J Stat Phys (2008) 130: 669–685
DOI 10.1007/s10955-007-9452-4

Curvilinear All-Atom Multiscale (CAM) Theory of Macromolecular Dynamics

Z. Shreif · P. Ortoleva

Received: 13 December 2006 / Accepted: 2 October 2007 / Published online: 10 November 2007

Abstract A method is introduced for simulating long timescale macromolecular structural fluctuations and transitions with atomic-scale detail. The N-atom Liouville equation for the macromolecule/host medium system provides the starting point for the analysis. Order parameters characterizing overall macromolecular architecture are demonstrated to be slowly evolving.

For single-stranded macromolecules, a curvilinear coordinate provides a way to introduce the order parameters. Using a multiscale approach, Fokker–Planck equations are derived. A nanocanonical method for constructing the lowest order solution to the Liouville equation and the equivalence of long-time and ensemble averages avoid the tedious bookkeeping needed to preserve the number of degrees of freedom (required in earlier methods). The method overcomes the large energy barriers that plague other approaches for estimating rates of transition between macromolecular conformations. A reduced dynamics is derived for the friction dominated limit. New experimental methods for observing macromolecular dynamics and medical sciences applications are discussed.

Keywords Macromolecular conformations · Multiscale analysis · Protein folding · RNA folding · DNA folding · Fokker–Planck equations

1 Background

Predicting macromolecular structural dynamics over long times requires a multiscale approach. Small molecules or ions interact with a macromolecule through atomic-scale processes. The complexing of two or more macromolecules occurs through atomic-scale configurations. While nanoscale architecture and dynamics are of central interest in many phenomena, it is atomic-scale effects which cumulatively give rise to them.

Z. Shreif · P. Ortoleva (✉)
Center for Cell and Virus Theory, Department of Chemistry, Indiana University, 800 E. Kirkwood Ave., Bloomington, IN 47405, USA
e-mail: ortoleva@indiana.edu

Macromolecular states are related to free-energy landscapes mediated by atomic-scale fluctuation entropic contributions. To use a calibrated interatomic force field as a basis for a parameter free macromolecular theory, an all-atom formulation is required. Yet, to fully understand the nanoscale structure and dynamics of a macromolecule, one must develop methods to simulate hundreds of thousands of atoms over the long timescales that typify biomacromolecular phenomena. Macromolecular complexes like viruses or other bionanostructures, and their interaction with drugs or cell surface receptors, involve even more atoms. In summary, macromolecular systems must be understood via an integration of effects spanning many scales in space and time. Computational molecular dynamics (MD) is discussed extensively in the literature [1–12] and methods were introduced to solve the problem of multiple time scales encountered in these simulations. Most of these methods [13–16] like TJMTS [13] use the separation of the equations of motion into fast and slow parts, running many small time steps for the rapidly changing terms keeping the slowly varying ones constant before updating the latter using a large time step. RESPA [14–16] uses the same idea while correcting for the errors encountered when approximating the equation of motion by adjusting the time step dynamically. MD is a powerful approach for smaller-size, shorter-time phenomena; however, it is presently impractical for simulating large macromolecules and other nanosystems over millisecond times or longer.

Another difficulty is the need to account for solvent effects. In biological systems, the solvent influence the structure and dynamics of the macromolecule; thus, omitting it from the simulation will not lead to realistic results. Models based on a continuum representation of the solvent have been developed [17] and proved to be a less expensive alternative to full atomistic models. These, however, face difficulties in quantitatively describing a number of essential effects such as the role of hydrogen-bonding and hydrophobicity. Other problems are encountered due to the fact that solvent effects change over different size scales (see Ref. [17] for a review).

The Smoluchowski equation has been applied to polymer dynamics [18]. This work yields insights into long timescale fluctuating polymer dynamics. However, it does not take full advantage of the multiple timescale character of these systems and therefore, does not integrate an interatomic force field. It is not based on the reduction of the Liouville equation. In addition, Kirkwood's generalized diffusion equation assumes slowly evolving variables. However, in the cited work above, it was used as a starting point to the method with fast variables such as the atomic positions (see however Sect. 4 below).

Advances in the modeling of many-atom systems hold great promise for addressing the above challenges. An all-atom multiscale analysis of the Liouville equation provides a method to derive generalized Fokker–Planck (FP) equations in cases where slow variables can be constructed explicitly [19, 20].

In previous multiscale analysis methods [21–25], FP equations were also derived starting from the Liouville equation. However, these only focused on particles devoid of atomic-scale structure in order to avoid difficulties encountered in preserving the total number of degrees of freedom. AMA (all-atom multiscale analysis) developed earlier [19] solve this problem by removing the secular behavior from the full N-atom probability density while introducing the order parameters; a direct analysis of the time dependence of the perturbation expansion was used instead of integration of fast variables. The nanocanonical ensemble was developed through entropy maximization constrained by the statistical average position of the center of mass of the nanoparticle. AMA was applied to the study of viral structural transitions [20] where the nanocanonical ensemble introduced was defined by the viral center of mass and dilatation.

The objective of the present work is to modify AMA to arrive at a theory of macromolecular dynamics. AMA provides a coarse-grained model, but does not neglect the internal

structure of the nanosystem as is essential for the macromolecular conformational studies discussed above. In summary, AMA captures both atomic and nanoscale behaviors simultaneously.

The systems of interest are sufficiently complex that the FP equations cannot always be written directly. Thus, we derive them using a modified AMA approach. AMA uses a seven-step algorithm.

Step 1: The system is described in terms of N classical atoms interacting via bonded and non-bonded forces.

Step 2: A set of order parameters, e.g. center of mass, orientation or overall deformation of each nanoscale component, is set forth; Newton's equations and statistical arguments are used to show that these variables evolve on timescales long relative to that of atomic vibrations or collisions.

Step 3: The N-atom kinetic and potential energies are expressed in terms of the slow variables $\underline{\Phi}$ and residual dependence on the set of atomic momenta and positions Γ.

Step 4: The solution of the Liouville equation (i.e. the probability density ρ for all atomic positions and momenta) is reformulated to make its dependence on Γ (both directly and through the slow variables $\underline{\Phi}$) explicit. Importantly, $\underline{\Phi}$ is not an additional set of dynamical variables. Through this reformulation, one avoids the tedious algebra to ensure there are only $6N$ degrees of freedom. Rather, our $\rho(\Gamma, \underline{\Phi}, t)$ formulation expresses the two distinct dependencies that capture the multiscale character of a nanosystem.

