Viral structural transitions: An all-atom multiscale theory

Yinglong Miao and Peter J. Ortoleva

Citation: The Journal of Chemical Physics **125**, 214901 (2006); doi: 10.1063/1.2400858
View online: http://dx.doi.org/10.1063/1.2400858
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/125/21?ver=pdfcov
Published by the AIP Publishing

Viral structural transitions: An all-atom multiscale theory

Yinglong Miao and Peter J. Ortoleva[a)]
Center for Cell and Virus Theory, Department of Chemistry, Indiana University, Bloomington, Indiana 47405

(Received 18 July 2006; accepted 31 October 2006; published online 1 December 2006)

An all-atom theory of viral structural transitions (STs) is developed based on a multiscale analysis of the *N*-atom Liouville equation. The approach yields an understanding of viral STs from first principles and a calibrated interatomic force field. To carry out the multiscale analysis, we introduce slow variables characterizing the whole-virus dynamics. Use of the "nanocanonical ensemble" technique and the fundamental hypothesis of statistical mechanics (i.e., the equivalence of long-time and ensemble averages) is shown to imply a Fokker-Planck equation yielding the coarse-grained evolution of the slow variables. As viral STs occur on long time scales, transition state theory is used to estimate the energy barrier of transition between free energy wells implied by observed hysteresis in viral STs. Its application to Nudaurelia capensis ω virus provides an upper bound on the free energy barrier when a single dilatational order parameter is used. The long time scale of viral STs is shown to follow from the aggregate effect of inertia, energy barrier, and entropic effects. Our formulation can be generalized for multiple order parameter models to account for lower free energy barrier pathways for transition. The theory with its all-atom description can be applied to nonviral nanoparticles as well. © *2006 American Institute of Physics*. [DOI: 10.1063/1.2400858]

I. INTRODUCTION

Viruses are composites of proteins, genetic material, and other constituents organized into a nanometer-scale structure. They are supramillion atom in size and hence dramatic changes in structure, i.e., large conformational changes or structural transitions (STs) with characteristics similar to macroscopic phase transitions, are expected and observed. The transitions occur in response to condition changes in the host medium, such as *p*H, ionic strength, and cation concentrations. For example, native cowpea chlorotic mottle virus (CCMV) capsid swells dramatically as *p*H is increased from 5.0 to 7.5 in the absence of divalent cations[1–5] (Fig. 1). Nudaurelia capensis ω virus[6–9] (NωV) and HK97 bacteriophage[10,11] undergo large conformational changes during capsid maturation. Poliovirus undergoes irreversible receptor-mediated STs upon cell entry.[12–16] More details of these STs will be discussed in Sec. III. The objective of this study is to place such phenomena in the framework of rigorous statistical mechanics, i.e., to show how they emerge from an all-atom description via a multiscale analysis of the Liouville equation. We seek to understand how the coherent responses of a nanoscale life form emerge from atomistic chaos. To do so we introduced novel methods for introducing structural order parameters and for deriving Fokker-Planck (FP) equations for their stochastic dynamics. In this way we place virology within the framework of chemical physics.

In principle, a viral ST can be understood in terms of the dynamics of a set of atoms evolving classically in an interatomic force field. The benefit of such an approach is that it enables a general, parameter-free predictive virus model. The challenge is to implement this approach as a practical computational algorithm. A central objective of the present approach is to capture atomic-scale detail and project the results into whole-virus scale responses.

Theoretical approaches have been developed to simulate viral processes. These include lumped models,[17] symmetry-constrained models,[18] computational molecular dynamics,[18–30] and normal mode analysis[31–34] as reviewed earlier.[35] However, they do not simultaneously capture atomic-scale detail and whole-virus, long-time behaviors in a nonsymmetry-constrained fashion or lead to a computational algorithm that accounts for the highly and often symmetry-breaking nonlinear nature of the phenomena of interest. While there has been an all-atom molecular dynamics simulation on a complete satellite tobacco mosaic virus with up to 1×10^6 atoms without symmetry constraint by using 256 processors and 128 gigabytes memory on an SGI 1024-processor Altix system with a performance of 1.1 ns/day,[36] the computation took 50 days to capture, about 55 ns of physical time, the time scale of which is still very short in comparison with that of viral STs (subseconds to minutes). Therefore there has not been an approach that enables the computation in virology needed for drug and vaccine design or for fundamental studies of interest, e.g., interactions of viruses with host cell receptors or the initiation and propagation of a capsid ST. Thus the challenge remains to develop methods that preserve the atomic/whole-virus interplay underlying many bionanostructural phenomena.

The great size of a virus relative to that of an atom suggests that key elements of viral life-cycle events display characteristic aspects of Brownian motion, i.e., arise out of the cross talk between slow nanoscale dynamics and rapidly fluctuating atomic behaviors taking place across multiple time and length scales. Brownian behavior has been shown

[a)]URL: http://sysbio.indiana.edu FAX: 812-855-8300. Electronic mail: ortoleva@indiana.edu

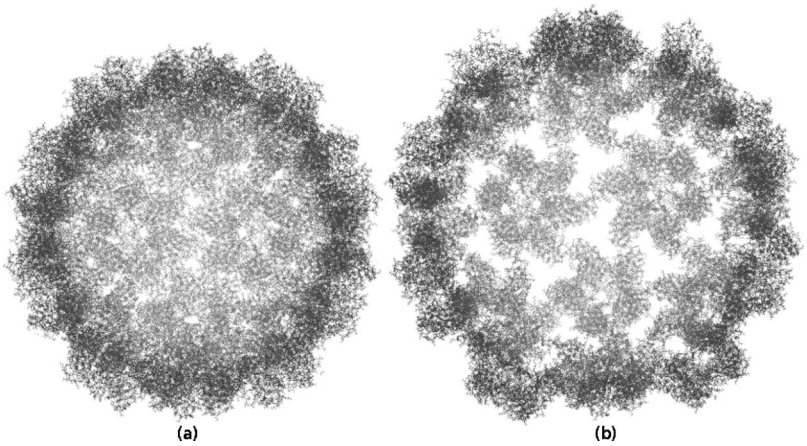

FIG. 1. (a) Native all-atom CCMV capsid and (b) swollen CCMV capsid: pentamers are translated by 24 Å relative to the c.m. of the whole capsid and rotated counterclockwise by 9° along their fivefold axes; hexamers are translated by 21 Å and rotated counterclockwise by 8° along their threefold axes (Ref. 3).

to follow from a reduction (i.e., coarse graining) of the Liouville equation, and this theme is reexamined here in the context of viral migration and STs. We examine the energetics and kinetics of viral processes as implications of their multiscale character. The basis of our approach[35] has its origins in a long history of multiscale analysis that started with Einstein[37] and continues to this century.

Deutch and Oppenheim[38] presented an approach to study structureless Brownian particle (BP) dynamics based on the use of projection operators and a perturbation scheme developed in the host particle/BP mass ratio. This approach sets the stage for a series of studies of BP dynamics based on the Liouville equation and BP models that did not account for internal molecular structure (i.e., were structureless). Shea and Oppenheim[39] derived FP and Langevin equations for a single structureless BP in a both of small particles via projection operators, a perturbation expansion in the mass ratio, and the assumption that gradients in the host medium are small. Peters[40,41] derived FP equations for the coupling of rotational and translational motions of a structureless nonspherical BP near a surface. Shea and Oppenheim[42] analyzed the case of multiple BPs, introducing a number of smallness parameters including the mass ratio. Ortoleva[43] presented an approach based on a formal multiple space-time scaling approach integrated with a statistical argument derived from the BP/host particle size ratio and BP geometry that allows for a united asymptotic expansion for solving the Liouville equation; the result is a FP equation for single and multiple BPs and intra-BP dynamics. It also was shown formally that this approach leads to a set of coupled FP equations, one for each slow host mode when the slow hydrodynamics of the host medium is accounted for.

While an all-atom approach to structured nanoparticles seems natural, technical difficulties have hindered progress. These stem from the need to preserve the total number of degrees of freedom while simultaneously introducing slow variables including the center of mass (c.m.) and overall orientation, as well as structural order parameters. Recently this problem was solved by introducing the notion that these slow variables are not to be thought of as new dynamical variables,[35] rather they are a way to make the multiscale character of the N-atom probability density explicit. Using this formulation, the chain rule, and a "nanocanonical ensemble" technique, the need for tedious bookkeeping to ensure conservation of the total number of degrees of freedom was avoided. A further advance of this nanocanonical approach is to recognize that the solvability conditions used to derive the FP equation can be obtained from an analysis of the time dependence of the terms in a multiscale perturbation analysis of the Liouville equation directly, and not by integration over a reduced set of atomic degrees of freedom (i.e., of the host atoms) that is only natural for structureless nanoparticles, but which cannot be readily carried out for the structured nanoparticles of interest. The fast variable integration is avoided in Ref. 35 by using the fundamental hypothesis of statistical mechanics, "the long-time and ensemble averages are equivalent near equilibrium for systems with slowly varying collective behaviors."

While the theory can be applied to all dynamical nanoparticles due to its all-atom description of the system,[35] here we focus on understanding viral STs. The essence of the physical picture adopted is suggested in Fig. 2. A virus is a massive and geometrically large aggregate of atoms immersed in a host medium that subjects it to frequent collisions. In the absence of macroscopic gradients in the host medium, these collisions result in a fluctuating force on the viral c.m. that averages out in time and across the surface of the virus but lead to a coupled Brownian translational/rotational/ST dynamics. Our approach preserves the dual na-

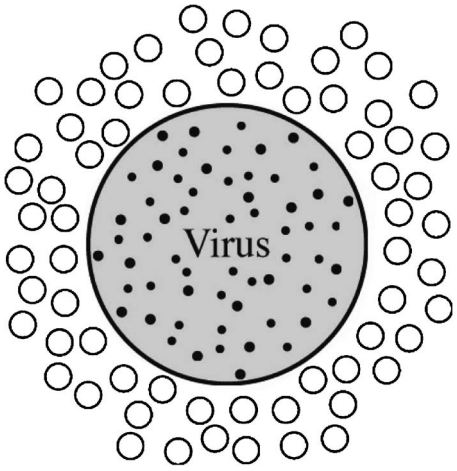

FIG. 2. A virus with a massive and geometrically large aggregate of atoms immersed in a host medium that subjects it to frequent collisions.

ture of the composite virus/host system. It is only a myth of our interest in the virus that it is treated as special. In reality this system is a collection of N atoms wherein any subsystem (e.g., the virus) exchanges momentum and energy with the remainder (e.g., the host) mediated by the interatomic force field.

