Slow-Down in Diffusion in Crowded Protein Solutions Correlates with Transient Cluster Formation

Grzegorz Nawrocki[1#], Po-hung Wang[2#], Isseki Yu[2,3], Yuji Sugita[2,3,4,5], Michael Feig[1,4,*]

[1]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, United States
[2]RIKEN Theoretical Molecular Science Laboratory, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan
[3]RIKEN iTHES, 2-1 Hirosawa, Wako-shi, Saitama 351-0198, Japan
[4]RIKEN Quantitative Biology Center, Integrated Innovation Building 7F, 6-7-1 Minaotojima-Minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan
[5]RIKEN Advanced Institute for Computational Science, 7-1-26 Minatojima-Minamimachi, Chuo-ku, Kobe, Hyogo 650-0047, Japan

DETAILED METHODOLOGY

Molecular Dynamics Simulations

All-atom MD simulations of the one- and eight-villin systems were performed using the simulation programs NAMD[1] (version 2.10) for equilibration and CHARMM[2] (version 42a1) together with OpenMM[3] for GPU acceleration in production runs. The 64-villin system was simulated using GENESIS[4].

After 1,000 steps of conjugate gradient minimization the eight-villin systems were equilibrated by gradually increasing the temperature from 10 to 290 K at increments of 10 K, and then, finally, to 298 K. During the heating phase, positional restraints using a force constant of 1 kcal/mol/Å^2 were applied to protein heavy atoms and the system was equilibrated in the NPT ensemble at a pressure of 1 bar using a Langevin piston barostat (200 fs oscillation time, 100 fs damping time and a noise temperature set equal to the system temperature). Subsequently, production simulations were carried out over 2 μs in the NVT ensemble at 298 K. A 2-fs time step was used with a leap-frog integrator. SHAKE[5] was applied to holonomically constrain bonds involving hydrogen atoms in the proteins. A Langevin thermostat with a friction coefficient of 0.01 ps^{-1} was applied. It is well known that large friction coefficients used in Langevin thermostats, even when applied to explicit water systems, lead to a significant slow-down in diffusive properties.[6] While the choice of the Langevin thermostat was limited by OpenMM's functionalities, small friction coefficients are needed to reproduce diffusive properties obtained with other thermostats and in the NVE ensemble.[6] We chose a value here that is even smaller than a value of 0.1 ps^{-1} shown previously to result in water diffusion similar to NVE simulations[6]. This value was further validated by comparing translational diffusion coefficients obtained from simulations of dilute villin and villin at different concentrations with Langevin (OpenMM), Berendsen[7] (NAMD), and Bussi[8] (GENESIS) thermostats (see Table S1). The resulting diffusion coefficients are generally similar, although diffusion coefficients estimated via the

Langevin/OpenMM simulations appear to be slightly faster than with the Bussi and Berendsen thermostats.

The 64-villin system was energy-minimized using the steepest descent algorithm for 10,000 steps and then heated to 298 K during three short simulations in the NVT ensemble using an integration time step of 1 fs: 10 ps at 100 K, 10 ps at 200 K, and 10 ps at 298 K. During the heating step, positional restraints of 1 kcal/mol/Å^2 were applied to Cα atoms. The system was then further equilibrated for 110 ns at a time step of 2 fs in the NPT ensemble at 298 K and 1 bar (using Berendsen pressure coupling[7] with a coupling constant of 1.0 ps). The production run at a time step of 2 fs using the velocity-Verlet integrator[9] continued in the NVT ensemble at the same temperature and pressure for 2 μs. Bussi's velocity-rescaling scheme[8] with a coupling constant of 0.1 ps was applied for temperature coupling in all procedures unless stated otherwise. The bond lengths involving hydrogen atoms in proteins were constrained using the SHAKE[5] and RATTLE[10] algorithms. The SETTLE algorithm was used for constraining rigid water molecules.[11]