Step 5: A perturbation parameter ε is identified (e.g. the ratio of the mass of a typical atom to that of a nanoparticle); statistical arguments about the sum of many terms of fluctuating sign and the assumption that the system is close to momentum equilibrium are adapted; with this, the Liouville equation becomes

$$\sum_{n=0}^{\infty} \varepsilon^n \left(\frac{\partial}{\partial t_n} - \mathcal{L}_n \right) \rho = 0, \tag{1.1}$$

where $t_n = \varepsilon^n t$ and $\mathcal{L}_n$ is the contribution to the Liouville operator that is $O(\varepsilon^n)$; $\mathcal{L}_n$ emerges naturally due to the way in which the slow variables, the length and mass ratios, and other physical factors appear in the Liouville equation.

Step 6: An expansion of ρ in powers of ε is introduced and the Liouville equation is solved order-by-order.

Step 7: The solution to various orders in ε is examined and, by asserting that the n-th order solution ρ_n is well behaved for large time t_0, the generalized FP equation is obtained; we do not ensure these conditions by integrating out the fast variables (e.g. the direct dependence of ρ on Γ); this approach of earlier studies leads to technical difficulties when one wishes to use an all-atom description of the nanoparticle. Rather, we use the statistical mechanical theorem "the long-time average is equivalent to the ensemble average".

Advantages and details of this seven-step procedure are explained elsewhere [19, 20].

AMA, as outlined above, facilitates the derivation of FP equations for nanometer scale structures. Here we implement AMA by introducing curvilinear coordinates that capture the overall structure of a macromolecule. A new automated method for constructing the order parameters is presented which eliminates the need to develop special software to account for the detailed atomic structure of the monomer units. These order parameters capture overall features of single-stranded macromolecules and are used to define the new nanocanonical ensemble as explained in Appendix 1. Innovations also include introducing two different

scales of the order parameters in the development of the ensemble, the impact of which is discussed.

Full FP and reduced (friction-dominated, Smoluchowski-type) equations that we derive here provide a method for practical simulations through their implied Langevin equations. Though atomic details are preserved, simulation time will tremendously decrease since the number of order parameters is much less than that of atoms in the system. The thermally-averaged forces and friction coefficients in these equations can be estimated using molecular dynamics [20]. The derivation presented here implies the structure of the cross-friction coefficients matrix and gives statistical mechanical formulas for these coefficients and the thermal average forces.

In what follows, we develop the theory and computational algorithm that make these advances in macromolecular simulation for large systems evolving over biologically relevant timescales feasible. Formal developments are provided in Sects. 2 to 4 and conclusions are drawn in Sect. 5.

2 Curvilinear Coordinates, Order Parameters, and Scaling

The conformation of single-stranded macromolecules is described here in terms of order parameters. The latter are shown to evolve slowly in time relative to the 10^{-12} second timescale of atomic collision and vibration, an essential property of an order parameter. The potential energy of the composite macromolecule/host system is then expressed in terms of these order parameters and residual, single-atom coordinate dependencies. The overall structure of a single-stranded macromolecule is described by a curve $\vec{r}$ that depends on an arc length-like parameter τ. This $\vec{r}/\tau$ relationship is parameterized by factors that ultimately turn out to be viable order parameters capturing overall macromolecular structure. In addition, we retain all-atom resolution by keeping certain residual atomistic variables.

We define τ such that $\tau = 1$ corresponds to one end (i.e. $\vec{r}$ at $\tau = 1$ is close to the center-of-mass (CM) of the first monomer), while L is the value of τ near the CM of the terminal monomer of the L-monomer macromolecule. A set of $k_{\max}$ vector order parameters $\underline{\vec{\Phi}}^* = \{\vec{\Phi}_1^*, \ldots, \vec{\Phi}_{k_{\max}}^*\}$ is introduced such that the curve characterizing overall macromolecular conformation takes the form

$$\vec{r}(\tau, \underline{\vec{\Phi}}^*) = \sum_{k=1}^{k_{\max}} \vec{\Phi}_k^* u_k(\tau). \tag{2.1}$$

We take τ to increase monotonically from 1 to L as $\vec{r}(\tau, \underline{\vec{\Phi}}^*)$ progresses along the strand. In particular, τ_ℓ is designed such that $\vec{r}(\tau_\ell, \underline{\vec{\Phi}}^*)$ is the point on the curve lying closest to the CM of the ℓ-th monomer. If the u_k are polynomials of low order, then $\vec{r}(\tau, \underline{\vec{\Phi}}^*)$ bends smoothly as τ varies from 1 to L; if the u_k are oscillatory functions of τ, then $\vec{r}(\tau, \underline{\vec{\Phi}}^*)$ has coiled or globular structure. Thus, proper choice of the u_k can capture much of the character of a macromolecule with only a few terms, i.e. $k_{\max} \ll N$ (see Fig. 1 for example). However, the choice of the order parameters $\vec{\Phi}_k^*$ as coefficients of specific functions u_k may be too restrictive in some problems; thus the relationship between $\vec{r}(\tau, \underline{\vec{\Phi}}^*)$ and the order parameters can be nonlinear to advantage (e.g. an order parameter could be the frequency of a periodic function). Such nonlinear relationships can be included in the present approach, but will not be considered here.

If the basis functions $u_k(\tau)$ vary on the scale of the macromolecule, i.e. do not have short-scale variations in the interval $1 < \tau < L$, then we expect the $\vec{\Phi}_k^*$ to evolve slowly

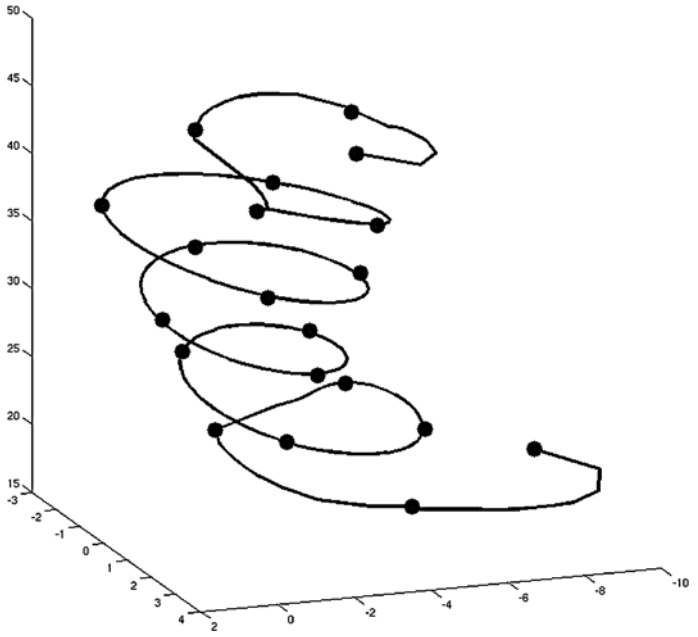

Fig. 1 $\vec{r}(\tau, \underline{\vec{\Phi}}^*)$ for Alamethicin with $k_{\max} = 19$. The black dots are the position of the CM of each monomer. See more details about the plot in Appendix 2

in time since they simultaneously involve many atoms, as is demonstrated in detail below. However, macromolecules like proteins can have sharp bends. In that case, we suggest that the u_k can be chosen as piecewise smooth functions that match at the "joints". This mixed smooth/joint structure will be addressed elsewhere.