In this study, the all-atom Liouville equation for a virus in a host medium is solved perturbatively in a small parameter to "unfold" the various scales characterizing the state of the virus; we then "recompose" the perturbation hierarchy into a FP equation for viral migration and ST. The notion of operator scaling[44] is used in obtaining the FP equation yielding the reduced probability density for the slow viral dynamics. Thermodynamic forces and frictional effects mediating migration and a ST emerge naturally in the course of deriving the FP equation. The stochastic dynamics provides the fluctuating kinetic features of the theory associated with barrier crossing. Free energy landscapes and frictional energy exchange are accounted for, and transition state (TS) theory estimates of transition energy barrier are obtained from experimental data. In Sec. II slow variables capturing a viral dilatational ST are introduced and a self-consistent FP equation is derived for these slow variables via a multiscale analysis. In Sec. III we discuss the computation of free energy and the friction coefficients; a double-well free energy profile is shown to be compatible with the capsid ST during NωV maturation using observed hysteretic effects in the ST. TS theory for obtaining approximate solutions to the FP equation and estimating the free energy barrier of viral STs, as well as an application of the theory to NωV capsid ST, are presented in Sec. IV. Conclusions and a plan for future studies are presented in Sec. V.

II. ALL-ATOM MULTISCALE ANALYSIS OF A VIRAL DILATATIONAL ST

An all-atom multiscale analysis of the Liouville equation starts with the identification of a set of slow variables. In hydrodynamics, conserved quantities are the traditional choice. For the present application, we suggest that slow variables can be chosen according to the following criteria:

(A) They are expressible in terms of the $6N$ atomic momenta and positions;
(B) they can be shown via Newton's equations to evolve on a time scale that is long compared to that of atomic vibrations or collisions;
(C) they capture the nanoscale phenomena of interest;
(D) the set of slow variables is complete, i.e., they are not strongly coupled to other slow variables not included in the model; and
(E) the energy of the system can be expressed in terms of these variables and residual rapidly fluctuating atomistic variables.

As there are initial data or applied fields that can make any variable change rapidly, a corollary to criterion (B) is that the special conditions on problems of interest are met. Such slow variables capturing viral dilatation are constructed below and a FP equation for their long time scale dynamics is derived.

First, denote the c.m. position of the virus $\mathbf{R}^*$:

$$\mathbf{R}^* = \sum_{i=1}^{N} \frac{m_i \mathbf{r}_i}{m^*} \Theta_i, \quad m^* = \sum_{i=1}^{N} m_i \Theta_i, \tag{2.1}$$

where $\Theta_i = 1$ if atom i is in the virus and 0 otherwise, and m^* is the total viral mass. Newton's equations imply that $d\mathbf{R}^*/dt = -\mathcal{L}\mathbf{R}^*$, in which $\mathcal{L}$ is the Liouville operator

$$\mathcal{L} = -\sum_{i=1}^{N} \left[\frac{\mathbf{p}_i}{m_i} \cdot \frac{\partial}{\partial \mathbf{r}_i} + \mathbf{F}_i \cdot \frac{\partial}{\partial \mathbf{p}_i} \right], \tag{2.2}$$

where $\mathbf{F}_i = -(\partial V/\partial \mathbf{r}_i)_{\mathbf{r}_{j\neq 1}}$ is the force on atom i and $V(\mathbf{r}_1, \ldots, \mathbf{r}_N)$ is the N-atom potential. With this and introducing the total viral momentum $\mathbf{P}^*$, we obtain $d\mathbf{R}^*/dt = \mathbf{P}^*/m^*$ for $\mathbf{P}^* = \Sigma_{i=1}^{N} \mathbf{p}_i \Theta_i$. Similarly, $d\mathbf{P}^*/dt = -\mathcal{L}\mathbf{P}^* = \Sigma_{i=1}^{N} \mathbf{F}_i \Theta_i$ (the net force on the virus).

The virus contains many atoms and hence has a large size relative to that of an atom; thus we take it to have a diameter of $O(\varepsilon^{-1})$ for small parameter ε. Since we are interested in significant migration distances (i.e., greater than or equal to the viral diameter), we scale $\mathbf{R}^*$ to be $O(\varepsilon^{-1})$. The scaling $\varepsilon^2 = m/m^*$ for typical atomic mass m is adopted under the assumption that the virus is empty, i.e., we develop a theory for the dynamics of a viral capsid. Other cases may easily be considered as well [e.g., m^* is $O(\varepsilon^{-3})$ for a complete virus with genome in the capsid's cavity]. Under the assumption that the system is near equilibrium, a typical viral c.m. kinetic energy $P^{*2}/2m^*$ is $O(k_B T)$. Thus $\mathbf{P}^*$ scales as $\sqrt{m^*}$, i.e., is $O(\varepsilon^{-1})$. With this we adopt the scaled variables $\mathbf{R}$ and $\mathbf{P}$ such that

$$\mathbf{R}^* = \varepsilon^{-1}\mathbf{R}, \quad \mathbf{P}^* = \varepsilon^{-1}\mathbf{P}. \tag{2.3}$$

Let $\mathbf{f} = \varepsilon^{-1} \Sigma_{i=1}^{N} \mathbf{F}_i \Theta_i$ be the scaled net force on the virus; this scaling emerges from the near-equilibrium assumption and that the viral diameter is $O(\varepsilon^{-1})$ and hence its surface area is $O(\varepsilon^{-2})$. For short range interactions with the host medium, the number of contributions to the net force on the virus is proportional to its surface area; however, near equilibrium there is much cancellation of these forces so that the residual force over and above its zero average is small and hence scaled to be on the order of 1 over the square root of the number of individual atomic contributions, i.e., $O(\varepsilon)$. This is one example of the corollary to criterion (B), i.e., the net force is sufficiently small that the momentum evolves slowly. In contrast, if the host medium is experiencing a shock wave, the number of host atoms on one side of the virus would be much larger than on the other and hence $\mathbf{P}$ would change rapidly as the shock is passing. In summary, under the above assumptions Newton's equations imply

$$\frac{d\mathbf{R}}{dt} = \varepsilon^2 \frac{\mathbf{P}}{m}, \quad \frac{d\mathbf{P}}{dt} = \varepsilon^2 \mathbf{f}, \tag{2.4}$$

proving that $\mathbf{R}$ and $\mathbf{P}$ are slowly varying.

To characterize intraviral structural dynamics, we introduce relative coordinates $\mathbf{s}_i$ for viral atoms such that

$$\mathbf{r}_i = \mathbf{s}_i + \Theta_i \mathbf{R}^*. \tag{2.5}$$

Next an order parameter Φ is introduced to describe viral dilatation. In particular, we take Φ to be a measure of the viral size relative to that of a reference structure, e.g., an x-ray cryostructure. To this end, consider the definition

$$\Phi = \frac{\sum_{i=1}^{N} m_i \mathbf{s}_i \cdot \vec{\mathbf{X}} \mathbf{s}_i^0 \Theta_i}{\sum_{i=1}^{N} m_i s_i^{02} \Theta_i}, \tag{2.6}$$

where $\vec{\mathbf{X}}$ is a length-preserving rotation matrix (see below) and $s_i^0 = |\mathbf{s}_i^0|$. We now demonstrate that this order parameter and the associated momentum Π are slow variables. This Φ, Π order parameter description will serve as the basis for a simple viral dilation phase transition theory.

The equation of motion of Φ is

$$\frac{d\Phi}{dt} = -\mathcal{L}\Phi \frac{\varepsilon^4}{m'} \sum_{i=1}^{N} \left(\frac{\mathbf{p}_i}{m_i} - \frac{\mathbf{P}^*}{m^*} \right) \cdot \vec{\mathbf{X}} \mathbf{s}^0 m_i \Theta_i,$$
$$m' \equiv \varepsilon^4 \sum_{i=1}^{N} m_i s_i^{02} \Theta_i. \tag{2.7}$$

The scaling of m' is consistent with our earlier assumption that the number of atoms in the virus is $O(\varepsilon^{-2})$ and its diameter is $O(\varepsilon^{-1})$. Introducing relative velocities $\boldsymbol{\pi}_i/m_i = (\mathbf{P}_i/m_i) - (\mathbf{P}^*/m^*)$ for viral atoms yields

$$\frac{d\Phi}{dt} = -\mathcal{L}\Phi = \varepsilon^2 \frac{\Pi}{m'}, \quad \Pi = \varepsilon^2 \Pi^*, \quad \Pi^* \equiv \sum_{i=1}^{N} \boldsymbol{\pi}_i \cdot \vec{\mathbf{X}} \mathbf{s}_i^0 \Theta_i. \tag{2.8}$$

The scaling of Π is based on the assumption that while there are $O(\varepsilon^{-2})$ atoms in the virus, the contributions to Π^* are of fluctuating sign, but each term has an $\mathbf{s}_i^0$ factor which, like the viral diameter, is $O(\varepsilon^{-1})$; thus Π^* is $O(\varepsilon^{-2})$. With this

$$\frac{d\Pi}{dt} = -\mathcal{L}\Pi = \varepsilon^2 \sum_{i=1}^{N} \mathbf{F}_i \cdot \left(\vec{\mathbf{X}} \mathbf{s}_i^0 - \sum_{j=1}^{N} \frac{m_j \vec{\mathbf{X}} \mathbf{s}_j^0}{m^*} \Theta_j \right) \Theta_i. \tag{2.9}$$

The j sum in (2.9) is over a large number of vector contributions which tend to cancel; as the m^* factor in this term is proportional to the number of atoms in the virus, the j sum is small relative to $\vec{\mathbf{X}} s_i^0$. Thus to good approximation we rewrite (2.9) as

$$\frac{d\Pi}{dt} = \varepsilon^2 g, \quad g = \sum_{i=0}^{N} \mathbf{F}_i \cdot \vec{\mathbf{X}} \mathbf{s}_i^0 \Theta_i \tag{2.10}$$

for "dilatation force" g that is the analog of the c.m. force $\mathbf{f}$. The scaling of g as implied in (2.10) differs from that of $\mathbf{f}$ as each term in g has an additional $\mathbf{s}_i^0$ factor [which is $O(\varepsilon^{-1})$]. With this we conclude that Φ and Π are also slow variables.