Analysis

Translational diffusion coefficient, D_t, was corrected for periodic boundary artifacts[12] according to:

$$D_{t,PBC} = \frac{k_B T}{6\pi \eta L}\left(\xi - \frac{4\pi R_h^2}{3L^2}\right) \qquad (S1)$$

with $\xi = 2.837$, the Boltzmann constant, k_B, the temperature, T=298 K, the length of the cubic simulation box, L, the hydrodynamic radius of villin, R_h=1.386 nm, estimated with HYDROPRO[13] from the native structure (PDB ID: 1VII), and the shear viscosity of the solvent, η. For crowded systems, the viscosity was adjusted from the viscosity of pure solvent, η_w, based on the volume fraction of the villin molecules (determined from R_h), φ, according to:[14]

$$\eta = \eta_w(1 + 2.5\varphi) \qquad (S2)$$

Corrections $D_{t,PBC}$ according to Eq. S1 withtout applying Eq. S2 amount to 0.272, 0.146, 0.190, 0.225, and 0.117 nm²/ns at dilute, 8, 16, 32 (8 copies) and 32 (64 copies) mM concentrations, respectively. With the viscosity correction according to Eq. S2, the total corrections become 0.272, 0.129, 0.145, 0.146, 0.076 nm²/ns at dilute, 8, 16, 32 (8 copies) and 32 (64 copies) mM concentrations, respectively. The total corrections account for for 59%, 32%, 47%, 66%, and 48% of the overall D_t values for 10-100 ns reported in Table 2 and similarly for the other time regimes. We note that the finite-size correction according to Eq. S1 was derived for low molecular weight compounds in relatively dilute coniditons and we are assuming here that the correction can be applied without modification to concentrated villin systems.

The kinetics of contact formation was extracted using correlation analysis based on the following conditional function that describes whether a contact initially present at time t_0 is still present at time $t_0+\Delta t$:[15]

$$P(\Delta t) = \frac{1}{N-k}\frac{1}{N_p}\sum_{j=1}^{N-k}\sum_{i=1}^{N_p}\delta_i(t_j)\delta_i(t_j + \Delta t) \qquad (S3)$$

where N is the number of snapshots, Δt is the k-th time interval ($\Delta t=k*0.1$ ps, $k=4500$), N_p is the number of protein pairs, and the function $\delta_i(t)$ assumes values of either 1 or 0 depending on whether or not a protein pair i is in contact at time t. Contacts of a given pair that are separated by at least one snapshot without the contact are assumed to be uncorrelated and excluded from the summation.

Coarse-Grained Simulations

Coarse-grained simulations were also performed using CHARMM. Coarse-grained protein models were built by replacing residues with beads at the Cα positions. The beads were harmonically restrained to the native structure using a force constant of 10 kcal/mol/Å² after least-squares superposition but without restricting rigid-body motions. Bonds between beads were constrained to

3.818 nm length (the average distance between subsequent Cα atoms) using a force constant of 20 kcal/mol/Å^2. The attractive interactions between beads were model via a Lennard-Jones potential (see Eq. 1) using a well depth of ε=2 kcal/mol. The neutral interactions between beads were modeled by setting ε=0.1 kcal/mol. The repulsive interactions were modeled through a Coulombic potential where small charges of q=0.1e were applied to each bead and ε=0.1 kcal/mol was used again as the well-depth in the Lennard-Jones potential. A distance of the Lennard-Jones potential minimum R^{min} was set to 3.49 nm for all the interaction types to matches the volume of the all-atom model. The simulations were performed using Langevin dynamics with a friction coefficient equal to 1 ps^{-1} for 100 ns each and time step 5 ps. The short-range non-bonded interactions were shifted to zero at 0.9 nm and the long-range were calculated using PME method[16] under PBC. The simulations were carried out at 300 K temperature and constant volume.

Table S1. Translational diffusion (obtained for 1-10 ns time regime) with different thermostats and simulation software for simulations of villin

Concentration [mM]	λ	Langevin OpenMM[1] γ=0.01 ps^{-1}	Berendsen NAMD[1] t=1 ps	Bussi GENESIS[2] t=0.1 ps
dilute	1.00	0.485 *(0.016)*		0.464 *(0.021)*
dilute	1.05	0.538 *(0.014)*		0.478 *(0.015)*
dilute	1.10	0.502 *(0.008)*		0.477 *(0.022)*
8	1.05	0.396 *(0.012)*		0.328 *(0.017)*
16	1.00	0.260 *(0.011)*		0.227 *(0.011)*
16	1.05	0.304 *(0.008)*		0.260 *(0.020)*
32	1.05	0.219 *(0.005)*		0.200 *(0.007)*
16	1.10	0.320 *(0.008)*	0.312 *(0.005)*	0.313 *(0.010)*