According to AMA [19, 20], order parameters must be expressed in terms of the nanostructure's atomic configuration. Here, this is accomplished through a mass-weighted least-squares approach, i.e. by minimizing the deviation D defined via

$$D = \frac{1}{2} \sum_{\ell=1}^{L} \frac{m_\ell^*}{m^{**}} |\vec{r}(\tau_\ell, \underline{\vec{\Phi}}^*) - \vec{R}_\ell^*|^2, \tag{2.2}$$

where $\vec{R}_\ell^*$ and m_ℓ^* are the CM and mass of monomer ℓ, respectively; and m^{**} is the mass of the whole macromolecule.

The $\vec{\Phi}_k^*$ are chosen to minimize D; setting $\partial D / \partial \vec{\Phi}_q^* = \vec{0}$ yields

$$\sum_{\ell=1}^{L} m_\ell^* \vec{R}_\ell^* u_q(\tau_\ell) = \sum_{k=1}^{k_{\max}} \sum_{\ell=1}^{L} m_\ell^* \vec{\Phi}_k^* u_k(\tau_\ell) u_q(\tau_\ell), \tag{2.3}$$

constituting $k_{\max}$ linear equations from which the $\vec{\Phi}_k^*$ are to be obtained.

As the $\vec{R}_\ell^*$ depend on the coordinates of the atoms in monomer ℓ, the solution of (2.3) has the form $\vec{\Phi}_k^* = \vec{\Phi}_k'$ where the $\vec{\Phi}_k'$ depend on the locations of all the atoms in the macro-

molecule. It is convenient to introduce an inner product via

$$(u_q, u_k) \equiv \sum_{\ell=1}^{L} \frac{m_\ell^*}{m^{**}} u_q(\tau_\ell) u_k(\tau_\ell). \tag{2.4}$$

Choosing the basis functions to be orthonormal, i.e. $(u_k, u_q) = \delta_{kq}$, yields

$$\vec{\Phi}_k^* = (\vec{R}^*, u_k) \equiv \sum_{\ell=1}^{L} \vec{R}_\ell^* u_k(\tau_\ell) \frac{m_\ell^*}{m^{**}} \equiv \vec{\Phi}_k'. \tag{2.5}$$

With this, $\underline{\vec{\Phi}}^*$ is now shown to serve as the set of order parameters on which a multiscale analysis can be based.

When the basis functions are properly selected, the order parameters $\vec{\Phi}_k^*$ are slowly varying as follows. Newton's equations imply

$$\frac{d\vec{\Phi}_k^*}{dt} = -\mathcal{L}\vec{\Phi}_k^*, \tag{2.6}$$

where $\mathcal{L}$ is the Liouville operator for the N-atom system (macromolecule plus host medium). This implies

$$\frac{d\vec{\Phi}_k^*}{dt} = \sum_{i=1}^{N} \frac{\vec{p}_i}{m_i} \left(\frac{\partial \vec{R}_{\ell(i)}^*}{\partial \vec{r}_i} \right) \frac{m_{\ell(i)}^*}{m^{**}} u_k(\tau_{\ell(i)}) \Theta_i \tag{2.7}$$

where $\ell(i)$ indicates the monomer in which atom i resides; m_i and $\vec{p}_i$ are the mass and momentum of atom i; and Θ_i is zero except when atom i is in the macromolecule, in which case it is one. By definition,

$$\vec{R}_\ell^* = \sum_{i=1}^{N} \frac{m_i}{m_\ell^*} \vec{r}_i \Theta_i^\ell, \tag{2.8}$$

where Θ_i^ℓ is zero except when atom i is in the ℓ-th monomer, in which case it is one.

With this

$$\frac{d\vec{\Phi}_k^*}{dt} = \frac{1}{m^{**}} \sum_{\ell=1}^{L} \vec{P}_\ell^* u_k(\tau_\ell) \equiv \frac{\vec{\Pi}_k^*}{m^{**}}, \tag{2.9}$$

$$\vec{P}_\ell^* = \sum_{i=1}^{N} \vec{p}_i \Theta_i^\ell. \tag{2.10}$$

This completes the introduction of the order parameters $\underline{\vec{\Phi}}^*$ and associated momenta $\underline{\vec{\Pi}}^*$.

Let $\varepsilon^2 = m/m^{**}$ where m is the mass of a typical atom in the macromolecule. Thus, $\varepsilon^{-2} = N_m$ where N_m is the number of atoms in the macromolecule. Assume that the macromolecule's range of migration and the monomers' mass are $\mathrm{O}(\varepsilon^{-1})$. Since there are L monomers, the above imply that L and $\vec{R}_\ell^*$ are $\mathrm{O}(\varepsilon^{-1})$ also. $\vec{\Phi}_k^*$ is a sum of $\mathrm{O}(\varepsilon^{-1})$ terms, each of $\mathrm{O}(\varepsilon^0)$; hence $\vec{\Phi}_k^*$ is $\mathrm{O}(\varepsilon^{-1})$. If the total momentum of monomer ℓ, $\vec{P}_\ell^*$, is thermalized (i.e. P_ℓ^{*2}/m_ℓ^* is typically $k_B T$), then $\vec{P}_\ell^*$ is $\mathrm{O}(\varepsilon^{-1/2})$. As $\vec{\Pi}_k^*$ is a sum of L terms of fluctuating direction, then for the thermalized system, $\vec{\Pi}_k^*$ scales as the square root of the

contributing terms times the scaling of one term; thus it scales as ε^{-1}, not $\varepsilon^{-3/2}$ (i.e. many of the L terms in $\vec{\Pi}_k^*$ cancel each other). Therefore,

$$\frac{d\vec{\Phi}_k^*}{dt} = \varepsilon \frac{\vec{\Pi}_k}{m}, \qquad \frac{d\vec{\Phi}_k}{dt} = \varepsilon^2 \frac{\vec{\Pi}_k}{m}, \tag{2.11}$$

for $\vec{\Phi}_k = \varepsilon \vec{\Phi}_k^*$, $\vec{\Pi}_k = \varepsilon \vec{\Pi}_k^*$. We conclude that both $\vec{\Phi}_k$ and $\vec{\Phi}_k^*$ are slow variables. As discussed further below, the $\vec{\Phi}_k^*$ account for configurational changes that create large potential energy variations, while the $\vec{\Phi}_k$ account for large-scale conformational, rotational, and migration changes that do not create appreciable changes in nearest-neighbor interatomic distances.