Viral rotation can also be shown to be slow due to the large moment of inertia. First relate the rotation matrix $\vec{\mathbf{X}}$ to the relative atomic configuration $\{\mathbf{s}_i, \Theta_i = 1\}$. If the intraviral state only changes due to overall rotation and isotropic dilatation, then the α_2 component of $\mathbf{s}_i$ is given by

$$s_{i\alpha_2} = \Phi \sum_{\alpha_1=1}^{3} X_{\alpha_2\alpha_1} s_{i\alpha_1}^0. \tag{2.11}$$

Multiplying by $s_{i\alpha_3}^0 \Theta_i$ and summing over i yield

$$Y_{\alpha_3\alpha_2} = \Phi \sum_{\alpha_1=1}^{3} X_{\alpha_2\alpha_1} Y_{\alpha_3\alpha_1}^0, \tag{2.12}$$

$$Y_{\alpha_3\alpha_2} = \sum_{i=1}^{N} s_{i\alpha_3}^0 s_{i\alpha_2} \Theta_i, \quad Y_{\alpha_3\alpha_1}^0 = \sum_{i=1}^{N} s_{i\alpha_3}^0 s_{i\alpha_1}^0 \Theta_i. \tag{2.13}$$

Taking d/dt of both sides of (2.12) and recalling that $\mathbf{s}_i^0$ is a reference configuration (and not a dynamical variable) yield

$$\sum_{i=1}^{N} \frac{s_{i\alpha_3}^0 \pi_{i\alpha_2} \Theta_i}{m_i} = \Phi \sum_{\alpha_1=1}^{3} \frac{dX_{\alpha_2\alpha_1}}{dt} Y_{\alpha_3\alpha_1}^0 + \sum_{\alpha_1=1}^{3} X_{\alpha_2\alpha_1} Y_{\alpha_3\alpha_1}^0 \frac{d\Phi}{dt}. \tag{2.14}$$

Since the virus diameter is $O(\varepsilon^{-1})$, then so are $\mathbf{s}_i$ and $\mathbf{s}_i^0$ for viral atoms. This implies that $Y_{\alpha_3\alpha_1}^0$ is $O(\varepsilon^{-4})$ as it is a sum of $O(\varepsilon^{-2})$ positive terms (at least for the diagonal elements). Consider the trace of the left hand side of (2.14), i.e., the sum of $\mathbf{s}_i^0 \cdot \boldsymbol{\pi}_i / m_i$ terms. Using an argument as for Π^*, it is expressed to be $O(\varepsilon^{-2})$. With this, and that $d\Phi/dt$ is $O(\varepsilon^2)$, $d\vec{\mathbf{X}}/dt$ is seen to be $O(\varepsilon^2)$ so that rotation is slow.

Since icosahedral viruses (e.g., CCMV, NωV, and poliovirus) are nearly spherical (see Fig. 1), this symmetry and the conjecture that centrifugal forces do not have significant effect on STs, we expect that effects of overall rotation (i.e., $\vec{\mathbf{X}}$ and $\vec{\boldsymbol{\Omega}}$) should decouple from those of STs. Hence the reduced distribution for the slow variables $W(\mathbf{P}, \mathbf{R}, \Pi, \Phi, t)$ can be factorized into $\vec{\mathbf{X}}$, $\vec{\boldsymbol{\Omega}}$, $\mathbf{P}$, $\mathbf{R}$, Π, and Φ dependent parts. With this, the N-atom probability density ρ is taken to have the dependence $\rho(\Gamma; \mathbf{P}, \mathbf{R}, \Pi, \Phi, t_0, t)$ with $\Gamma \equiv \{\mathbf{p}_1, \mathbf{r}_1, \ldots, \mathbf{p}_N, \mathbf{r}_N\}$, while $\vec{\mathbf{X}}$ and $\vec{\boldsymbol{\Omega}}$ are ignored henceforth.

Introduction of the slow variables in ρ does not imply an increase in the number of degrees of freedom beyond $6N$,[35] rather this ansatz indicates a special dependence of ρ on Γ through the slow variables that enable the multiscale analysis. With the slow variable dependences of the N-atom potential and kinetic energy as well as the lowest order probability density analyzed in Appendix A, we obtain a FP equation for $W(\mathbf{P}, \mathbf{R}, \Pi, \Phi, t)$ from the multiscale formulation of the N-atom Liouville equation in Appendix B by following the approach outlined earlier:[35]

$$\frac{\partial W}{\partial \tau} = -\left[\frac{\mathbf{P}}{m} \cdot \frac{\partial}{\partial \mathbf{R}} + \mathbf{f}^{\text{th}} \cdot \frac{\partial}{\partial \mathbf{P}} + \frac{\Pi}{m'} \frac{\partial}{\partial \Phi} + g^{\text{th}} \frac{\partial}{\partial \Pi} \right] W$$
$$+ \varepsilon^2 \left[\vec{\boldsymbol{\gamma}}_{ff} \frac{\partial}{\partial \mathbf{P}} \left(\beta \frac{\mathbf{P}}{m} + \frac{\partial}{\partial \mathbf{P}} \right) + \boldsymbol{\gamma}_{fg} \cdot \frac{\partial}{\partial \mathbf{P}} \left(\beta \frac{\Pi}{m'} + \frac{\partial}{\partial \Pi} \right) \right.$$
$$\left. + \boldsymbol{\gamma}_{gf} \cdot \frac{\partial}{\partial \Pi} \left(\beta \frac{\mathbf{P}}{m} + \frac{\partial}{\partial \mathbf{P}} \right) + \gamma_{gg} \frac{\partial}{\partial \Pi} \left(\beta \frac{\Pi}{m'} + \frac{\partial}{\partial \Pi} \right) \right] W, \tag{2.15}$$

where $\mathbf{f}^{\text{th}}$ and g^{th} are thermal average forces: $\vec{\gamma}_{ff}$, $\boldsymbol{\gamma}_{fg}$, $\boldsymbol{\gamma}_{gf}$, and γ_{gg} are friction tensors given by (B18); and $\tau=\varepsilon^2 t$.

In this FP equation the c.m. and ST order parameters are coupled in three ways. The thermally averaged forces $\mathbf{f}^{\text{th}}$ and g^{th} depend on $\mathbf{R}$ and Φ (except for a virus in an otherwise uniform medium wherein there is only Φ dependence). Coupling is also provided through the $\mathbf{R}$, Φ dependence of the friction coefficients. The cross-friction coefficients $\boldsymbol{\gamma}_{fg}$ and $\boldsymbol{\gamma}_{gf}$ provide additional coupling. However, for a virus in an isotropic medium, $\boldsymbol{\gamma}_{fg}$ and $\boldsymbol{\gamma}_{gf}$ vanish because the thermal average of any vector [e.g., $(\mathbf{f}(0)g(t))^{\text{th}}$] is zero for an isotropic system.

To deduce some of the implications of (2.15) for viral STs, consider an isolated virus in an isotropic medium and that the viral c.m. momentum is at equilibrium. In this case, $W=\tilde{W}\exp(-\beta P^2/2m)/Q_{\text{c.m.}}$, where $Q_{\text{c.m.}}$ ensures normalization for the c.m. part of W. In this case, the reduced order parameter distribution $\tilde{W}(\Pi,\Phi,\tau)$ satisfies

$$\frac{\partial\tilde{W}}{\partial\tau}=-\left[\frac{\Pi}{m'}\frac{\partial}{\partial\Phi}+g^{\text{th}}\frac{\partial}{\partial\Pi}\right]\tilde{W}+\varepsilon^2\gamma_{gg}\frac{\partial}{\partial\Pi}\left(\beta\frac{\Pi}{m'}+\frac{\partial}{\partial\Pi}\right)\tilde{W}, \tag{2.16}$$

as can be verified upon substituting the product solution for W in (2.15). Let $\ln Q=-\beta F$ for Φ-dependent free energy F. Then $g^{\text{th}}=-\partial F/\partial\Phi$, from which we see that (2.16) has the equilibrium solution that is proportional to $\exp[-\beta(\Pi^2/2m'+F)]$. Note that Π ranges over all positive and negative values, while Φ is strictly positive and $F\to\infty$ as $\Phi\to 0$ (as $\Phi=0$ represents a virus wherein all atoms are at the c.m.).

III. FREE ENERGY AND FRICTION COEFFICIENTS

To implement the multiscale analysis for viral STs, one must construct the free energy and friction coefficients for the range of slow variable values of interest. In this section, an approach for using constrained evolution to generate the nanocanonical ensemble for computing viral free energy and friction coefficients is presented and the computational feasibility of the approach is discussed. Hysteresis and irreversibility in viral STs from several experimental studies are reviewed. A double-well free energy profile is deduced for the ST in NωV capsid.