[1]Simulation length 2 μs; [2]simulation length 1 μs

Table S2. Generalized internal order parameter S_i^2 averaged over residues and time blocks or protein copies in simulations of diluted and concentrated villin solutions, respectively, with different protein-waters scaling factors λ

λ	dilute	8 mM	16 mM	32 mM
1.00	0.810 *(0.010)*	0.800 *(0.005)*	0.803 *(0.013)*	0.799 *(0.085)*
1.05	0.804 *(0.004)*	0.761 *(0.019)*	0.794 *(0.016)*	0.779 *(0.020)*
1.09	0.781 *(0.008)*	0.780 *(0.009)*	0.800 *(0.008)*	0.794 *(0.006)*
1.10	0.793 *(0.010)*	0.555 *(0.108)*	0.751 *(0.030)*	0.616 *(0.090)*
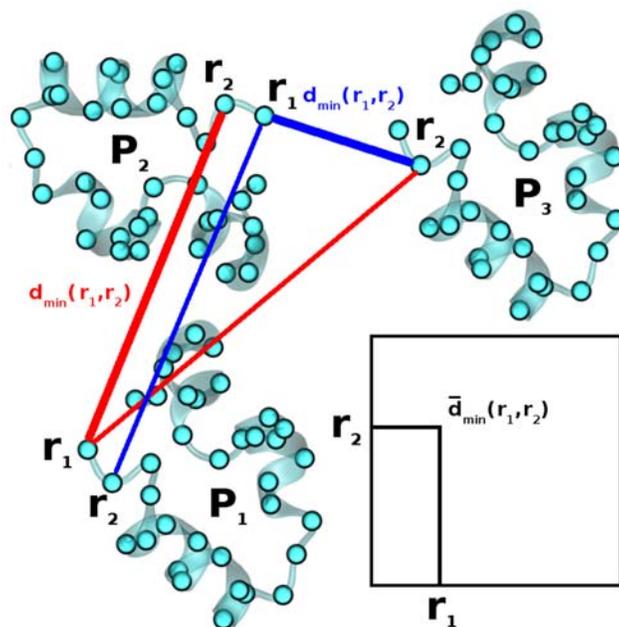
Figure S1. Minimum distance calculation for a residue pair (r_1, r_2). For residue r_1 of protein P_1 the closest r_2 is in P_2 (red line), but the closest residue r_2 from r_1 in P_2 is in protein P_3 (blue line).