Newton's equations imply that the $\vec{\Pi}_k^*$ evolve via

$$\frac{d\vec{\Pi}_k^*}{dt} = \sum_{i=1}^{N} u_k(\tau_{\ell(i)}) \vec{F}_i \Theta_i \equiv \vec{f}_k, \tag{2.12}$$

where $\vec{F}_i$ is the force on atom i. $\vec{f}_k$ is assumed to scale as $O(\varepsilon^0)$ due to cancellation of the individual forces when the system is near equilibrium. With this and the introduction of the scaled quantity $\vec{\Pi}_k$, one obtains

$$\frac{d\vec{\Pi}_k}{dt} = \varepsilon \vec{f}_k, \tag{2.13}$$

so that $\vec{\Pi}_k$ evolves slowly, although $\vec{\Pi}_k^*$ does not. It is concluded that the $\vec{\Phi}_k^*$, $\vec{\Phi}_k$, and $\vec{\Pi}_k$ constitute a set of $3k_{\max}$ vector order parameters on which our multiscale analysis of macromolecular dynamics can be built. Note that in the above scaling arguments, it is assumed that the u_k did not fluctuate rapidly in sign. This is consistent with our picture that the macromolecule bends smoothly, i.e. the radius of curvature is large compared to the typical nearest monomer distance.

Individual atomic positions $\vec{r}_i$ are taken to have coherent behavior from the order parameters, and a residual (incoherent) component $\vec{\sigma}_i$ for atom i such that

$$\vec{r}_i = \vec{r}(\tau_{\ell(i)}, \underline{\vec{\Phi}}^*)\Theta_i + \vec{\sigma}_i. \tag{2.14}$$

Given that (2.5) provides a relationship between the order parameters and the atomic positions, this becomes a relationship between the residuals $\vec{\sigma}_i$ and the atomic coordinates.

From the above, we conclude that $\underline{\vec{\Phi}}^*$ and $\underline{\vec{\Pi}}$ co-evolve on the ε^{-1} timescale, while $\underline{\vec{\Phi}}$ evolves on the ε^{-2} scale. Thus, relative to the slowly varying order parameter $\underline{\vec{\Phi}}$, the pair $\underline{\vec{\Phi}}^*$, $\underline{\vec{\Pi}}$ are expected to be highly fluctuating quantities that should be treated stochastically, while $\underline{\vec{\Phi}}$ displays quasi-macroscopic behavior, as noted in Appendix 1.

The N-atom potential $V(\vec{r}_1, \ldots, \vec{r}_N)$ is next expressed in terms of coherent and residual dependencies via $\underline{\vec{\Phi}}^*$, $\underline{\vec{\Phi}}$, and $\vec{\sigma}(\vec{r}_1, \ldots, \vec{r}_N)$. V has both short and long-range contributions. Short-range interactions include the bonded forces and the r^{-12} repulsive core of the Lennard-Jones potential, while coulomb and r^{-6} potentials are common long-range forces. In reality, the interactions change character smoothly as distance between atoms increases. Thus, rather than writing V as a sum of long and short-range parts, we adopt the more realistic form $V(\vec{r}_1, \ldots, \vec{r}_N; \varepsilon\vec{r}_1, \ldots, \varepsilon\vec{r}_N)$. Using (2.14), which relates $\vec{r}_i$ to the residual and coherent parts, we have the function

$$Ul(\vec{r}_1,\ldots,\vec{r}_N;\underline{\vec{\Phi}}^*,\underline{\vec{\Phi}};\varepsilon)$$
$$= V\left(\left\{\Theta_i\left[\sum_{k=1}^{k_{\max}}\vec{\Phi}_k^* u_k(\tau_{\ell(i)})+\vec{\sigma}_i\right]+(1-\Theta_i)\vec{r}_i,\ i=1,\ldots,N\right\};\right.$$
$$\left.\left\{\Theta_i\left[\sum_{k=1}^{k_{\max}}\vec{\Phi}_k u_k(\tau_{\ell(i)})+\varepsilon\vec{\sigma}_i\right]+\varepsilon(1-\Theta_i)\vec{r}_i,\ i=1,\ldots,N\right\}\right). \tag{2.15}$$

From the potential one may write the force $\vec{F}_i$ on atom i in the form

$$\vec{F}_i = \vec{F}_{i(0)} + \varepsilon\vec{F}_{i(1)}, \tag{2.16}$$

where $\vec{F}_{i(0)}$ arises from the $\vec{r}_i$-dependence of V while $\vec{F}_{i(1)}$ is from the $\varepsilon\vec{r}_i$-dependence. Note that $\vec{F}_{i(0)}$ and $\vec{F}_{i(1)}$ have dependence on ε through the long-range dependency of V. Thus $\vec{f}_k$ is given by $\vec{f}_k = \vec{f}_{k(0)} + \varepsilon\vec{f}_{k(1)}$, where

$$\vec{f}_{k(0)} = -\left(\frac{\partial U}{\partial\vec{\Phi}_k^*}\right)_{\vec{\Phi}^*_{k'\neq k},\underline{\vec{\Phi}};\vec{r}_1,\ldots,\vec{r}_N}, \qquad \vec{f}_{k(1)} = -\left(\frac{\partial U}{\partial\vec{\Phi}_k}\right)_{\vec{\Phi}_{k'\neq k},\underline{\vec{\Phi}}^*;\vec{r}_1,\ldots,\vec{r}_N}. \tag{2.17}$$

In this way, the k-force is divided into a strong part $\vec{f}_{k(0)}$ (that one would expect to have zero thermal average as the system rapidly adjusts) and a more persistent but weaker force $\vec{f}_{k(1)}$. In the remainder of this presentation we assume that the host medium only has short-range interactions and that the $\varepsilon\vec{r}_i$ term in the long-range dependence can be neglected as ε is small. Thus

$$U = U(\vec{r}_1,\ldots,\vec{r}_N;\underline{\vec{\Phi}}^*,\underline{\vec{\Phi}}) \tag{2.18}$$

is the form of the N-atom potential. Under this assumption, $\vec{F}_{i(0)}$ and $\vec{F}_{i(1)}$ of (2.16) have no ε dependence, but rather depend on $\vec{r}_1,\ldots,\vec{r}_N,\underline{\vec{\Phi}}^*$, and $\underline{\vec{\Phi}}$ only.

The kinetic energy can also be expressed in terms of coherent and residual momentum contributions. However, as in earlier [19, 20], this has little impact on the multiscale development for the quasi-equilibrium conditions of interest here.

3 A CAM Fokker–Planck Equation for Macromolecular Dynamics

The development to follow has three major advances over our earlier analysis [19, 20]. First, we introduce an automated procedure to generate the order parameters as described in Sect. 2. Second, we build the multiscale analysis on the order parameters $\underline{\vec{\Phi}}^*$ and $\underline{\vec{\Phi}}$ simultaneously; this is key to allowing for the interaction of multiple nanosubunits, i.e. with $\underline{\vec{\Phi}}$ alone, one would encounter overlapping atomic configurations and the forces generated thereby could lead to violations of the multiscale ansatz. Finally, the latter advance allows us to build the analysis on the more realistic assumption that while the short-range forces are large their thermal average acting on a nanoscale structure is small. This allows one to avoid the assumption that the force acting on the nanostructure is small instantaneously.