A. Constrained evolution for nanocanonical ensemble

Configurations and corresponding energies for fixed values of slow variables can be generated using the following noninertial dynamics approach. The approach can also be generalized to generate the inertial dynamics needed to construct the friction coefficients. The ensemble of detailed atomic configurations for free energy computations is generated by solving the pseudodynamics equations

$$\frac{dr_{i\alpha}}{dt}=\sum_{i'\alpha'}B_{i\alpha i'\alpha'}\left[-\frac{\partial V}{\partial r_{i'\alpha'}}+A_{i'\alpha'}\right], \tag{3.1}$$

where $r_{i\alpha}$ is the αth position coordinate of atom i, V is the N-atom potential, and $A_{i'\alpha'}$ is a random force. The B matrix is introduced to ensure that the $\mathbf{r}_i$ only explore configurations consistent with given values of the slow variables. Consider a set of M slow variables $\underline{\Phi}\equiv\{\Phi_k,k=1,\ldots,M\}$. From each slow variable we construct the $3N$ length column vectors $\partial\Phi_k/\partial r_{i\alpha}$, $i=1,\ldots,N$; $\alpha=1,2,3$. From these we form a set of M orthonormalized row $\langle k|$ and column $|k\rangle$ vectors and thereby the projection matrix $|k\rangle\langle k|$. Then, in matrix notation, $\underline{\underline{B}}=\underline{\underline{I}}-|1\rangle\langle 1|-\ldots-|M\rangle\langle M|$. With this and (3.1), one finds that $d\Phi_k/dt=0$, $k=1,\ldots,M$. The ensemble generation can be stabilized by replacing V with $U(\underline{\Phi};\mathbf{r}_1,\ldots,\mathbf{r}_N)$ defined in Sec. II. Thus (3.1) drives the system to a state with $\underline{\Phi}$-dependent minimum and there is no gradual drift of the $\underline{\Phi}$ due to numerical round-off. The ensembles so generated can be enriched by using multiple initial datasets for the $r_{i\alpha}$ to avoid being trapped in a local minimum of U. Space warping can be used to generate these configurations as it facilitates transitions between configurations with significantly different properties. The utility of this approach in minimization has already been demonstrated.[45] The ensemble for each choice of slow variables can be used as a basis of a Monte Carlo integration algorithm to compute the thermal average forces needed for the FP equation.

B. Computational feasibility

As reviewed in Sec. I, an all-atom molecular dynamics simulation on a complete satellite tobacco mosaic virus has been achieved on an SGI 1024-processor Altix system by using 256 processors and 128 gigabytes memory with a performance of 1.1 ns/day (7.85×10^{-2} s/time step).[36] We believe that it should take about 10 000 such cycles with intermittent major moves using our space-warping technique to obtain a free energy variation with 20 wisely chosen slow variable values. This would take 4.36 h on a computer platform with similar performance to the Altix system noted above. A more complete temporal dynamics will be used in a similar way to construct thermal average forces and friction coefficients. These computations will benefit from published methods.[41,46–52]

C. Hysteresis, irreversibility, and the free energy profile in viral STs

Hysteresis of capsid assembly/disassembly in hepatitis B virus (HBV) has been observed and studied with a chemical kinetic model.[53] CCMV capsid undergoes a pH and metal ion dependent reversible swelling transition between close native and open swollen forms.[1–5] This expansion is reversible, yet it has not been determined whether there is hysteresis in the process. However, hysteretic effects are found in the maturation of NωV.[8] NωV undergoes large conformational changes from a procapsid form (480 Å in diameter) to a compact capsid form (410 Å in diameter) when pH is decreased from 7.6 to 5.0. The transition takes less than 100 ms and is accompanied by a slow autoproteolyis (taking hours) corresponding to the cleavage of 70 kDa coat proteins to 62 and 8 kDa proteins. The conformational rearrangement is initially reversible until about 15% of the cleavage events are completed, at which point the particles are locked into the capsid conformation, regardless of pH.[6,7,9] A further study on a cleavage-defective mutant (N570T) of NωV showed that

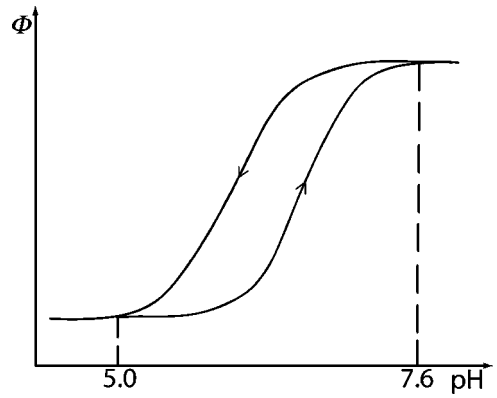

FIG. 3. The cleavage-defective mutant of NωV undergoes hysteretic ST when pH in the host medium changes. The curve indicates values of Φ for which the virus/host system's free energy is a minimum.

the transition from procapsid to capsid in the mutant is reversible, and that the reverse process is much slower, with some capsids never reexpanded after 4 days of dialysis against pH 7.6 buffer, and the reexpanded procapsids display slightly different properties than the original capsid, suggesting hysteresis but not complete reversibility in the transition.[8] Irreversible poliovirus conformational changes are found in receptor-mediated cell entry, during which its coat protein VP4 and the N terminus of VP1 are externalized. Two putative cell entry intermediates (135S and 80S particles) are formed and they are both about 4% larger than the native virions (160S particles).[12–16] The viral receptor behaves as a classic TS theory catalyst, facilitating the ST from native virions to 135S intermediate particles by lowering the activation energy for the process by 50 kcal/mol.[15]

Hysteresis in the NωV ST implies that there are two structurally distinct states of the virus for a given range of host pH. Such states correspond to free energy wells, suggesting a free energy landscape with double-well character. Viruses make transitions between fluctuating states defined by deep, local free energy wells. The residence time within a given well is determined by the intensity of thermal fluctuations, the mass and size of the virus, the number of states within the well, and the height of free energy barriers that must be crossed in existing the well. These conclusions are shown in Figs. 3 and 4 to follow naturally from our formulation, as discussed in Sec. IV.

As suggested in Fig. 3, when pH changes in the host medium, an order parameter Φ (i.e., the relative overall size of the virus as defined in Sec. II) of the cleavage-defective mutant of NωV can make a hysteric transition between two states with distinct structures, i.e., residing within different free energy wells. Underlying the transition is a double-well free energy profile as suggested in Fig. 4. In more complex systems, e.g., for HK97 bacteriophage,[10,11] there are multiple wells in a multidimensional order parameter space and transitions between them. Even for systems supporting a single transition, there are likely several key order parameters so that a true picture of the ST must follow from an analysis of landscapes in higher dimensional order parameter space. A single order parameter model supporting two states with distinct structures implied by different ranges of Φ is suggested in Fig. 4 along with the associated probability distribution. In Sec. IV, we will explore TS theory to estimate the free energy barrier of the transition between the two structures.

IV. TS THEORY AND AN APPLICATION TO NωV CAPSID ST

In this section, TS theory for obtaining approximate solutions to the FP equation (2.16) in the inertial limit and estimating the free energy barrier of viral STs with their free energy profile sketched in Sec. III is presented as below. The TS ansatz is designed for the case where the barrier height is much greater than k_BT. Thus the system resides for an appreciable time in a given free energy well and one may thus define a viral state as the subset of detailed slow variable configurations residing within the well. TS theory can be used to compute the evolving statistics of the subset of viruses whose state resides within a given well.

As suggested in Fig. 4, the value of Φ in the TS is Φ_c, the location of the barrier separating the left $(-)$ and right $(+)$ wells. To start the analysis, define the total probabilities $\tilde{W}_\pm$ of occupation of the $(-)$ and $(+)$ wells via

$$\tilde{W}_-(\tau) = \int_0^{\Phi_c} d\Phi \int d\Pi \tilde{W}(\Pi,\Phi,\tau),$$
$$\tilde{W}_+(\tau) = \int_{\Phi_c}^{\infty} d\Phi \int d\Pi \tilde{W}(\Pi,\Phi,\tau). \tag{4.1}$$

Integrating (2.16) over Φ from Φ_c to ∞ (and similarly, from 0 to Φ_c) and over all Π yields

$$\frac{d\tilde{W}_\pm}{d\tau} = \pm \int d\Pi \frac{\Pi}{m'} \tilde{W}(\Pi,\Phi_c,\tau). \tag{4.2}$$

In obtaining (4.2) we used the fact that because $F \to \infty$ as $\Phi \to 0$ or ∞, the probability of finding a virus in such extreme

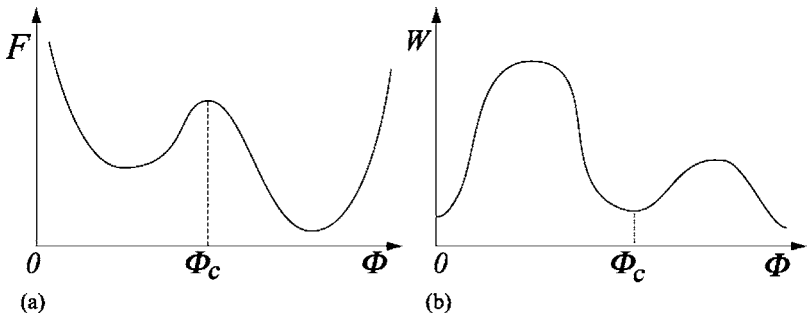

FIG. 4. (a) A double-well free energy profile F vs Φ implied by the hysteresis in NωV capsid ST and (b) the associated probability distribution W vs Φ.

configurations is zero, i.e., $\tilde{W}(\Pi,0,\tau)=\tilde{W}(\Pi,\infty,\tau)=0$.

