PHYSICAL REVIEW E **81**, 030903(R) (2010)

Coarse-grained protein-protein stiffnesses and dynamics from all-atom simulations

Stephen D. Hicks and C. L. Henley
Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853-2501, USA
(Received 22 June 2009; published 11 March 2010)

Large protein assemblies, such as virus capsids, may be coarse-grained as a set of rigid units linked by generalized (rotational and stretching) harmonic springs. We present an *ab initio* method to obtain the elastic parameters and overdamped dynamics for these springs from all-atom molecular-dynamics simulations of one pair of units at a time. The computed relaxation times of this pair give a consistency check for the simulation, and we can also find the corrective force needed to null systematic drifts. As a first application we predict the stiffness of an HIV capsid layer and the relaxation time for its breathing mode.

DOI: 10.1103/PhysRevE.81.030903 PACS number(s): 87.14.E−, 87.10.Pq, 87.15.ap, 87.15.hg

I. INTRODUCTION

Large protein assemblies are pertinent to most of the soft-matter physics in cells; how can one calculate their elastic properties and corresponding dynamics? Such assemblies are too large to handle by all-atom simulations, but numerical coarse graining techniques are opening the door to direct simulations [1]. An ideal coarse graining would involve rather simple parametrizations for the purposes of human understanding, analytic treatment, transmission to other researchers, and building up coarse-grained models [2]. In this Rapid Communication we present an approach to extract such simplified parameters from all-atom molecular-dynamics (MD) simulations of small subsystems. Our approach should generalize well to any system in which the proteins are primarily structural. Moreover, because our simulations involve only a few proteins, they are tractable even without supercomputers.

Our program is to break up a large assembly as a network of many discrete units connected by several kinds of (typically) pairwise interactions modeled as generalized springs. We treat each unit—typically a single protein or domain—as a rigid body, which thus has only six degrees of freedom. We approach each interaction, *one kind at a time*, by simulating just the pair of interacting units and measuring the trajectory $\boldsymbol{x}(t)$ of the positions and orientations of each unit. Our aim is, from these observed trajectories, to extract the parameters for an effective Hamiltonian and equation of motion for each spring, and then reassemble the springs into the coarse-grained network.

We model $\boldsymbol{x}(t)$ as an overdamped random walk in a (biased) harmonic potential. This walk is parametrized primarily by two important tensors: one to describe the shape of the harmonic well, and the other to describe the (mainly hydrodynamic) damping and the associated stochastic noise. Combining these tensors gives a matrix whose eigenvalues are the relaxation rates. Thus by measuring short-time dynamic correlation functions together with the position distribution, we can identify whether the simulation has sufficiently sampled the equilibrium ensemble during the simulation time, or in case of an overall drift, we can compute the external forces needed to shift the equilibrium to the biologically proper configuration. This is similar in spirit to computing a potential of mean force or free-energy landscape with Jarzynski's equality [3], except that our coarse-grained $\boldsymbol{x}$ has more than one component, and (at minimum) represents angular degrees of freedom in addition to stretching. As an application, we simulate several interdomain interactions in the HIV capsid and estimate the Young's modulus (which can be measured experimentally) and Poisson ratio of the capsid lattice, as well as the relaxation rate of the breathing mode.

II. COARSE GRAINED STOCHASTIC DYNAMICS

We represent our system as a vector of generalized coordinates x_i, $i=1\ldots N$, where N is far smaller than the number of atoms and is obtained by some form of coarse graining. Our objective is to parametrize and determine from simulation (i) an effective free-energy potential function $U(\boldsymbol{x})$, and (ii) an equation of motion, for the coarse-grained coordinates.

We assume the coarse-grained degrees of freedom are overdamped: this is true at time scales much longer than the "ballistic scale" ($t_{\mathrm{bal}} \sim \Gamma m \sim 1$ ps, where $m \sim 10$ kDa is the mass of a protein) and the period of local bond vibrations (~ 0.1 ps). Then the dynamics is a continuous-time random walk,