The order parameters introduced in Sect. 2 are now used to develop a multiscale theory of macromolecular dynamics based on the N-atom probability density ρ and the seven-step scheme outlined in Sect. 1. First, we make the ansatz that ρ has the dependence

$$\rho = \rho(\Gamma,\underline{\vec{\Phi}}^*,\underline{\vec{\Phi}},\underline{\vec{\Pi}},t_0,\underline{t}), \tag{3.1}$$

where $\underline{t} = \{t_1, t_2, \ldots\}$ and $t_n = \varepsilon^n t$. In this way, ρ depends on $\Gamma = \{\vec{p}_1, \vec{r}_1, \ldots, \vec{p}_N, \vec{r}_N\}$ both directly, and indirectly through $\underline{\vec{\Phi}}^*, \underline{\vec{\Phi}}$, and $\underline{\vec{\Pi}}$. The chain rule implies that the Liouville equation $\partial\rho/\partial t = \mathcal{L}\rho$ takes the form (1.1). In the CAM framework this implies

$$\sum_{n=0}^{\infty} \varepsilon^n \frac{\partial \rho}{\partial t_n} = (\mathcal{L}_0 + \varepsilon\mathcal{L}_1 + \varepsilon^2\mathcal{L}_2)\rho, \tag{3.2}$$

$$\mathcal{L}_0 = -\sum_{i=1}^{N}\left[\frac{\vec{p}_i}{m_i}\cdot\frac{\partial}{\partial\vec{r}_i} + \vec{F}_{i(0)}\cdot\frac{\partial}{\partial\vec{p}_i}\right], \tag{3.3}$$

$$\mathcal{L}_1 = -\sum_{k=1}^{k_{\max}}\left[\frac{\vec{\Pi}_k}{m}\cdot\frac{\partial}{\partial\vec{\Phi}_k^*} + \vec{f}_{k(0)}\cdot\frac{\partial}{\partial\vec{\Pi}_k}\right], \tag{3.4}$$

$$\mathcal{L}_2 = -\sum_{k=1}^{k_{\max}}\left[\frac{\vec{\Pi}_{\mathrm{k}}}{m}\cdot\frac{\partial}{\partial\vec{\Phi}_k} + \vec{f}_{k(1)}\cdot\frac{\partial}{\partial\vec{\Pi}_{\mathrm{k}}}\right] - \Xi, \tag{3.5}$$

$$\Xi = \frac{1}{\varepsilon}\sum_{i=1}^{N}\vec{F}_{i(1)}\cdot\frac{\partial}{\partial\vec{p}_i}\Theta_i. \tag{3.6}$$

When operating on ρ in the form (3.1), derivatives with respect to the $\vec{r}_i$ and $\vec{p}_i$ in $\mathcal{L}_0$ are taken at constant $\underline{\vec{\Phi}}^*, \underline{\vec{\Phi}}$, and $\underline{\vec{\Pi}}$, while those with respect to $\vec{\Phi}_k^*$, $\vec{\Phi}_k$, and $\vec{\Pi}_k$ in $\mathcal{L}_1$ and $\mathcal{L}_2$ are to be taken at constant Γ. In writing the operator Ξ, we have assumed that the slowly varying potential is only weakly correlated with the atomic momenta so that Ξ should be treated as $\mathrm{O}(\varepsilon^0)$ due to the many cancellations that occur when it operates on a quasi-equilibrium distribution.

Expanding ρ in an integer power series in ε, the lowest-order solution to (3.2) is found to be

$$\rho_0 = \frac{e^{-\beta H_0}}{Q(\beta, \underline{\vec{\Phi}}^*, \underline{\vec{\Phi}})} W(\underline{\vec{\Phi}}^*, \underline{\vec{\Phi}}, \underline{\vec{\Pi}}, \underline{t}), \tag{3.7}$$

where W is the coarse-grained probability distribution for $\underline{\vec{\Phi}}^*, \underline{\vec{\Phi}}$, and $\underline{\vec{\Pi}}$ that only depends on the slow times $\underline{t}$. H_0 and the nanocanonical partition function Q are given in Appendix 1.

Continuing the perturbation analysis to $\mathrm{O}(\varepsilon)$, we get

$$\rho_1 = \int_0^t dt' e^{\mathcal{L}_0(t-t')}\left[-\hat{\rho}\frac{\partial W}{\partial t_1} + \mathcal{L}_1\hat{\rho}W\right]. \tag{3.8}$$

Using (3.4), the statistical mechanical postulate "the longtime and ensemble average for equilibrium systems are equal", and removing the secular behavior, as developed in detail earlier [19, 20], implies

$$\frac{\partial W}{\partial t_1} = -\sum_{k=1}^{k_{\max}}\frac{\vec{\Pi}_k}{m}\cdot\frac{\partial W}{\partial\vec{\Phi}_k^*}, \tag{3.9}$$

upon recalling that $\vec{f}_{k(0)}$ is a large amplitude force and is assumed to have zero thermal average:

$$\beta\vec{f}_{k(0)}^{th} = \frac{\partial \ln Q}{\partial\vec{\Phi}_k^*} = \vec{0}. \tag{3.10}$$

With this, we obtain the solution

$$\rho_1 = -\hat{\rho}\int_0^t dt' e^{\mathcal{L}_0(t-t')}\sum_{k=1}^{k_{\max}} \vec{f}_{k(0)}\cdot\left(\beta\frac{\vec{\Pi}_k}{m}+\frac{\partial}{\partial\vec{\Pi}_k}\right)W. \tag{3.11}$$

To $O(\varepsilon^2)$, we get

$$\rho_2 = \int_0^t dt' e^{\mathcal{L}_0(t-t')}\left[-\hat{\rho}\frac{\partial W}{\partial t_2}-\frac{\partial\rho_1}{\partial t_1}+\mathcal{L}_1\rho_1+\mathcal{L}_2\rho_0\right]. \tag{3.12}$$

Using (3.4), (3.5), (3.7), and (3.11), and continuing as above, implies

$$\frac{\partial W}{\partial t_2} = -\sum_{k=1}^{k_{\max}}\left[\frac{\vec{\Pi}_k}{m}\cdot\frac{\partial}{\partial\vec{\Phi}_k}+\vec{f}^{\,th}_{k(1)}\cdot\frac{\partial}{\partial\vec{\Pi}_k}\right]W+\sum_{k,k'=1}^{k_{\max}}\vec{\vec{\gamma}}_{kk'}\frac{\partial}{\partial\vec{\Pi}_k}\cdot\left[\beta\frac{\vec{\Pi}_{k'}}{m}+\frac{\partial}{\partial\vec{\Pi}_{k'}}\right]W. \tag{3.13}$$