According to TS theory, in the right well $\tilde{W}$ is essentially an equilibrium distribution and hence is proportional to $\exp[-\beta(\Pi^2/2m'+F)]$; therefore the $\tilde{W}$-weighted average of Π is zero in the well, showing that there is no contribution to $d\tilde{W}_+/d\tau$ except from a small region of Φ near Φ_c. The TS ansatz for $\tilde{W}(\Pi,\Phi,\tau)$ is developed as follows. When the barrier is much higher than k_BT, the two domains (Φ to the left or right of Φ_c) act as if they were independent; hence $\tilde{W}$ is essentially its equilibrium value for the left and right wells independently. Only infrequently is the barrier surmounted. Thus $\tilde{W}$ is proportional to $\exp[-\beta(\Pi^2/2m'+F)]/Z_-$ for $\Phi<\Phi_c$ and $\exp[-\beta(\Pi^2/2m'+F)]/Z_+$ for $\Phi>\Phi_c$, where

$$Z_- = \int_0^{\Phi_c} d\Phi \int d\Pi e^{-\beta(\Pi^2/2m'+F)},$$
$$Z_+ = \int_{\Phi_c}^{\infty} d\Phi \int d\Pi e^{-\beta(\Pi^2/2m'+F)}. \tag{4.3}$$

With this and the definition of $\tilde{W}_\pm$ in (4.1), the proportionality constants are determined and we obtain

$$\tilde{W} = \begin{cases} \exp[-\beta(\Pi^2/2m'+F)]\dfrac{\tilde{W}_-}{Z_-}, & \Phi<\Phi_c \\ \exp[-\beta(\Pi^2/2m'+F)]\dfrac{\tilde{W}_+}{Z_+}, & \Phi>\Phi_c. \end{cases} \tag{4.4}$$

However, near Φ_c infrequent barrier crossing takes place. The TS theory suggests that at Φ_c the right-going viruses are predominantly coming from the left near-equilibrium well, and vice versa. Thus

$$\tilde{W}(\Pi,\Phi_c,\tau) = e^{-\beta(\Pi^2/2m'+F_c)}\left\{\frac{\tilde{W}_-}{Z_-}\theta(\Pi)+\frac{\tilde{W}_+}{Z_+}\theta(-\Pi)\right\}, \tag{4.5}$$

$$\theta(x) = \begin{cases} 0, & x<0 \\ 1, & x>0, \end{cases} \tag{4.6}$$

and F_c is F evaluated at Φ_c. With this ansatz, (4.2) becomes

$$\frac{d\tilde{W}_+}{d\tau} = \frac{1}{2}\omega\left[e^{-\beta\Delta F_-^c}\frac{\tilde{W}_-}{Y_-} - e^{-\beta\Delta F_+^c}\frac{\tilde{W}_+}{Y_+}\right], \tag{4.7}$$

where

$$\omega = \left(\frac{2}{\pi\beta m'}\right)^{1/2}, \tag{4.8}$$

$$Y_- = \int_0^{\Phi_c} d\Phi e^{-\beta\Delta F_-}, \quad Y_+ = \int_{\Phi_c}^{\infty} d\Phi e^{-\beta\Delta F_+}, \tag{4.9}$$

$$\Delta F_\pm = F - F_\pm, \tag{4.10}$$

where F_- and F_+ are F evaluated at the minimum of the left and right wells, respectively. The superscript c on $\Delta F_\pm^c$ indicates evaluation at Φ_c. The quantity ω is the thermal average of $|\Pi/m'|$ and hence has dimensions of s^{-1}. As $\tilde{W}_+ + \tilde{W}_- = 1$, a second equation for $\tilde{W}_-$ is not needed. Recalling that $\tau=\varepsilon^2 t$, one should add a factor $\varepsilon^2 = m/m^*$ to the right hand side of (4.7) to get $d\tilde{W}_+/dt (=\varepsilon^2 d\tilde{W}_+/d\tau)$.

Time-resolved small-angle x-ray scattering (TR-SAXS) experiments show that the capsid ST during NωV maturation occurs on time scales in the range of 0.1–100 s with the detection of a fast-forming intermediate in the transition.[7] To evaluate this in our formulation via TS theory, consider the purely forward contribution to (4.7) as recast in the form

$$\frac{d\tilde{W}_+}{dt} = \frac{1}{2Y_-}\varepsilon^2\omega\exp\left[-\frac{\Delta F_-^c}{k_BT}\right](1-\tilde{W}_+). \tag{4.11}$$

Several factors are seen from (4.11) to limit the rate of a viral ST: (1) a factor 1/2 as only viruses undergoing purely forward transitions at the TS contribute, (2) the ε^2 factor expressing the great size of a virus relative to a single atom (i.e., $\varepsilon^2 \approx 10^{-6}$), (3) $\exp[-\Delta F_-^c/k_BT]$ expressing a free energy barrier effect, and (4) Y_- introducing entropic effects that account for the number of Φ configurations collectively labeled the $(-)$ state in the left free energy well.

The rate law (4.11) is in the form $d\tilde{W}_+/dt = k(1-\tilde{W}_+)$. From the aforementioned experimental data, the rate coefficient k varies from 10 to 0.01 s^{-1}. To use these data in estimating ΔF_-^c, we adopt the following assumptions. Y_- is the range of Φ in the $(-)$ well for which $\Delta F_-^c > k_BT$. As a crude estimate, we use the fact that a change in the radius upon transition is about 16% for NωV capsid and then take Y_- to be smaller than this, in particular, 1% of the capsid radius, i.e., $Y_- \approx 0.01$. Using the carbon atom mass as typical for the average atom in the capsid, a capsid radius of 198 Å and $\varepsilon^2 = m/m^* \approx 10^{-6}$, m' is given by (2.7), i.e.,

$$\begin{aligned} m' &\equiv \varepsilon^4 \sum_{l=1}^{N} m_i s_i^{0.2}\Theta_i \approx \varepsilon^2 \frac{12.01\times10^{-3}\ \text{kg mol}^{-1}}{6.022\times10^{23}\ \text{mol}^{-1}} \\ &\quad\times(198\times10^{-10}\ \text{m})^2 \\ &= 7.73\times10^{-48}\ \text{kg m}^2. \end{aligned} \tag{4.12}$$

This yields

$$\begin{aligned} \omega &= \left(\frac{2}{\pi\beta m'}\right)^{1/2} \\ &\approx \left(\frac{2\times1.38\times10^{-23}\ \text{m}^2\ \text{kg s}^{-2}\ \text{K}^{-1}\times298.15\ \text{K}}{\pi\times7.73\times10^{-48}\ \text{kg m}^2}\right)^{1/2} \\ &= 1.84\times10^{13}\ \text{s}^{-1}. \end{aligned} \tag{4.13}$$

Using this and (4.11) to evaluate ΔF_-^c in kcal/mol, we have

$$\begin{aligned} \Delta F_-^c &= -\frac{N}{4186\ \text{J kcal}^{-1}} k_BT \ln\frac{2kY_-}{\varepsilon^2\omega} \\ &\approx \begin{cases} 10.86\ \text{kcal/mol}, & k=10\ \text{s}^{-1} \\ 14.95\ \text{kcal/mol}, & k=0.01\ \text{s}^{-1}. \end{cases} \end{aligned} \tag{4.14}$$

The transition of native poliovirus (160S particle) to an infectious intermediate (135S particle) during cell entry is determined to have a free energy barrier of 30 kcal/mol without the cell receptor.[15,54] Our calculations based on crude

estimates show that the NωV capsid ST has a lower free energy barrier, and it depends on how we change the *p*H to induce the ST. When *p*H is lowered from 7.5 to 5.0, it occurs on 0.1 s time scale with a free energy barrier of 10.86 kcal/mol; with a less drastic *p*H change (*p*H lowered to 5.8 or 5.5 instead of 5.0), the ST occurs more slowly on a 100 s time scale with a higher free energy barrier of 14.95 kcal/mol. The activation enthalpy for the poliovirus ST is known to be lowered by 50 kcal/mol in the presence of a cell receptor, and capsid-binding drugs for poliovirus are shown to inhibit the receptor-mediated ST through a combination of enthalpic and entropic effects. We suggest that viral STs must be understood in terms of free energy barriers and the entropics of viral states, as well as inertial effects [manifested in the ε^2 factor in (4.11)].

V. CONCLUSIONS

The all-atom multiscale approach enables one to analyze viral STs from first principles and a calibrated interatomic force field. For a virus in a homogeneous isotropic medium, STs can be modeled via a FP equation for a set of order parameters characterizing the intraviral state. For more complex host media (i.e., with a cell membrane or a solid surface), there is a coupling between the order parameters and the c.m. variables manifest in both the free energy and the cross-friction coefficients ($\boldsymbol{\gamma}_{fg}$ and $\boldsymbol{\gamma}_{gf}$ for the single order parameter model of Sec. II).

Observations on STs in viral capsids and the multiscale analysis of the *N*-atom problem suggest that there are three distinct aspects to the dynamics of viral STs. Inertial effects emerge from the large contrast in the mass of a whole virus, or one of its structural units, and that of an atom. This is one way in which the time scale of viral migration or overall structural change is much longer than that of individual atomic vibrations and collisions. However, there is also the effect of energy barrier crossing from one free energy well to another, which implies that a virus resides in a given well for an extended time, only occasionally crossing the barrier separating wells where the virus experiences an unlikely thermal fluctuation. Finally, entropics can also inhibit the rate of exiting a free energy well, i.e., a virus spends an extended period of time exploring many detailed configurations before making a ST. For example, this is why capsid assembly is much slower than disassembly. These three effects are accounted for in the TS theory presented in Sec. IV.