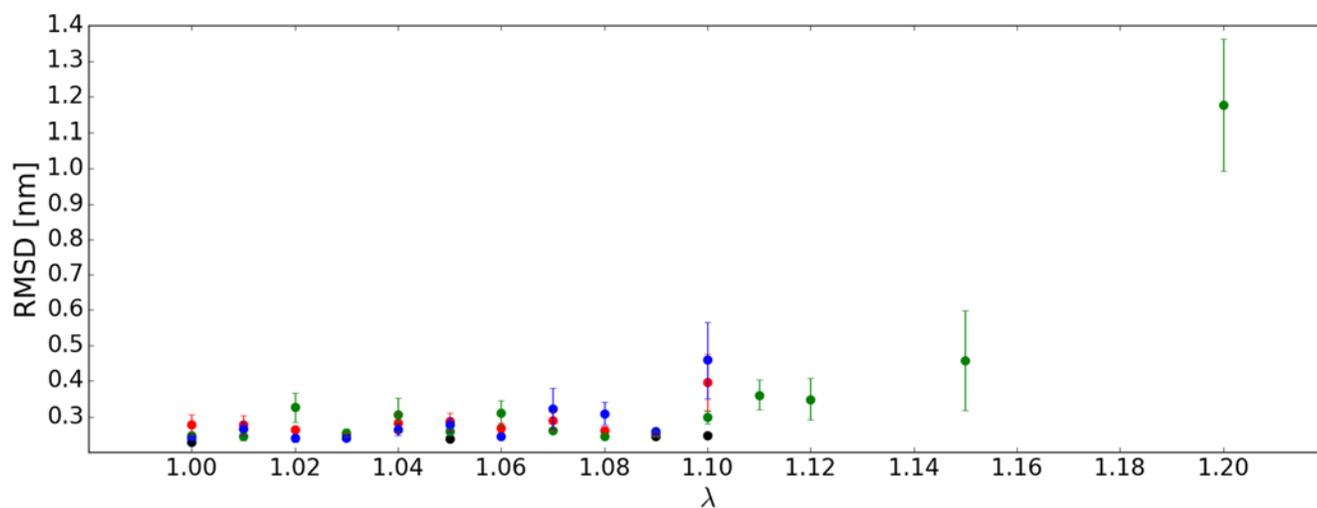
Figure S2. Average Cα RMSD of villin with respect to NMR structure (PDB: 1VII) at different concentrations (black: dilute, blue: 8 mM, green: 16 mM, red: 32 mM) as a function of the protein-water Lennard-Jones interaction scaling factor (see Eq. 1). Error bars indicate variations between different villin copies for concentrated systems and results from block averaging for the dilute system using 100 ns-blocks.
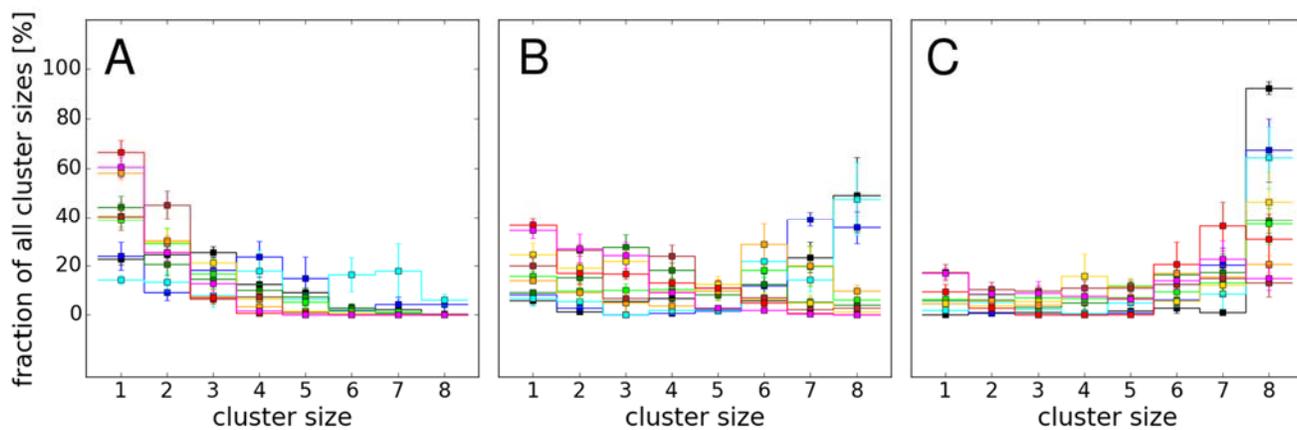
Figure S3. Distribution of clusters for villin simulations at 8 mM (A), 16 mM (B), and 32 mM (C) with different protein-water scaling factors (1.00 (black), 1.01 (blue), 1.02 (cyan), 1.03 (dark green), 1.04 (lime green), 1.05 (yellow), 1.06 (orange), 1.07 (brown), 1.08 (red) 1.09 (magenta)). Error bars are obtained from block averaging.