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{\Gamma} \boldsymbol{f}(\boldsymbol{x},t) + \boldsymbol{\zeta}(t), \tag{1}$$

where $\boldsymbol{\Gamma}$ is the (symmetric) *mobility tensor*, $\boldsymbol{f}(\boldsymbol{x},t)$ is the force, $\boldsymbol{\zeta}(t)$ is a (Gaussian) stochastic function satisfying

$$\langle \boldsymbol{\zeta}(t) \otimes \boldsymbol{\zeta}(t') \rangle = 2\mathbf{D}\,\delta(t-t'), \tag{2}$$

and $\mathbf{D}$ is the *diffusion tensor*. For detailed balance, $\mathbf{D} = k_{\mathrm{B}} T \boldsymbol{\Gamma}$ at temperature T. We can expand the potential to second order about a point $\boldsymbol{x}_*$,

$$U(\boldsymbol{x}) = U_0 - \boldsymbol{f}_* \cdot (\boldsymbol{x}-\boldsymbol{x}_*) + \frac{1}{2}(\boldsymbol{x}-\boldsymbol{x}_*)\mathbf{K}(\boldsymbol{x}-\boldsymbol{x}_*), \tag{3}$$

where $\mathbf{K}$ is the (symmetric) *stiffness tensor*; then the force in Eq. (1) is $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}_* - \mathbf{K}(\boldsymbol{x}-\boldsymbol{x}_*)$. From measuring coordinate covariances in the simulation, we obtain $\mathbf{K}$,

$$\mathbf{G} \equiv \langle [\boldsymbol{x}-\boldsymbol{x}_*] \otimes [\boldsymbol{x}-\boldsymbol{x}_*] \rangle = k_{\mathrm{B}} T \mathbf{K}^{-1}. \tag{4}$$

If the static effective potential were our only interest, and if our runs were always long enough to equilibrate our sys-

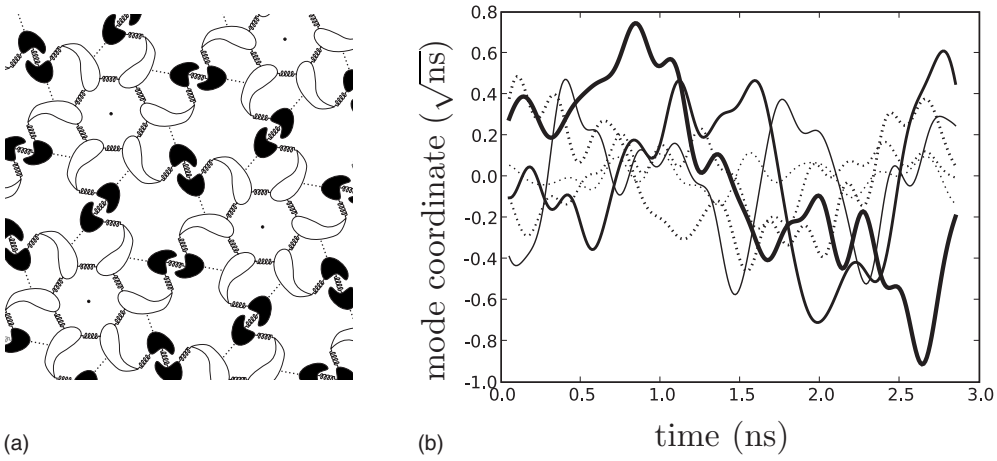

FIG. 1. (a) Diagram of interactions in the HIV capsid lattice. The black and white shapes represent the dimer-forming CTD and the hexamer-forming NTD, respectively. Springs represent the three different bonds we are interested in, and dotted lines represent the fourth bond we are ignoring. (b) Relaxation mode trajectories of linker. The mode coordinate has units of $\sqrt{\text{ns}}$ because it has been normalized by the noise. The slower modes are drawn with thicker lines. Note that the slowest mode has a very small drift, and we could correct this by applying an external force. The traces have been smoothed with a low-pass filter for readability.

tem, there would have been no need to model the dynamics [Eq. (1)]. As we do need the dynamics, we determine the diffusion tensor $\mathbf{D}$ (and hence $\mathbf{\Gamma}$) by measuring the correlation function at short times between the ballistic and relaxation time scales (see below) during which the deterministic term in Eq. (1) is less important than the noise:

$$\mathbf{D} = \frac{\langle [\boldsymbol{x}(t') - \boldsymbol{x}(t)] \otimes [\boldsymbol{x}(t') - \boldsymbol{x}(t)] \rangle}{2|t' - t|} \equiv \frac{\mathbf{W}(t' - t)}{2|t' - t|}. \quad (5)$$

We calculate D by fitting $\mathbf{W}(\Delta t)$ to a line over offsets $t_{\text{bal}} \ll \Delta t \ll \tau$, the relaxation times, weighting each point by $\sigma^2_{\mathbf{W}}(\Delta t) \propto (\Delta t)^3$. Notice that since $\mathbf{\Gamma}$ pertains to short-time dynamics, it is correctly measured even in runs too short to equilibrate in the potential well.