The TS estimate of the free energy barrier of NωV capsid ST provides an upper bound, i.e., the single order parameter model likely misses lower energy barrier pathways for a viral capsid transition. However, it can also be argued that an overall dilatational fluctuation could provide more room for capsid structural units to translate and rotate relative to their neighbor units and thereby initiate a ST. In either case, the single order parameter model of Sec. II can be used as a base line to interpret experimental results in terms of free energy barrier and entropic effects that can then be compared with estimates from more complete models.

While we estimate the free energy barrier of the NωV capsid ST roughly, for a more predictive and accurate approach, the free energy and friction coefficients must be computed from a calibrated interatomic force field and the FP equation solved to realize the full implications of the all-atom multiscale approach for virology and medical sciences or biotechnology. In future studies, a set of slow variables characterizing viral STs will be explored and a constrained nanocanonical ensemble method[35] will be implemented for computing the free energy and friction coefficients. Monte Carlo and related methods[41,46–52] for carrying out such computations will be adopted for viral STs as they were shown in Sec. III to be feasible. The notion of space warping[45] was introduced to treat the self-assembly, deformation, and rotation of many-atom systems and will greatly accelerate the computation of free energy and friction coefficients; it can also be integrated into our all-atom multiscale analysis to construct multiple order parameter models that account for lower free energy barrier pathways for viral STs missed by single order parameter models.[43] We suggest that the numerical implementation of our approach to viral phenomena will create a parameter-free, general simulator for viruses and other nanostructures.

ACKNOWLEDGMENTS

We thank Professor J. E. Johnson (Department of Molecular Biology, The Scripps Research Institute) for valuable discussions on viral capsid STs and Dr. R. Sawafta (QuarTek International) for suggestions on the biotechnical importance of this work. We also thank Z. Shreif for her efforts in verifying the equations. We appreciate the support of the U.S. Department of Energy, Office of Science, Indiana University's College of Arts and Sciences for general activities at the Center for Cell and Virus Theory, and the Office of the Vice President for Research.

APPENDIX A: SLOW VARIABLE DEPENDENCES OF THE *N*-ATOM ENERGY AND THE LOWEST ORDER PROBABILITY DENSITY

To complete the multiscale analysis, the *N*-atom potential V and kinetic energy K must be expressed in terms of $\mathbf{R}$, Φ, $\vec{\mathbf{X}}$, the residual dependence on $\mathbf{r}_1,\ldots,\mathbf{r}_N$, and the associated momenta. To accomplish this for V we write

$$\mathbf{r}_i = \boldsymbol{\sigma}_i + \Theta_i(\varepsilon^{-1}\mathbf{R} + \Phi\vec{\mathbf{X}}\mathbf{s}_i^0). \tag{A1}$$

The $\mathbf{R}^*$ term accounts for migration of the viral c.m., the $\Phi\vec{\mathbf{X}}$ contribution generates an atomic displacement due to rotation and dilation, and $\boldsymbol{\sigma}_i$ is the residual (incoherent) displacement over and above the coherent effects of $\mathbf{R}$, Φ, and $\vec{\mathbf{X}}$. With this we introduce the explicit potential function U via

$$U(\mathbf{R},\Phi,\vec{\mathbf{X}};\mathbf{r}_1,\ldots,\mathbf{r}_N) = V(\{\boldsymbol{\sigma}_i + \Theta_i(\varepsilon^{-1}\mathbf{R} + \Phi\vec{\mathbf{X}}\mathbf{s}_i^0);\ i=1,\ldots,N. \tag{A2}$$

The gradient of V with respect to Φ is given by the derivative of U with respect to Φ keeping $\mathbf{R}$, $\vec{\mathbf{X}}$, and the residual $\mathbf{r}_i$ dependence fixed; this shows the utility in introducing U, i.e., the residual dependence on the $\mathbf{r}_i$ from the $\boldsymbol{\sigma}_i$ is explicit.

With this the scaled net force on the virus $\mathbf{f}$ and dilation force g are given by

$$\mathbf{f} = \varepsilon^{-1}\sum_{i=1}^{N} \mathbf{F}_i\Theta_i = -\left(\frac{\partial U}{\partial \mathbf{R}}\right)_{\Phi,\vec{\mathbf{X}},\mathbf{r}_1,\ldots,\mathbf{r}_N}, \tag{A3}$$

$$g = -\left(\frac{\partial U}{\partial \Phi}\right)_{\mathbf{R},\vec{\mathbf{X}},\mathbf{r}_1,\ldots,\mathbf{r}_N} = \sum_{i=1}^{N} \mathbf{F}_i \cdot \vec{\mathbf{X}}\mathbf{s}_i^0\Theta_i. \tag{A4}$$

The kinetic energy K of the virus is given by

$$K = \sum_{i=1}^{N} \frac{p_i^2}{2m_i}\Theta_i. \tag{A5}$$

Letting $\mathbf{p}_i^{\mathrm{inc}}/m_i = -\mathcal{L}\boldsymbol{\sigma}_i$ and $\mathbf{p}_i^{\mathrm{coh}}/m_i = -\mathcal{L}(\mathbf{r}_i - \boldsymbol{\sigma}_i)$, we obtain

$$K = \sum_{i=1}^{N} \frac{1}{2m_i}[(p_i^{\mathrm{inc}})^2 + 2\mathbf{p}_i^{\mathrm{inc}} \cdot \mathbf{p}_i^{\mathrm{coh}} + (p_i^{\mathrm{coh}})^2]\Theta_i. \tag{A6}$$

Consider the coh/inc cross term for a small cluster (say 100) of the million atoms in the virus. The $\mathbf{p}_i^{\mathrm{coh}}$ in this cluster are in the same direction, but the $\mathbf{p}_i^{\mathrm{inc}}$ will have varying directions. Even if there is a small residual contribution from the cross term of one cluster, it is likely that there is another to cancel it. With this we take the cross term to be negligible.

Letting $\varepsilon^2\vec{\mathbf{\Omega}} = -\mathcal{L}\vec{\mathbf{X}}$, we find that

$$\mathbf{p}_i^{\mathrm{coh}} = m_i\Theta_i\left[\varepsilon\frac{\mathbf{P}}{m} + \varepsilon^2\frac{\Pi}{m'}\vec{\mathbf{X}}\mathbf{s}_i^0 + \varepsilon^2\Phi\vec{\Omega}\mathbf{s}_i^0\right]. \tag{A7}$$

Note that the Π contribution corresponds to motion along the $\vec{\mathbf{X}}\mathbf{s}_i^0$ direction, while the $\vec{\mathbf{\Omega}}$ contribution is a rotation about the c.m.; thus the Π and $\vec{\mathbf{\Omega}}$ contributions are orthogonal. With this, and recalling that $|\vec{\mathbf{X}}\mathbf{s}_i^0| = s_i^0$, we obtain

$$K^{\mathrm{coh}} = \sum_{i=1}^{N} \frac{(p_i^{\mathrm{coh}})^2}{2m_i}\Theta_i = \frac{1}{2}\sum_{i=1}^{N} m_i\Theta_i\left[\varepsilon^2\frac{P^2}{m^2} + \varepsilon^4\frac{\Pi^2}{m^2}s_i^{02} + \varepsilon^4\Phi^2|\vec{\mathbf{\Omega}}\mathbf{s}_i^0|^2\right]. \tag{A8}$$

Gross terms of $\mathbf{P}$ with Π or $\vec{\mathbf{\Omega}}$ are dropped as the latter contributions are in many directions so that the i sum makes them negligible. Note that K^{coh} is then independent of the residual variations in Γ (i.e., those over and above that in $\mathbf{P}$, Π, and $\vec{\mathbf{\Omega}}$) so that the partition function Q (see below) can be written as a configuration part (i.e., one dependent on $\mathbf{R}$, ϕ, and $\vec{\mathbf{\Omega}}$) times $\exp(-\beta K^{\mathrm{coh}})$. The incoherent part of K can be written as

$$K^{\mathrm{inc}} = \sum_{i=1}^{N} \frac{(p_i^{\mathrm{inc}})^2}{2m_i}\Theta_i = \sum_{i=1}^{N} \frac{\Theta_i}{2m_i}|\mathbf{p}_i - \mathbf{p}_i^{\mathrm{coh}}|^2. \tag{A9}$$

In the lowest order solution to the scaled Liouville equation as analyzed in (B5) in Appendix B, we have the following factor:

$$\hat{\rho} = \frac{e^{-\beta(K^{\mathrm{inc}}+V)}}{Q^{\mathrm{inc}}(\beta,\mathbf{R},\Phi,\vec{\mathbf{X}})}, \tag{A10}$$

$$Q^{\mathrm{inc}} = \int d^{6N}\Gamma\Delta\,\exp[-\beta(K^{\mathrm{inc}} + V)], \tag{A11}$$

where Δ is a product of delta functions of $\mathbf{R}$, Φ, and $\vec{\mathbf{\Omega}}$ centered around their Γ-dependent values. As $\mathbf{p}_i^{\mathrm{inc}} = \mathbf{p}_i - \mathbf{p}_i^{\mathrm{coh}}$ and $\mathbf{p}_i^{\mathrm{coh}}$ is $O(\varepsilon)$, $\partial\hat{\rho}/\partial\mathbf{P}$, $\partial\hat{\rho}/\partial\Pi$, and $\partial\hat{\rho}/\partial\vec{\mathbf{\Omega}}$ are zero.