The thermal average force $\vec{f}^{\,th}_{k(1)}$ can be computed using molecular dynamics and the long-time average, i.e.

$$\vec{f}^{\,th}_{k(1)} = \lim_{t\to\infty}\frac{1}{t}\int_{-t}^{0} dt' e^{-\mathcal{L}_0 t'}\vec{f}_{k(1)}, \tag{3.14}$$

$$\vec{f}_{k(1)} = \sum_{i=1}^{N}\vec{F}_{i(1)}u_k(\tau_{\ell(i)})\Theta_i. \tag{3.15}$$

The friction coefficients are related to force auto-correlation functions:

$$\gamma_{k\alpha k'\alpha'} = \int_0^\infty dt_0 (f_{k\alpha(0)}(0) f_{k'\alpha'(0)}(t_0))^{th}, \quad k,k' = 1,\ldots,k_{\max},\ \alpha,\alpha' = 1,2,3. \tag{3.16}$$

The result (3.13) appears to be in error as additional terms from $\frac{\partial\hat{\rho}}{\partial\vec{\Phi}_k}$ are not presented. It is found that these terms would have violated detailed balance and conservation of total probability (i.e. the integral of W over all $\vec{\Pi}_k$ and order parameters should be a constant). However, the omitted terms drop from the analysis when an additional contribution to ρ_1 that is a null vector of the operator $\mathcal{L}_0 - \frac{\partial}{\partial t_0}$ is included in ρ_1. This theme will be developed in more detail in future work.

Combining (3.9) and (3.13), we arrive at a reconstituted equation for W, the FP equation. A special case is when $\vec{f}_{k(0)}$ depends on $\underline{\vec{\Phi}}$ rather than both $\underline{\vec{\Phi}}$ and $\underline{\vec{\Phi}}^*$. In that event, there are solutions for which W is independent of $\underline{\vec{\Phi}}^*$. With this, we obtain

$$\frac{\partial W}{\partial\tau} = \mathcal{D}'W-\sum_{k=1}^{k_{\max}}\left[\frac{\vec{\Pi}_k}{m}\cdot\frac{\partial}{\partial\vec{\Phi}_k}+\vec{f}^{\,th}_{k(1)}\cdot\frac{\partial}{\partial\vec{\Pi}_k}\right]W, \tag{3.17}$$

$$\mathcal{D}' = \sum_{k,k'=1}^{k_{\max}}\vec{\vec{\gamma}}_{kk'}\frac{\partial}{\partial\vec{\Pi}_k}\cdot\left[\beta\frac{\vec{\Pi}_{k'}}{m}+\frac{\partial}{\partial\vec{\Pi}_{k'}}\right], \tag{3.18}$$

where $\tau = t_2 = \varepsilon^2 t$.

4 Friction-Dominated (Smoluchowski) Dynamics

When the macromolecule is immersed in a viscous medium or is in intimate content with itself or surrounding macromolecules, we expect that friction effects could dominate the inertial terms in (3.17). Examples include RNA or DNA in a viral capsid or the genomic macromolecules in a bacterium or eukaryotic nucleus. The derivation of the Smoluchowski equation from an FP equation for a simpler case has been demonstrated earlier [26]. In this section, we examine the friction-dominated limit via a perturbation analysis as applied to (3.17).

In this limit, the $\vec{\vec{\gamma}}_{kk'}$ are large. To express this, we rewrite (3.17) in the form

$$\mathcal{D}W = \eta \frac{\partial W}{\partial \tau} + \eta \sum_{k=1}^{k_{\max}} \left[\frac{\vec{\Pi}_k}{m} \cdot \frac{\partial W}{\partial \vec{\Phi}_k} + \vec{f}^{th}_{k(1)} \cdot \frac{\partial W}{\partial \vec{\Pi}_k} \right], \tag{4.1}$$

$$\mathcal{D} = \sum_{k,k'=1}^{k_{\max}} \vec{\vec{g}}_{kk'} \frac{\partial}{\partial \vec{\Pi}_k} \cdot \left[\beta \frac{\vec{\Pi}_k}{m} - \frac{\partial}{\partial \vec{\Pi}_{k'}} \right], \tag{4.2}$$

where $\vec{\vec{g}}_{kk'} = \eta \vec{\vec{\gamma}}_{kk'}$, η^{-1} being a typical value of the $\gamma_{k\alpha k'\alpha'}$ $(\alpha, \alpha' = 1, 2, 3)$. In this framework, η is small in the friction-dominated regime.

Expanding W in a power series in η,

$$W = \sum_{j=0}^{\infty} W_j \eta^j, \tag{4.3}$$

we carry out an order-by-order analysis.

To $O(\eta^0)$ the FP equation (4.1) implies $\mathcal{D}W_0 = 0$. This admits the solution

$$W_0 = \frac{\exp[-\beta \sum_{k=1}^{k_{\max}} \Pi_k^2/2m]}{B} \Psi(\underline{\vec{\Phi}}, \tau_0, \tau_1, \ldots) \equiv \hat{W}\Psi, \tag{4.4}$$

for reduced distribution Ψ (to be determined in the higher order analysis), $\tau_n = \eta^n \tau$ $(n = 0, 1, \ldots)$, and factor B that normalizes $\hat{W}$.

Terms of $O(\eta^1)$ yield

$$\frac{\partial W_0}{\partial \tau_0} = -\sum_{k=1}^{k_{\max}} \left[\frac{\vec{\Pi}_k}{m} \cdot \frac{\partial W_0}{\partial \vec{\Phi}_k} + \vec{f}^{th}_{k(1)} \cdot \frac{\partial W_0}{\partial \vec{\Pi}_k} \right] + \mathcal{D}W_1. \tag{4.5}$$