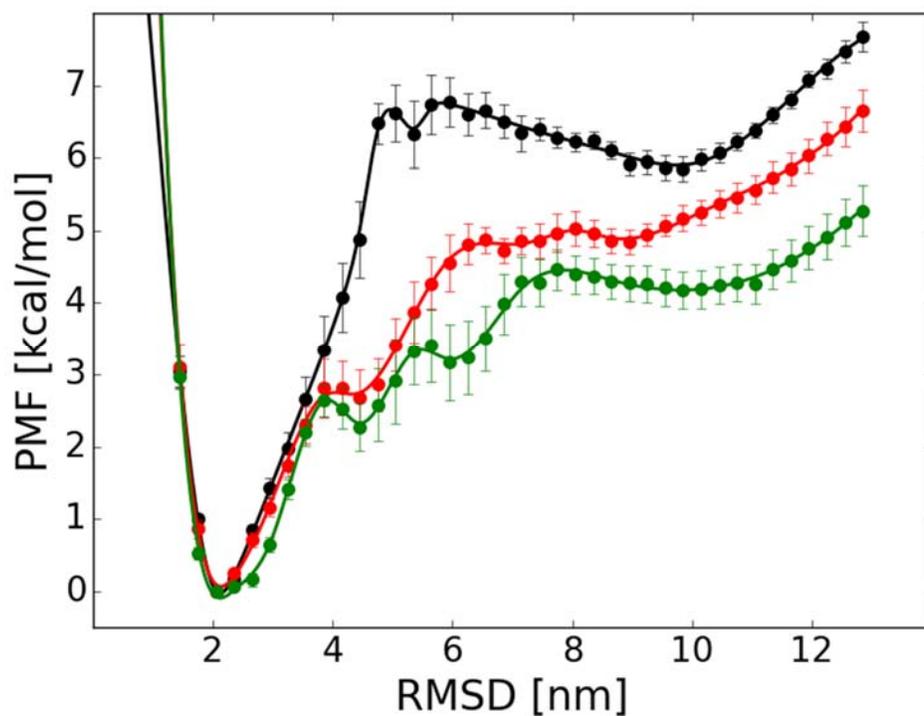
Figure S4. Potential of mean force (PMF) as a function of root-mean-square displacement (RMSD) obtained from sampling folding-unfolding of villin with Hamiltonian replica exchange molecular dynamics (H-REMD) simulations using $\lambda =1.00$ (black)[17], $\lambda =1.05$ (red), and $\lambda=1.10$ (green).**Figure S5.** Correlation functions for scalar product between random vectors fixed to the molecular frame used for describing rotational diffusion (see *Methods*) of individual villin copies (color lines) at 8 mM (A), 16 mM (B), 32 mM with eight copies (C), and 32 mM with 64 copies (D). Correlation functions averaged over villin copies are shown with error bars (black lines).
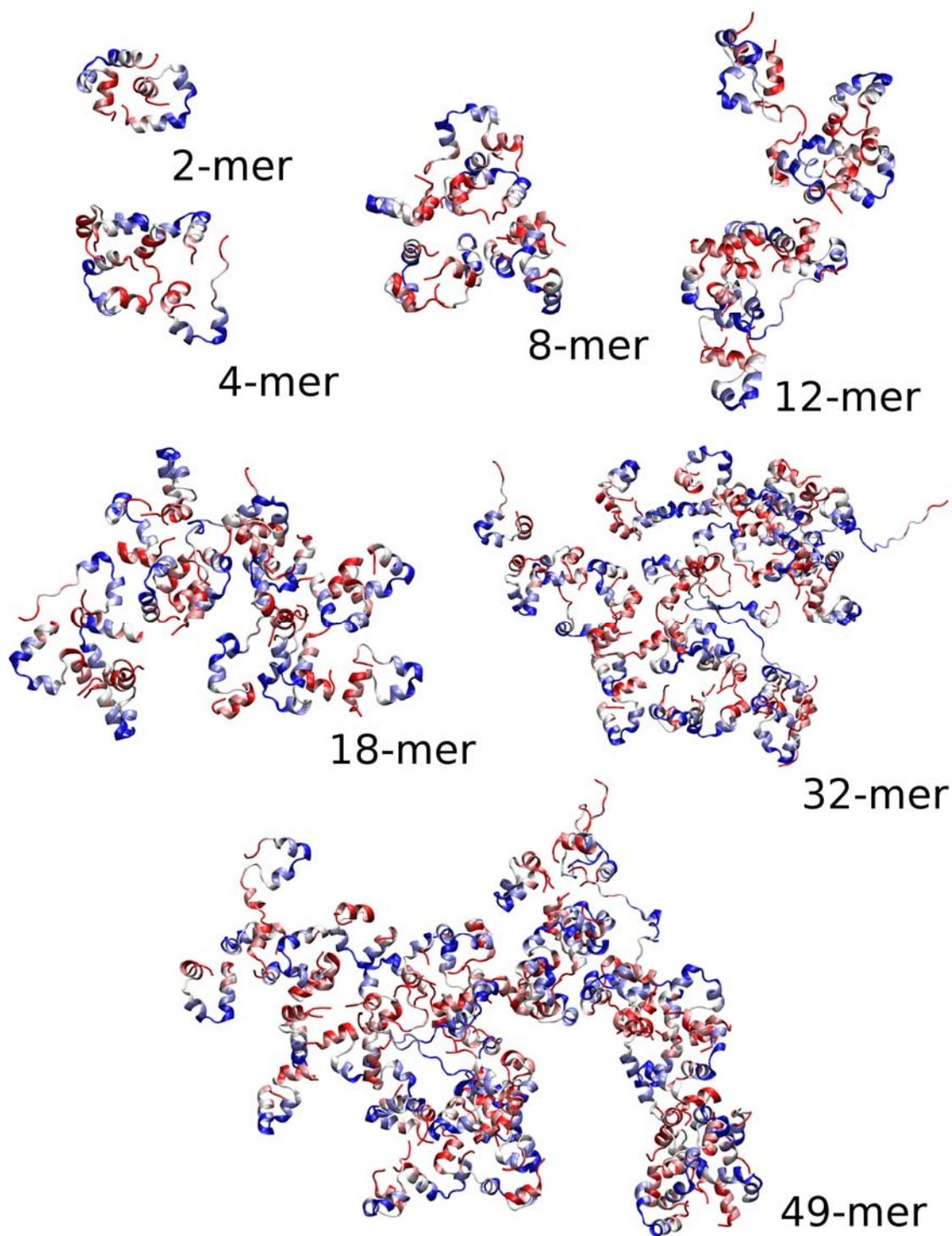