If we transform into coordinates $\tilde{\boldsymbol{x}} \equiv \mathbf{\Gamma}^{-1/2}\boldsymbol{x}$ then the equation of motion becomes

$$\frac{d\tilde{\boldsymbol{x}}}{dt} = \mathbf{\Gamma}^{1/2}\boldsymbol{f}_* - \mathbf{R}(\tilde{\boldsymbol{x}} - \tilde{\boldsymbol{x}}_*) + \tilde{\boldsymbol{\zeta}}(t), \quad (6)$$

where

$$\langle \tilde{\zeta}_\alpha(t)\tilde{\zeta}_\beta(t') \rangle = 2k_{\text{B}}T\delta_{\alpha\beta}\delta(t - t'), \quad (7)$$

and the *relaxation matrix* $\mathbf{R} = \mathbf{\Gamma}^{1/2}\mathbf{K}\mathbf{\Gamma}^{1/2}$ (which has units $[\text{time}]^{-1}$) is simply the stiffness tensor in our transformed frame. The eigenvalues of $\mathbf{R}$ are the decay rates τ_α^{-1} for the relaxation normal modes α.

The correlation time for a mode is the same as its relaxation time, so the relative error in $\mathbf{K}$ for mode α is of order $\sqrt{\tau_\alpha / \tau_{\text{run}}}$, where τ_{run} is the total run time. Thus, if all the $\tau_\alpha \ll \tau_{\text{run}}$, our estimate [Eq. (4)] of $\mathbf{K}$ is valid. But if $\tau_\alpha \sim \tau_{\text{run}}$ for some direction, not only are errors large but the initial deviation may still be relaxing over the entire run, which is often visible as a steady drift of the coordinates with mean velocity $\bar{\boldsymbol{v}}$. Averaging over time gives a large spurious variance in the drifting directions, leading to an underestimate of the corresponding stiffness.

III. APPLICATION TO HIV CAPSID

The elastic and dynamic properties of viruses in general are of particular importance in understanding the mechanisms by which they assemble and disassemble. The assembly must be reliable enough to produce capsids capable of surviving the harsh intercellular environment, while still being able to disassemble upon entering a new host cell. HIV in particular is unique because of its characteristic conical capsids [4], whose mechanism of formation is yet unsettled.

The HIV capsid protein (CA) consists of two globular domains: the larger 145-amino acid N-terminal domain (NTD) has a radius 1.3nm and the smaller 70-amino acid C-terminal domain (CTD) has a radius 1.7nm; we treat these as two separate units. The NTD and CTD are connected covalently by a flexible linker; there is also an NTD-NTD interaction (which forms hexamers in the capsid structure), a CTD-CTD interaction (which forms symmetric dimers in the structure), and an NTD-CTD interaction between neighboring proteins around a hexamer. These four interactions are shown in Fig. 1(a). We believe the NTD-CTD interaction to be the weakest, and the known structure is also poorest, so we will ignore it from now on. We therefore simulate each other pair in isolation, using structures from the Protein Data Bank [5].

We carried out our simulations using a modified version of the NAMD [6] package with the CHARMM22 force field. Our proteins are in a periodic cell 5 to 9 nm to a side using the TIP3P model for explicit water and 0.1 M NaCl, run with 2 fs time steps for a total of 3 ns each. We do most of the work at constant pressure and temperature (NPT), using a Langevin piston barostat at $P = 1$ atm, and a Langevin thermostat at $T = 310$ K and damping rate $\gamma_{\text{L}} = 5$ ps^{-1}. The NPT simulations model the statics well, but the thermostat's damping leads to unphysical dynamics with increased relaxation rates. This allows shorter simulations to equilibrate, but prevents us from determining the rates we should expect to see in reality. We therefore do a second measurement of dif-

TABLE I. Effective stiffness eigenvalues for pair simulations: NTD dimer, CTD dimer, and the NTD-CTD linker within the CA protein.

