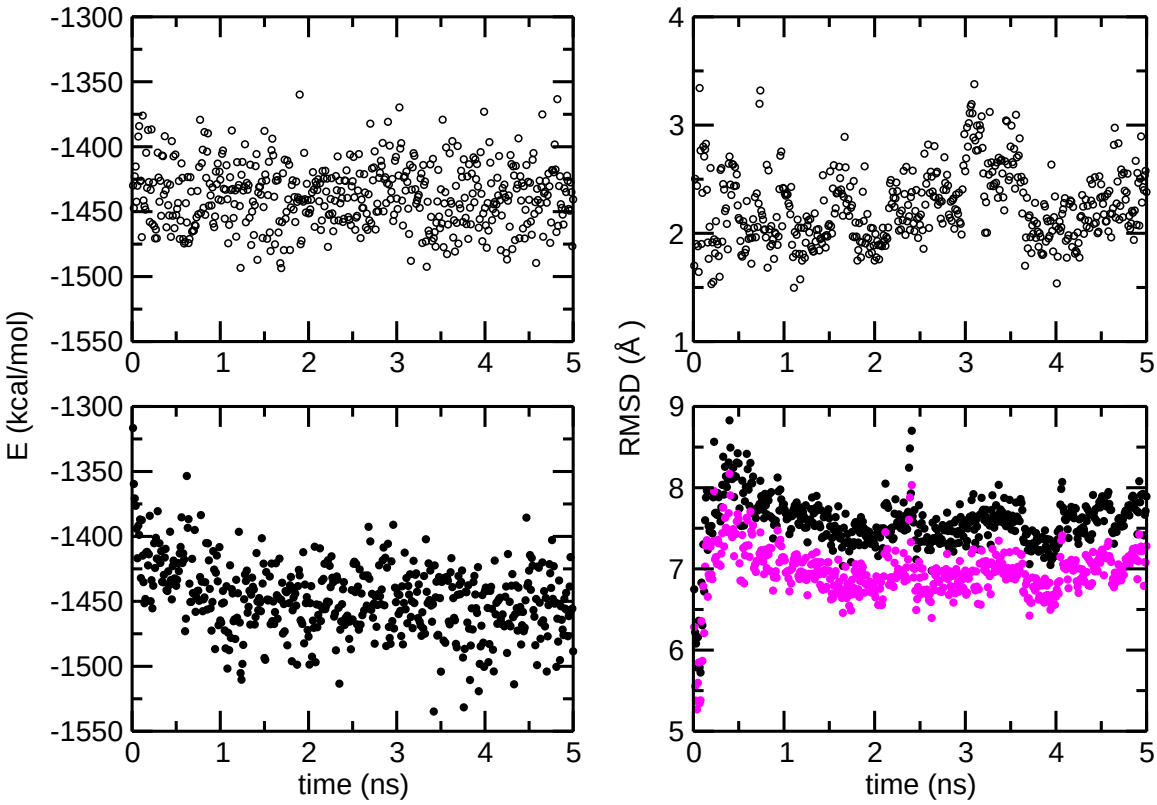

Figure S2: Convergence of the CPHMD simulations at 298 K and pH 7 starting from the X-ray crystal (top, open symbols) and NMR solution (bottom, filled symbols) structures. The total energy and main-chain heavy atom RMSD with respect to the starting structure of HP67 sampled every 10 ps are plotted as a function of simulation time. For the simulation initiated from the NMR structure, RMS deviations with respect to the X-ray structure are shown in magenta. The data were obtained from simulations initiated with the default velocity seed.

3 The N-terminal hydrophobic core is more accessible to solvent in the putative intermediate

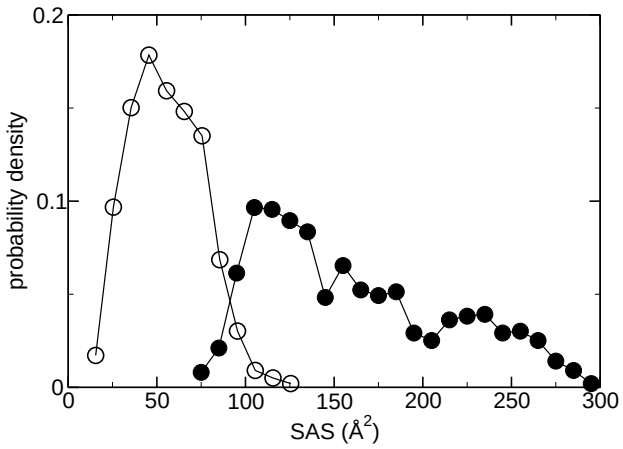

Figure S3: Probability density of solvent accessible surface area for the N-terminal hydrophobic core in state **1** (open symbols) and **2** (filled symbols). The bin width of the histograms is 10 Å^2.

References

[1] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983,** *4,* 187–217.

[2] Feig, M.; Karanicolas, J.; Brooks, III, C. L. *J. Mol. Graph. Model.* **2004,** *22,* 377–395.

[3] Lee, M. S.; Salsbury, Jr., F. R.; Brooks III, C. L. *Proteins* **2004,** *56,* 738–752.

[4] Khandogin, J.; Brooks III, C. L. *Biophys. J.* **2005,** *89,* 141–157.

[5] Khandogin, J.; Brooks, III, C. L. *Biochemistry* **2006,** *45,* 9363–9373.

[6] Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001,** *7,* 306–317.