Figure S5. Representative transient cluster structures from the 64-copy simulation at 32 mM. Each villin is colored by residue according to their average protein-protein contact propensity (frequent contacts: red, infrequent contacts: blue).
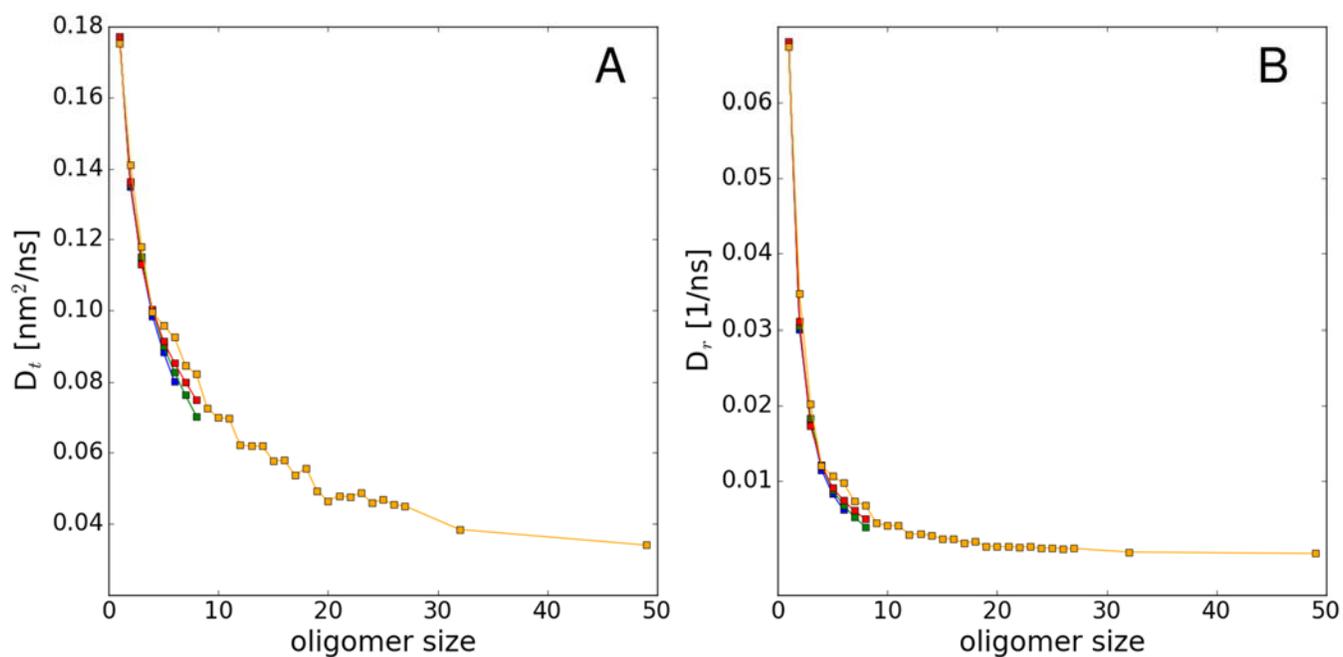
Figure S6. Translational (A) and rotational (B) diffusion calculated using HYDROPRO for cluster structures derived from the simulations at 8 mM (blue), 16 mM (green) and 32 mM (eight-copy: red; 64-copy: orange). For the clarity error bars obtained from block-averaging were omitted. None of the bars exceeded 2% for D_t or 5% for D_r.
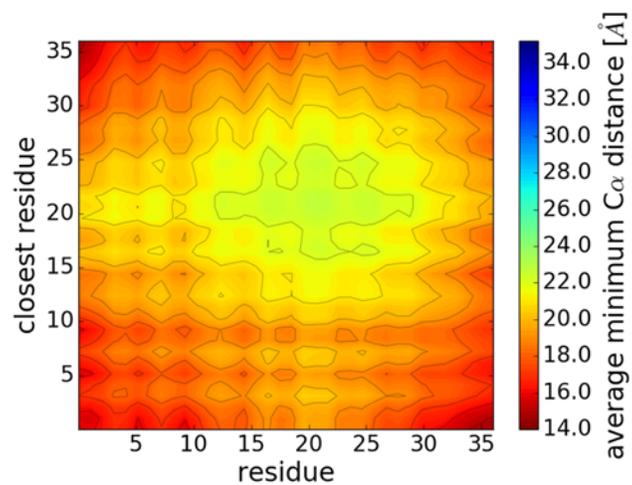
Figure S7. Contact map of residues between proteins at 32 mM in 64 copies system. Colors indicate the average Cα distances from a given residue to the closest residue of another nearby villin molecule.