	$K_{\text{stretch}}^{(\text{eff})}$ (k_BT/nm^2)	$\mathbf{K}_{\text{orient}}^{(\text{eff})}$ eigenvalues (k_BT)				
NTD-NTD	12	1300	2800	4500	10000	18000
CTD-CTD	9.9	210	340	1100	3900	8300
Linker	2.8	130	250	480	1100	3800

fusion at constant volume and energy (NVE).

The center of mass and global rotation of the pair accounts for six trivial degrees of freedom; the remaining six represent the relative position and orientation of the two units. Of these six, only one is a pure translation: the distance $r=|\boldsymbol{r}_2-\boldsymbol{r}_1|$ between the center of each unit. The orientation of unit m can be represented by a rotation matrix $\boldsymbol{\Omega}_m$ which rotates the unit from its reference orientation by an angle $|\theta_m|$ about the axis $\hat{\theta}_m$. The even and odd combinations $\theta_1 \pm \theta_2$ give six degrees of freedom that comprise the remaining five coordinates, along with an overall rotation due to the even combination about the interbody axis $\boldsymbol{r}_2-\boldsymbol{r}_1$.

As we simulate just one pair of units from a protein complex, we omit the forces and torques on them due to the other units in the lattice, which generically had a nonzero resultant. In order to expand the free energy around the physiologically relevant configuration, we must add external forces to compensate; in light of Eq. (1) the correct force to impose is given by $\boldsymbol{f}_*=-\boldsymbol{\Gamma}^{-1}\overline{\boldsymbol{v}}$, where $\overline{\boldsymbol{v}}$ is the drift velocity measured in the absence of the compensating force. This was not important for the pairs reported in our results.

IV. RESULTS

The results for each simulation were similar, and the trajectory of the linker in the transformed relaxation mode coordinates is shown in Fig. 1(b), which is characteristic of all the observed trajectories. Once we have an equilibrated segment of a trajectory we use Eq. (4) to determine the 6×6 stiffness tensor $\mathbf{K}$; different components have different units, so it would be mathematically meaningless to diagonalize it directly. Instead, we define reduced stiffness tensors, representing the free-energy cost if we optimize r for a fixed set of angles and vice versa. Given

$$\mathbf{K}=\begin{pmatrix} K_{\text{rr}} & \mathbf{K}_{\text{r}\theta} \\ \mathbf{K}_{\theta\text{r}} & \mathbf{K}_{\theta\theta} \end{pmatrix}, \tag{8}$$

then integrating out the orientations (for K_{stretch}) and the stretch (for $\mathbf{K}_{\text{orient}}$) gives

$$K_{\text{stretch}}^{(\text{eff})}=K_{\text{rr}}-\mathbf{K}_{\text{r}\theta}\mathbf{K}_{\theta\theta}^{-1}\mathbf{K}_{\theta\text{r}} \tag{9}$$

$$\mathbf{K}_{\text{orient}}^{(\text{eff})}=\mathbf{K}_{\theta\theta}-\mathbf{K}_{\theta\text{r}}K_{\text{rr}}^{-1}\mathbf{K}_{\text{r}\theta}. \tag{10}$$

The eigenvalues of these reduced tensors are given in Table I.

We computed the stiffness tensor implicitly in the relative coordinates between the two bodies, but the absolute coordinates are the natural frame for computing the noise. Measuring the diffusion of a single body in an NVE simulation yields a mean $D_{\text{CTD}}^{(\text{rot})}=0.11$ rad^2/ns and $D_{\text{NTD}}^{(\text{rot})}=0.044$ rad^2/ns. If we approximate each domain as a solid sphere then Stokes' law gives a rotational diffusion constant $D^{(\text{rot})}=k_BT/(8\pi\eta r^3)$ [7]. We thus expect $D_{\text{CTD}}^{(\text{rot})}=0.11$ rad^2/ns and $D_{\text{NTD}}^{(\text{rot})}=0.050$ rad^2/ns using a viscosity $\eta^{(310\text{ K})}=0.69$ cP. The accepted TIP3P viscosity $\eta^{(\text{TIP3P})}=0.31$ cP gives poorer agreement.

TABLE II. NPT time constants for the relaxation modes of each pair.