Integration of both sides of (4.5) over all $\vec{\Pi}_k$ shows that Ψ is independent of τ_0. Upon using (4.4) for W_0, (4.5) becomes

$$0 = -\sum_{k=1}^{k_{\max}} \frac{\vec{\Pi}_k}{m} \cdot \left[\frac{\partial \Psi}{\partial \vec{\Phi}_k} - \beta \vec{f}^{th}_{k(1)} \Psi \right] + \mathcal{D}W_1^{\mathrm{m}}, \tag{4.6}$$

where $W_1 = \hat{W} W_1^{\mathrm{m}}$. Consider a solution in the form $W_1^{\mathrm{m}} = \vec{A}_1 \cdot \vec{\Pi}_1 + \cdots + \vec{A}_{k_{\max}} \cdot \vec{\Pi}_{k_{\max}}$. With this, we find

$$\sum_{q=1}^{k_{\max}} \sum_{\alpha=1}^{3} \frac{\Pi_{q\alpha}}{m} \left[\frac{\partial \Psi}{\partial \Phi_{q\alpha}} - \beta f^{th}_{q\alpha(1)} \Psi - \beta G A_{q\alpha} - \beta \sum_{\hat{q}=1}^{k_{\max}} \sum_{\hat{\alpha}=1}^{3} g_{\hat{q}\hat{\alpha}q\alpha} A_{\hat{q}\hat{\alpha}} \right] = 0, \tag{4.7}$$

$$G = \sum_{k=1}^{k_{\max}} \mathrm{tr}(\vec{\vec{g}}_{kk}), \tag{4.8}$$

where tr indicates the trace. As this must hold for arbitrary $\Pi_{q\alpha}$, we conclude

$$\beta \sum_{\hat{q}=1}^{k_{\max}} \sum_{\hat{\alpha}=1}^{3} [G\delta_{q\hat{q}}\delta_{\alpha\hat{\alpha}} + g_{\hat{q}\hat{\alpha}q\alpha}]A_{\hat{q}\hat{\alpha}} = \frac{\partial \Psi}{\partial \Phi_{q\alpha}} - \beta f^{th}_{q\alpha}\Psi, \tag{4.9}$$

constituting $3k_{\max}$ equations to be solved for the $A_{k\alpha}$.

Collecting terms of $\mathrm{O}(\eta^2)$, yields

$$\frac{\partial W_0}{\partial \tau_1} = -\sum_{k=1}^{k_{\max}} \left[\frac{\vec{\Pi}_k}{m} \cdot \frac{\partial W_1}{\partial \vec{\Phi}_k} + \vec{f}^{th}_{k(1)} \cdot \frac{\partial W_1}{\partial \vec{\Pi}_k} \right] + \mathcal{D} W_2. \tag{4.10}$$

Integration over all $\vec{\Pi}_k$ yields

$$\frac{\partial \Psi}{\partial \tau_1} = -\frac{1}{\beta} \sum_{k=1}^{k_{\max}} \sum_{\alpha=1}^{3} \frac{\partial A_{k\alpha}}{\partial \Phi_{k\alpha}}. \tag{4.11}$$

Let $\chi_{\hat{q}\hat{\alpha}q\alpha} = G\delta_{q\hat{q}}\delta_{\alpha\hat{\alpha}} + g_{\hat{q}\hat{\alpha}q\alpha}$, and $\mathcal{J}_{q\alpha} = \frac{\partial \Psi}{\partial \Phi_{q\alpha}} - \beta f^{th}_{q\alpha}\Psi$, then (4.11) becomes

$$\beta \sum_{\hat{q}=1}^{k_{\max}} \sum_{\hat{\alpha}=1}^{3} \chi_{\hat{q}\hat{\alpha}q\alpha} A_{\hat{q}\hat{\alpha}} = \mathcal{J}_{q\alpha}, \tag{4.12}$$

which can be rewritten as

$$-\frac{1}{\beta} A_{q\alpha} = \sum_{\hat{q}=1}^{k_{\max}} \sum_{\hat{\alpha}=1}^{3} \chi^{-1}_{\hat{q}\hat{\alpha}q\alpha} \mathcal{J}_{\hat{q}\hat{\alpha}}, \tag{4.13}$$

where $(\chi^{-1})_{q\alpha\hat{q}\hat{\alpha}}(\chi)_{q\alpha\hat{q}\hat{\alpha}} = -\beta^2 \delta_{q\hat{q}}\delta_{\alpha\hat{\alpha}}$. With this, we get

$$-\frac{1}{\beta} \frac{\partial A_{q\alpha}}{\partial \Phi_{q\alpha}} = \sum_{\hat{q}=1}^{k_{\max}} \sum_{\hat{\alpha}=1}^{3} \chi^{-1}_{q\alpha\hat{q}\hat{\alpha}} \frac{\partial \mathcal{J}_{\hat{q}\hat{\alpha}}}{\partial \Phi_{q\alpha}}. \tag{4.14}$$

Inserting (4.14) in (4.11) yields

$$\frac{\partial \Psi}{\partial \tau_1} = \sum_{q,\hat{q}=1}^{k_{\max}} \sum_{\alpha,\hat{\alpha}=1}^{3} \chi^{-1}_{q\alpha\hat{q}\hat{\alpha}} \frac{\partial}{\partial \Phi_{q\alpha}} \left[\frac{\partial \Psi}{\partial \Phi_{\hat{q}\hat{\alpha}}} - \beta f^{th}_{\hat{q}\hat{\alpha}}\Psi \right]. \tag{4.15}$$

Thus, Ψ satisfies a diffusion-like equation wherein $\vec{f}^{th}_{k(1)}$ acts like an applied field that tends to focus density in $\underline{\vec{\Phi}}$-space to places where U^{th} is a minimum. In addition, Ψ is driven to the equilibrium distribution that is proportional to $\exp(-\beta U^{th})$ so that if the macromolecule is in one free energy well, it gradually readjusts itself so that it has probability of being in others via a diffusion-like spreading of Ψ.

5 Conclusions

The all-atom, multiscale analysis and the curvilinear coordinate order parameters hold great promise for the efficient simulation of large macromolecules over long periods of time. Since the formulation is all-atom, interactions of macromolecules with drug molecules or cell surface receptors can be investigated; as our framework is multiscale, conclusions on whole macromolecule conformational dynamics induced by these interactions can be derived. The friction-dominated case (Sect. 4) provides Smoluchowski-type approximations relevant for systems such as the viral genome with its capsid, a macromolecule in a viscous host medium, or genomic macromolecules in a bacterium, mitochondrium, or nucleus.

Advances in molecular dynamics make the thermal average force and friction-tensor computations feasible (see Ref. [20] for a brief review). Freedom to choose the basis functions $u_k(\tau)$, or other representation of $\vec{r}(\tau, \underline{\vec{\Phi}}^*)$, allows for the optimization of the method by tailoring it to particular phenomena. For example, one could use basis functions to characterize two conformations (and their rotational equivalents) and then the theory would provide an estimate of the rate of transition between them.