Movie S1. Center of mass motion of villin molecules in the 64-copy, 32mM simulation. One highlighted villin molecule (in black) initially clusters with other villins shown in yellow, leaves the cluster, diffuses freely, and clusters with a different set of villins in magenta.

REFERENCES

1.   Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781-1802.

2.   Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S., et al., CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545-1614.

3.   Eastman, P.; Friedrichs, M. S.; Chodera, J. D.; Radmer, R. J.; Bruns, C. M.; Ku, J. P.; Beauchamp, K. A.; Lane, T. J.; Wang, L. P.; Shukla, D., et al., OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation. *J. Chem. Theory Comput.* **2013**, *9*, 461-469.

4.   Jung, J.; Mori, T.; Kobayashi, C.; Matsunaga, Y.; Yoda, T.; Feig, M.; Sugita, Y., GENESIS: A Hybrid-Parallel and Multi-Scale Molecular Dynamics Simulator with Enhanced Sampling Algorithms for Biomolecular and Cellular Simulations. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2015**, *5*, 310-323.

5.   Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C., Numerical-Integration of Cartesian Equations of Motion of a System with Constraints - Molecular-Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.

6.   Basconi, J. E.; Shirts, M. R., Effects of Temperature Control Algorithms on Transport Properties and Kinetics in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2013**, *9*, 2887-2899.

7.   Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular Dynamics with Coupling to an External Bath. *J. Phys. Chem. B* **1984**, *81*, 3684.

8.   Bussi, G.; Donadio, D.; Parrinello, M., Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

9.   Swope, W. C.; Andersen, H. C., A Computer Simulation Method for the Calculation of Equilibrium Constants for the Formation of Physical Clusters of Molecules: Application to Small Water Clusters. *J. Chem. Phys.* **1982**, *76*, 637-649.

10.   Andersen, H. C., Rattle - a Velocity Version of the Shake Algorithm for Molecular-Dynamics Calculations. *J. Comput. Phys.* **1983**, *52*, 24-34.

11.   Miyamoto, S.; Kollman, P. A., Settle - an Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952-962.

12.   Yeh, I. C.; Hummer, G., System-Size Dependence of Diffusion Coefficients and Viscosities from Molecular Dynamics Simulations with Periodic Boundary Conditions. *J. Phys. Chem. B* **2004**, *108*, 15873-15879.

13.   Ortega, A.; Amorós, D.; García de la Torre, J., Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. *Biophys. J.* **2011**, *101*, 892-898.

14.   Tanford, C., *Physical Chemistry of Macromolecules*. John Wiley & Sons: New York, 1961.

15.   Brunne, R. M.; Liepinsh, E.; Otting, G.; Wüthrich, K.; van Gunsteren, W. F., Hydration of Proteins: A Comparison of Experimental Residence Times of Water Molecules Solvating the Bovine Pancreatic Trypsin Inhibitor with Theoretical Model Calculations. *J. Mol. Biol.* **1993**, *231*, 1040-1048.

16.   Darden, T. A.; York, D.; Pedersen, L. G., Particle-Mesh Ewald: An N log(N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

17.   Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, *14*, 71-73.