	Relaxation times τ_α (ps)					
NTD-NTD	120	23	18	9.3	6.0	4.4
CTD-CTD	76	26	24	7.8	5.4	4.1
Linker	190	140	80	76	22	8.3

The translational diffusion constant is slightly harder to measure, since it is influenced significantly by the finite-size effect [8]. This can be corrected for by measuring the diffusion at several box side lengths L and using a linear fit of $D^{(\text{tr})}$ versus $1/L$ to extrapolate to $1/L=0$. Doing so yields $D_{\text{CTD}}^{(\text{tr})}=55$ Å^2/ns and $D_{\text{NTD}}^{(\text{tr})}=27$ Å^2/ns. Stokes' law gives expected $D_{\text{CTD}}^{(\text{tr})}=56$ Å^2/ns and $D_{\text{NTD}}^{(\text{tr})}=43$ Å^2/ns using $\eta^{(\text{TIP3P})}=0.31$ cP. The measured $D^{(\text{tr})}$ has a significantly larger relative error than $D^{(\text{rot})}$, due to the finite-L extrapolation.

We can diagonalize the relaxation matrix to compute the relaxation modes for each linkage. The NPT relaxation times from this calculation are listed in Table II. All the times are significantly shorter than the simulation time, so we can be confident that the simulations are equilibrated.

Finally, we can compose these generalized springs together into a triangular lattice as shown in Fig. 1(a), with an NTD hexamer at each vertex, a CTD dimer at the midpoint of each edge, and a spring connecting each domain, whose free energy is given by the relative positions multiplied into the appropriate stiffness tensor. We can then determine the free-energy minimum as a function of periodic cell dimensions to find a lattice constant of $a=9.1$ nm. This is slightly smaller than the experimentally measured 10.7 nm [4], which may be largely due to our sheet being flat, rather than curved into a tube. Computing the free energy of simple extension yields a two-dimensional Young's modulus of $0.92k_BT/\text{Å}^2=0.39$ N/m and a Poisson ratio of 0.30. Assuming homogeneity and a thickness of 5 nm, we find a three-dimensional Young's modulus of 77 MPa (compared with 115 MPa measured using atomic force microscopy [9]).

Furthermore, we can estimate the relaxation rate of the full-capsid breathing mode in water by further coarse graining to a single coordinate a representing a uniform dilation in the plane, which has dynamics given by Eq. (1) with stiffness and mobility constants K_a and Γ_a. The projected stiffness is given by the bulk modulus $K_a=4K_{[2\text{d}]}=2.6k_BT/\text{Å}^2$, calculated from the $2d$ Young's modulus and Poisson ratio. To project the damping term, we observe that all the actual motion in the breathing mode of a virus capsid of radius r is in

the radial direction, and we thus need to scale the capsid protein's translational diffusion constant by $(da/dr)^2$ to find the diffusion constant for a. Using the detailed balance condition,

$$\Gamma_a = \frac{16\pi\sqrt{3}}{N} \frac{D_{\mathrm{NTD}}^{(\mathrm{tr})} + D_{\mathrm{CTD}}^{(\mathrm{tr})}}{k_{\mathrm{B}}T} \frac{\eta^{(\mathrm{TIP3P})}}{\eta^{(310\ \mathrm{K})}}, \tag{11}$$

where $N=16\pi\sqrt{3}r^2/a^2$ is the total number of capsid proteins [10]. Taking $N=1500$ proteins as the average size for an HIV capsid thus gives a relaxation rate of 6.1 ns^{-1} for the breathing mode.

V. DISCUSSION

In conclusion, we have reported a model of overdamped random walks in which the statics and dynamics are described respectively by complementary "stiffness" and "mobility" tensors. From these two tensors a "relaxation matrix" can be formed, the eigenvalues of which give the relaxation rates and provide a convergence test for simulations. We demonstrated the usefulness of this model in extracting coarse-grained elastic constants from molecular-dynamics trajectories of pairs of interacting units.

While our relaxation formalism is unusual in combining stochastic dynamics with a realistic multicomponent spring, it bears some similarities to certain more familiar techniques. Normal mode analysis, and in particular, Gaussian network models, replace interactions (either between atoms or groups of atoms) with springs of uniform stiffness [11]. While these techniques have been successful in determining the soft degrees of freedom to explain reaction pathways such as virus maturation [12,13], the frequencies themselves are well known to be artificial because they omit the damping forces of the surrounding water (this has been addressed by Lamm and Szabo [14] with their "Langevin modes"). Additionally, most of these techniques are insensitive to point mutations or environmental conditions. On the other hand, "essential dynamics" (or "principal component analysis") [15] uses all-atom simulations to determine the soft modes of a system. Moreover, Hayward *et al.* [16] suggested specifying important modes *a priori*, which is our starting point. We extend these approaches by using the same principal components (to wit, the position and orientation) for each unit to connect the different interactions, and by using the relaxation modes to predict the dynamics *ab initio*.