APPENDIX B: DERIVATION OF THE FP EQUATION

With the N-atom probability density ρ taken to have the dependence $\rho(\Gamma;\mathbf{P},\mathbf{R},\Pi,\Phi,t_0,t)$ with $\Gamma \equiv \{\mathbf{p}_1,\mathbf{r}_1,\ldots,\mathbf{p}_N,\mathbf{r}_N\}$, we introduce a set of scaled times $t_n = \varepsilon^{2n}t, n = 0,1,\ldots$. With this and the chain rule, the Liouville equation takes the form

$$\sum_{n=0}^{\infty} \varepsilon^{2n}\frac{\partial\rho}{\partial t_n} = (\mathcal{L}_0 + \varepsilon^2\mathcal{L}_1)\rho, \tag{B1}$$

$$\mathcal{L}_0 = -\sum_{i=1}^{N}\left[\frac{\mathbf{p}_i}{m_i}\cdot\frac{\partial}{\partial\mathbf{r}_i} + \mathbf{F}_i\cdot\frac{\partial}{\partial\mathbf{p}_i}\right], \tag{B2}$$

$$\mathcal{L}_1 = -\frac{\mathbf{P}}{m}\cdot\frac{\partial}{\partial\mathbf{R}} - \mathbf{f}\cdot\frac{\partial}{\partial\mathbf{P}} - \frac{\Pi}{m'}\frac{\partial}{\partial\Phi} - g\frac{\partial}{\partial\Pi}. \tag{B3}$$

The operators $\mathcal{L}_0$ and $\mathcal{L}_1$ arise from the chain rule in the course of accounting for the direct and indirect dependences of ρ on Γ; therefore these operators only have meaning when acting on ρ in its present form as a function of both Γ and the slow variables.

At first sight one might suggest choosing the relative variables $\boldsymbol{\pi}_i, \mathbf{S}_i$ of Sec. II for the viral atoms to express $\mathcal{L}_0$ and $\mathcal{L}_1$ and develop the theory, rather than the $\mathbf{p}_i, \mathbf{r}_i$ defined in the laboratory frame as used here. However, $\Sigma_{i=1}^{N}\boldsymbol{\pi}_i\Theta_i = \mathbf{0}$ and $\Sigma_{i=1}^{N} m_i\mathbf{S}_i\Theta_i = \mathbf{0}$. This implies that the $\{\boldsymbol{\pi}_i, \mathbf{S}_i; \Theta_i = 1\}$ do not constitute a set of independent variables. In contrast, the $\mathbf{p}_i, \mathbf{r}_i$ are independent. While the slow variables depend on Γ, the partial derivatives of ρ with respect to the $\mathbf{p}_i, \mathbf{r}_i$ at constant slow variables in $\mathcal{L}_0$, and conversely for derivatives in $\mathcal{L}_1$, only reflect the use of the chain rule to account for the direct and indirect dependences of ρ on Γ. It does not constitute a validation of the number of degrees of freedom. In contrast, ignoring the fact that the $\boldsymbol{\pi}_i$ and $\mathbf{S}_i$ are not independent variables would introduce a violation of the number of degrees of freedom.

Development of the theory proceeds by constructing a perturbative solution for ρ, i.e.,

$$\rho = \sum_{n=0}^{\infty} \rho_n\varepsilon^{2n}. \tag{B4}$$

To lowest order, with the assumption that ρ_0 is independent of the microscopic time t_0, we arrive at $\mathcal{L}_0\rho_0 = 0$; this implies that the biological phenomena of interest have quasiequilibrium character (e.g., we are not interested in processes on the 10^{-12} s time scale). The nanocanonical quasiequilibrium solution of this equation is found to be[35]

$$\rho_0 = \frac{e^{-\beta H_0}}{Q(\beta,\mathbf{R},\Phi)} W(\mathbf{P},\mathbf{R},\Pi,\Phi,\underline{t}) \equiv \hat{\rho} W. \tag{B5}$$

This illustrates the multiple dependences on Γ, i.e., direct and indirect through H_0 and indirectly through the slow variables appearing in Q and W. The reduced distribution W can be shown to be the average of the microscopic density $\delta(\mathbf{p}-\mathbf{P})\delta(\mathbf{r}-\mathbf{R})\delta(\pi-\Pi)\delta(\phi-\Phi)$ for the lowest order N-atom distribution ρ_0. The variables $\mathbf{p}$, $\mathbf{r}$, π, and ϕ are particular values of the dynamical variables $\mathbf{P}$, $\mathbf{R}$, Π, and Φ but are not dynamical variables (i.e., Γ-dependent quantities themselves). In contrast,

$$\mathbf{P} \equiv \varepsilon \sum_{l=1}^{N} \mathbf{p}_i \Theta_i, \quad \mathbf{R} \equiv \varepsilon \sum_{l=1}^{N} \frac{m_l \mathbf{r}_l}{m^*} \Theta_i,$$

$$\Pi \equiv \varepsilon^2 \sum_{i=1}^{N} \boldsymbol{\pi}_i \cdot \vec{\mathbf{X}} \mathbf{s}_i^0 \Theta_i, \quad \Phi \equiv \frac{\sum_{i=1}^{N} m_i \mathbf{s}_i \cdot \vec{\mathbf{X}} \mathbf{s}_i^0 \Theta_i}{\sum_{i=1}^{N} m_i s_i^{02} \Theta_i}. \tag{B6}$$

The $\mathbf{r},\phi$-dependent partition function Q is given by

$$Q(\beta,\mathbf{r},\phi) = \int d^{6N}\Gamma \delta(\mathbf{r}-\mathbf{R})\delta(\phi-\Phi) e^{-\beta H_0}, \tag{B7}$$

$$H_0 = \sum_{i=1}^{N} \frac{(p_i^{\mathrm{inc}})^2}{2m_i} \Theta_i + U(\mathbf{R},\Phi;\mathbf{r}_1,\ldots,\mathbf{r}_N), \tag{B8}$$

where $\mathbf{p}_i^{\mathrm{inc}}$ is defined in terms of the $\mathbf{p}_i$ and the slow variables of (B6) in (A9).

To $O(\varepsilon^2)$ the Liouville equation implies

$$\left(\frac{\partial}{\partial t} - \mathcal{L}_0\right)\rho_1 = -\frac{\partial \rho_0}{\partial t_1} + \mathcal{L}_1 \rho_0, \tag{B9}$$

where we drop the 0 on t_0 for simplicity henceforth. Assuming that initially ρ is near equilibrium, ρ_1 can be taken to be zero at $t=0$. With this, the $O(\varepsilon^2)$ equation admits the solution

$$\rho_1 = \int_0^t dt' e^{\mathcal{L}_0(t-t')}\left[-\hat{\rho}\frac{\partial W}{\partial t_1} + \mathcal{L}_1 \hat{\rho} W\right]. \tag{B10}$$

Applying $\mathcal{L}_1$ and noting that $\mathcal{L}_0\hat{\rho}=0$ yield

$$\rho_1 = -t\hat{\rho}\frac{\partial W}{\partial t_1} - \int_0^t dt' e^{\mathcal{L}_0(t-t')}\left[\frac{\mathbf{P}}{m}\cdot\frac{\partial}{\partial \mathbf{R}} + \mathbf{f}\cdot\frac{\partial}{\partial \mathbf{P}} + \frac{\Pi}{m'}\frac{\partial}{\partial \Phi} + g\frac{\partial}{\partial \Pi}\right]\hat{\rho} W. \tag{B11}$$

The statistical mechanical postulate "the long-time and ensemble averages for equilibrium systems are equal" implies

$$\lim_{t\to\infty} \frac{1}{t}\int_{-t}^{0} dt'' e^{-\mathcal{L}_0 t''} A = \int d^{6N}\Gamma A \hat{\rho} \equiv A^{\mathrm{th}} \tag{B12}$$

for any dynamical variable $A(\Gamma)$. Changing variables via $t''=t'-t$ in (B11), it is found that removal of the secular behavior in ρ_1 at large t implies

$$\frac{\partial W}{\partial t_1} = -\left[\frac{\mathbf{P}}{m}\cdot\frac{\partial}{\partial \mathbf{R}} + \mathbf{f}^{\mathrm{th}}\cdot\frac{\partial}{\partial \mathbf{P}} + \frac{\Pi}{m'}\frac{\partial}{\partial \Phi} + g^{\mathrm{th}}\frac{\partial}{\partial \Pi}\right] W, \tag{B13}$$

where $\mathbf{f}^{\mathrm{th}}$ and g^{th} are the $\hat{\rho}$-weighted thermal average quantities and the equivalence of long time and thermal averages have been assumed. With this

$$\rho_1 = -\hat{\rho}\int_0^t dt' e^{\mathcal{L}_0(t-t')}\left\{(\mathbf{f}-\mathbf{f}^{\mathrm{th}})\cdot\left(\beta\frac{\mathbf{P}}{m} + \frac{\partial}{\partial \mathbf{P}}\right) + (g-g^{\mathrm{th}})\left(\beta\frac{\Pi}{m'} + \frac{\partial}{\partial \Pi}\right)\right\} W. \tag{B14}$$

Note that the use of the long-time/ensemble average equivalence avoids the traditional use of the integration of the Liouville equation over Γ, which both erases ambiguities (i.e., do not include the Γ dependence in the slow variables) and ensures that all secular behavior in ρ is removed, not just that in a reduced distribution. This completes the $O(\varepsilon^2)$ analysis.