Our technique is not specific to virus capsids: a similar approach should be applicable to many other systems of interacting protein domains, such as microtubules or BAR domains. Among virus capsids, HIV was particularly amenable because all the important interactions are pairwise, while many other viruses are complicated by long tails in which all six molecules in the hexamer are entwined. In the future, we plan to look at the effect of point mutations, salinity, and pH on the resultant elasticity (which could then be verified by experiment); and to investigate further the properties of our generalized springs. We hope that the techniques presented here will provide a convenient middle ground between the atomistic and continuum pictures of many biological systems.

ACKNOWLEDGMENTS

We thank D. Murray, V. M. Vogt, M. Widom, H. Weinstein, D. Roundy, W. Sundquist, and M. Yeager. This work was supported by DOE Grant No. DE-FG02–89ER-45405. Computing facilities were provided through the Cornell Center for Materials Research under NSF Grant No. DMR-0079992.

[1] A. Arkhipov, P. Freddolino, and K. Schulten, Structure (London) **14**, 1767 (2006).

[2] R. Zandi *et al.*, Proc. Natl. Acad. Sci. U.S.A. **101**, 15556 (2004); S. D. Hicks and C. L. Henley, Phys. Rev. E **74**, 031912 (2006); M. F. Hagan and D. Chandler, Biophys. J. **91**, 42 (2006).

[3] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997); Eur. Phys. J. B **64**, 331 (2008).

[4] S. Li, C. P. Hill, W. I. Sundquist, and J. T. Finch, Nature (London) **407**, 409 (2000).

[5] For the full-length protein (linker simulation) we use cryo-EM structure 3DIK [B. Ganser-Pornillos, A. Cheng, and M. Yeager, Cell **131**, 70 (2007)]; for the NTD we use the NMR structure 1GWP [C. Tang, Y. Ndassa, and M. F. Summers, Nat. Struct. Biol **9**, 537 (2002)]; fitted to the homologous MLV hexamer crystal structure 1U7K [G. Mortuza *et al.*, Nature (London) **431**, 481 (2004)]; and for the CTD we use the crystal structure 1AUM [T. Gamble *et al.*, Science **278**, 849 (1997)].

[6] J. Phillips *et al.*, J. Comput. Chem. **26**, 1781 (2005).

[7] H. Lamb, *Hydrodynamics*, 6th ed., (Cambridge, Cambridge, 1932), p. 589.

[8] I. C. Yeh and G. Hummer, Biophys. J. **86**, 681 (2004).

[9] N. Kol *et al.*, Biophys. J. **92**, 1777 (2007).

[10] We have included a term $\eta^{(\mathrm{TIP3P})}/\eta^{(310\ \mathrm{K})}$ to correct for the TIP3P water model's incorrect viscosity (which affected the diffusion constants we measured in our simulations) so that we can estimate the relaxation rate in real water.

[11] M. M. Tirion, Phys. Rev. Lett. **77**, 1905 (1996); I. Bahar, A. R. Atilgan, M. C. Demirel, and B. Erman, *ibid.* **80**, 2733 (1998); F. Tama, M. Valle, J. Frank, and C. L. Brooks, Proc. Natl. Acad. Sci. U.S.A. **100**, 9319 (2003); M. Gibbons and W. Klug, J. Mater. Sci. **42**, 8995 (2007).

[12] A. Rader, D. Vlad, and I. Bahar, Structure **13**, 413 (2005).

[13] E. R. May (personal communication) has calibrated the stiffnesses in an elastic network model to all-atom MD of the HK97 phage (mature) capsid.

[14] G. Lamm and A. Szabo, J. Chem. Phys. **85**, 7334 (1986).

[15] T. Horiuchi and N. Go, Proteins **10**, 106 (1991); T. Ichiye and M. Karplus, *ibid.* **11**, 205 (1991); A. Amadei, A. Linssen, and H. Berendsen, *ibid.* **17**, 412 (1993).

[16] S. Hayward, A. Kitao, and H. Berendsen, Proteins **27**, 425 (1997).