To $O(\varepsilon^4)$ one finds

$$\left(\frac{\partial}{\partial t} - \mathcal{L}_0\right)\rho_2 = -\frac{\partial \rho_0}{\partial t_2} - \frac{\partial \rho_1}{\partial t_1} + \mathcal{L}_1 \rho_1. \tag{B15}$$

Again with $\rho_2=0$ at $t=0$, we find

$$\rho_2 = \int_0^t dt' e^{\mathcal{L}_0(t-t')}\left(-\hat{\rho}\frac{\partial W}{\partial t_2} - \frac{\partial \rho_1}{\partial t_1} + \mathcal{L}_1 \rho_1\right). \tag{B16}$$

Removal of the secular behavior in ρ_2 at large t implies

$$\frac{\partial W}{\partial t_2} = \left[\vec{\boldsymbol{\gamma}}_{ff}\frac{\partial}{\partial \mathbf{P}}\left(\beta\frac{\mathbf{P}}{m} + \frac{\partial}{\partial \mathbf{P}}\right) + \boldsymbol{\gamma}_{fg}\cdot\frac{\partial}{\partial \mathbf{P}}\left(\beta\frac{\Pi}{m'} + \frac{\partial}{\partial \Pi}\right) + \boldsymbol{\gamma}_{gf}\cdot\frac{\partial}{\partial \Pi}\left(\beta\frac{\mathbf{P}}{m} + \frac{\partial}{\partial \mathbf{P}}\right) + \gamma_{gg}\frac{\partial}{\partial \Pi}\left(\beta\frac{\Pi}{m'} + \frac{\partial}{\partial \Pi}\right)\right] W. \tag{B17}$$

The friction tensors are given by

$$\gamma_{ff,kl} \equiv \int_{-\infty}^{0} dt[(f_k(0)f_l(t))^{\mathrm{th}} - f_k^{\mathrm{th}} f_l^{\mathrm{th}}],$$

$$\gamma_{fg,k} \equiv \int_{-\infty}^{0} dt[(f_k(0)g(t))^{\mathrm{th}} - f_k^{\mathrm{th}} g^{\mathrm{th}}],$$

$$\gamma_{gf,k} \equiv \int_{-\infty}^{0} dt[(g(0)f_k(t))^{\mathrm{th}} - g^{\mathrm{th}} f_k^{\mathrm{th}}],$$

$$\gamma_{gg} \equiv \int_{-\infty}^{0} dt[(g(0)g(t))^{\mathrm{th}} - (g^{\mathrm{th}})^2],$$

$$k,l = 1,2,3. \tag{B18}$$

Thus the friction coefficients are related to force correlation functions as expected.

Letting $\tau=t_1$ we recompose the above results[43] to obtain a FP equation for the reduced distribution of viral slow variables $W(\mathbf{P},\mathbf{R},\Pi,\Phi,t)$ as

$$\frac{\partial W}{\partial \tau} = -\left[\frac{\mathbf{P}}{m}\cdot\frac{\partial}{\partial \mathbf{R}} + \mathbf{f}^{\mathrm{th}}\cdot\frac{\partial}{\partial \mathbf{P}} + \frac{\Pi}{m'}\frac{\partial}{\partial \Phi} + g^{\mathrm{th}}\frac{\partial}{\partial \Pi}\right]W + \varepsilon^2\left[\vec{\boldsymbol{\gamma}}_{ff}\frac{\partial}{\partial \mathbf{P}}\left(\beta\frac{\mathbf{P}}{m} + \frac{\partial}{\partial \mathbf{P}}\right) + \boldsymbol{\gamma}_{fg}\cdot\frac{\partial}{\partial \mathbf{P}}\left(\beta\frac{\Pi}{m'} + \frac{\partial}{\partial \Pi}\right) + \boldsymbol{\gamma}_{gf}\cdot\frac{\partial}{\partial \Pi}\left(\beta\frac{\mathbf{P}}{m} + \frac{\partial}{\partial \mathbf{P}}\right) + \gamma_{gg}\frac{\partial}{\partial \Pi}\left(\beta\frac{\Pi}{m'} + \frac{\partial}{\partial \Pi}\right)\right]W. \tag{B19}$$

[1] J. A. Speir, S. Munshi, G. Wang, T. S. Baker, and J. E. Johnson, Structure (London) **3**, 63 (1995).
[2] J. E. Johnson and J. A. Speir, J. Mol. Biol. **269**, 665 (1997).
[3] H. Liu, C. Qu, J. E. Johnson, and D. A. Case, J. Struct. Biol. **142**, 356 (2003).
[4] L. O. Liepold, J. Revis, M. Allen, L. Oltrogge, M. Young, and T. Douglas, Phys. Biol. **2**, S166 (2005).
[5] J. A. Speir, B. Bothner, C. Qu, D. A. Willits, M. J. Young, and J. E. Johnson, J. Virol. **80**, 3582 (2006).
[6] M. A. Canady, M. Tihova, T. N. Hanzlik, J. E. Johnson, and M. Yeager, J. Mol. Biol. **299**, 573 (2000).
[7] M. A. Canady, H. Tsuruta, and J. E. Johnson, J. Mol. Biol. **311**, 803 (2001).
[8] D. J. Taylor, N. K. Krishna, M. A. Canady, A. Schneemann, and J. E. Johnson, J. Virol. **76**, 9972 (2002).
[9] D. J. Taylor, Q. Wang, B. Bothner, P. Natarajan, M. G. Finn, and J. E. Johnson, Chem. Commun. (Cambridge) **2003**, 2770.
[10] K. K. Lee, H. Tsuruta, R. W. Hendrix, R. L. Duda, and J. E. Johnson, J. Mol. Biol. **352**, 723 (2005).
[11] W. R. Wikoff, J. F. Conway, J. Tang, K. K. Lee, L. Gan, N. Cheng, R. L. Duda, R. W. Hendrix, A. C. Steven, and J. E. Johnson, J. Struct. Biol. **153**, 300 (2006).
[12] C. E. Fricks and J. M. Hogle, J. Virol. **64**, 1934 (1990).
[13] D. M. Belnap, D. J. Filman, B. L. Trus *et al.*, J. Virol. **74**, 1342 (2000).
[14] D. M. Belnap, B. M. McDermott, D. J. Filman, N. Cheng, B. L. Trus, H. J. Zuccola, V. R. Racaniello, J. M. Hogle, and A. C. Steven, Proc. Natl. Acad. Sci. U.S.A. **97**, 73 (2000).
[15] S. K. Tsang, B. M. Mcdermott, V. R. Racaniello, and J. M. Hogle, J. Virol. **75**, 4984 (2001).
[16] J. M. Hogle, Annu. Rev. Microbiol. **56**, 677 (2002).
[17] D. Harries, S. May, W. M. Gelbart, and A. Ben-Shaul, Biophys. J. **75**, 159 (1998).
[18] B. Speelman, B. R. Brooks, and C. B. Post, Biophys. J. **80**, 121 (2001).
[19] J. Durup, J. Phys. Chem. **95**, 1817 (1991).
[20] M. Tuckerman and B. J. Berne, J. Chem. Phys. **95**, 8362 (1991).
[21] M. Tuckerman, B. J. Berne, and G. J. Martyna, J. Chem. Phys. **94**, 6811 (1991).
[22] A. Askar, B. Space, and H. Rabitz, J. Phys. Chem. **99**, 7330 (1995).
[23] S. Reich, Physica D **89**, 28 (1995).
[24] D. K. Phelps and C. B. Post, J. Mol. Biol. **254**, 544 (1995).
[25] D. K. Phelps, P. J. Rossky, and C. B. Post, J. Mol. Biol. **276**, 331 (1995).
[26] J. Elezgaray and Y. H. Sanejouand, Biopolymers **46**, 493 (1998).
[27] J. Elezgaray and Y. H. Sanejouand, J. Comput. Chem. **21**, 1274 (2000).
[28] K. A. Feenstra, B. Hess, and H. J. C. Berendsen, J. Comput. Chem. **20**, 786 (1999).
[29] M. R. Sorensen and A. F. Voter, J. Chem. Phys. **112**, 9599 (2000).
[30] H. M. Chun, C. E. Padilla, D. N. Chin *et al.*, J. Comput. Chem. **21**, 159 (2000).
[31] F. Tama and C. L. Brooks III, J. Mol. Biol. **318**, 733 (2002).
[32] F. Tama and C. L. Brooks III, J. Mol. Biol. **345**, 299 (2005).
[33] H. W. T. van Vlijmen and M. Karplus, J. Chem. Phys. **115**, 691 (2001).
[34] H. W. T. van Vlijmen and M. Karplus, J. Mol. Biol. **350**, 528 (2005).
[35] Y. Miao and P. Ortoleva, J. Chem. Phys. **125**, 044901 (2006).
[36] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten, Structure (London) **14**, 437 (2006).
[37] S. Chandrasekhar, Astrophys. J. **97**, 255 (1943).
[38] J. M. Deutch and I. Oppenheim, Faraday Discuss. Chem. Soc. **83**, 1 (1987).
[39] J. E. Shea and I. Oppenheim, J. Phys. Chem. **100**, 19035 (1996).
[40] M. H. Peters, J. Chem. Phys. **110**, 528 (1998).
[41] M. H. Peters, J. Stat. Phys. **94**, 557 (1999).
[42] J. E. Shea and I. Oppenheim, Physica A **247**, 417 (1997).
[43] P. Ortoleva, J. Phys. Chem. **109**, 21258 (2005).
[44] S. Bose and P. Ortoleva, J. Chem. Phys. **70**, 3041 (1979).
[45] K. Jaqaman and P. Ortoleva, J. Comput. Chem. **23**, 484 (2002).
[46] B. Dünweg and K. Kremer, J. Chem. Phys. **99**, 6983 (1993).
[47] A. Korpf, B. Dünweg, and W. Paul, J. Chem. Phys. **107**, 6945 (1997).
[48] P. Ahrichs and B. Dünweg, J. Chem. Phys. **111**, 8225 (1999).
[49] B. Liu and B. Dünweg, J. Chem. Phys. **118**, 8061 (2003).
[50] T. Ohmoria and Y. Kimura, J. Chem. Phys. **119**, 7328 (2002).
[51] A. Uvarov and S. Fritzsche, Phys. Rev. E **73**, 011111 (2006).
[52] A. Uvarov and S. Fritzsche, Macromol. Theory Simul. **13**, 241 (2004).
[53] S. Singh and A. Zlotnick, J. Biol. Chem. **278**, 18249 (2003).
[54] S. K. Tsang, P. Danthi, M. Chow, and J. M. Hogle, J. Mol. Biol. **296**, 335 (2000).