The conceptual framework developed here suggests the following simulation algorithm. Equivalent to the FP equation is a set of Langevin equations for the order parameters $\underline{\vec{\Phi}}$ and related momenta $\underline{\vec{\Pi}}$:

$$\frac{d\underline{\vec{\Phi}}}{dt} = \underline{\vec{\Pi}}, \tag{5.1}$$

$$\frac{d\underline{\vec{\Pi}}}{dt} = \underline{f}^{th} + \underline{f}^{fr} + \underline{A}, \tag{5.2}$$

for thermal average force $\underline{f}^{th}$, frictional forces $\underline{f}^{fr}$, and random forces $\underline{A}$ whose correlation is determined by the friction tensors. As $\underline{f}^{th}$ and $\underline{f}^{fr}$ change on a timescale much longer than that of the random force, modern solution techniques for solving stochastic differential equations [27] such as hybrid stochastic methods that divide the dynamics into fast and slow processes [28–30] can be used to solve (5.1, 5.2) using timesteps that are on the order of the characteristic time for changes in $\underline{\vec{\Phi}}$ and $\underline{\vec{\Pi}}$. The latter can be 10^{-3} seconds or longer, in sharp contrast to the 10^{-13} second timestep needed in molecular dynamics. Thus, this 10^{10} difference can easily absorb the extra computations needed at each timestep to carry out the ensemble averaging required to obtain $\underline{f}^{th}$ and $\underline{f}^{fr}$. In this way, the present framework can make large macromolecule, all-atom, long-time simulations feasible that cannot be achieved otherwise.

The ability to capture all-atom detail in long-time macromolecular simulations will allow one to develop a detailed understanding of experimental data. Consider a double-well free energy profile underlying two distinct macromolecular conformations. If a magnetic nanoparticle is attached to one end and an oscillatory magnetic field is applied, then several distinct resonances are expected. The latter include a harmonic-like oscillation within each well, and a tumbling of the macromolecule/nanoparticle complex. These resonances depend on the amplitude of the applied field, indicating the possibility of a resonant induced hopping between the free energy wells. Judicious positioning and choice of the mass of the attached magnetic nanoparticle could probe distinct resonances. If a quantum dot was also attached to the macromolecule, florescent fluctuations at the frequency of the applied magnetic field could improve signal-to-noise ratios, facilitating the detection of macromolecular structural transitions and fluctuations.

The present approach has a number of applications to key biological macromolecular phenomena. Complexing of a transcription factor at one gene could create long-range conformational changes that can regulate another. Attachment of a therapeutic agent to mRNA transcribed from a given gene could change its conformation and thus alter its translation kinetics. Detection of sequence abnormalities in the DNA via SNP could be analyzed by major conformational implications for the transcripts or translates of the affected gene.

Given the need for long-time simulation of macromolecular dynamics with atomic-scale resolution, it is our hope that the methods developed here will facilitate pure and applied research in macromolecular phenomena.

Appendix 1: The Nanocanonical Ensemble

An entropy maximization approach is now used to construct an ensemble for a nanosystem, and will serve as the lowest order statistical state for the multiscale analysis of the Liouville equation. The lowest order equation in the multiscale perturbation hierarchy of Sect. 3 is $\mathcal{L}_0\rho_0 = 0$. Recalling that $\mathcal{L}_0$ involves partial derivatives with respect to the set of atomic positions and momenta Γ at constant values of the order parameters $\underline{\vec{\Phi}}^*$, and the scaled variables $\underline{\vec{\Phi}}$ and $\underline{\vec{\Pi}}$, it is seen that ρ_0 can be any function of (1) variables in the null space of $\mathcal{L}_0$, and (2) $\underline{\vec{\Phi}}^*$, $\underline{\vec{\Phi}}$, and $\underline{\vec{\Pi}}$. As we seek quasi-equilibrium solutions of the Liouville equation, we conjecture that the relevant ensemble underlying ρ_0 can be deduced via maximization of the entropy S, $S = -k_B \int d^{6N}\Gamma \rho_0 \ln \rho_0$, subject to normalization and physical constraints relevant to the nanosystem of interest.

For isothermal systems, we take the ensemble to have a canonical flavor by constraining the entropy maximization to a given average energy. In our physical picture of the nanosystem, $\underline{\vec{\Phi}}^*$ represents intermediate space and time scale dynamics of the nanosystem, and therefore only its average can be constrained, and similarly for $\underline{\vec{\Pi}}$ which, from Sect. 2, evolves on the same timescale as $\underline{\vec{\Phi}}^*$. However, $\underline{\vec{\Phi}}$ is so slowly varying that its value can be considered known, i.e. is quasi-macroscopic, and not just its average value is known. Thus, we construct the "pre-nanocanonical ensemble" with probability $\rho_{\underline{\vec{\mu}}\underline{\vec{\kappa}}}$ determined by maximizing the auxiliary function $\tilde{S}$ defined via

$$\tilde{S} = S - k_B \int d^{6N}\Gamma \left\{ \lambda + \beta H_0 - \sum_{q=1}^{k_{\max}} \sum_{\alpha=1}^{3} (\mu_{q\alpha}\Phi'_{q\alpha} + \kappa_{q\alpha}\Pi'_{q\alpha}) \right\} \rho_{\underline{\vec{\mu}}\underline{\vec{\kappa}}} \tag{6.1}$$

where $\underline{\vec{\Phi}}'$ and $\underline{\vec{\Pi}}'$ are expressions for $\underline{\vec{\Phi}}^*$ and $\underline{\vec{\Pi}}$ in terms of Γ, H_0 is the Hamiltonian generating $\mathcal{L}_0$, and $\lambda, \beta, \underline{\vec{\mu}}$, and $\underline{\vec{\kappa}}$ are Lagrange multipliers. Maximizing $\tilde{S}$ over all $\rho_{\underline{\mu}\underline{\kappa}}$ implies

$$\rho_{\underline{\vec{\mu}}\underline{\vec{\kappa}}} = \frac{\exp\{-\beta H_0 + \underline{\vec{\mu}}\underline{\vec{\Phi}}' + \underline{\vec{\kappa}}\underline{\vec{\Pi}}'\}}{\mathrm{Z}(\beta, \underline{\vec{\mu}}, \underline{\vec{\kappa}}, \underline{\vec{\Phi}})}, \tag{6.2}$$

$$\mathrm{Z}(\beta, \underline{\vec{\mu}}, \underline{\vec{\kappa}}, \underline{\vec{\Phi}}) = \int d^{6N}\Gamma \exp\{-\beta H_0 + \underline{\vec{\mu}}\underline{\vec{\Phi}}' + \underline{\vec{\kappa}}\underline{\vec{\Pi}}'\}. \tag{6.3}$$

Proceeding as earlier [19], we seek a more general solution of $\mathcal{L}_0\rho_0 = 0$ expressed as a linear combination of the $\rho_{\underline{\vec{\mu}}\underline{\vec{\kappa}}}$. The final consequence of this computation is

$$\rho_0 = \frac{e^{-\beta H_0}}{Q(\beta, \underline{\vec{\Phi}}^*, \underline{\vec{\Phi}})} W(\underline{\vec{\Phi}}^*, \underline{\vec{\Phi}}, \underline{\vec{\Pi}}, \underline{t}), \tag{6.4}$$

where

$$Q(\beta, \underline{\vec{\Phi}}^*, \underline{\vec{\Phi}}) = \int d^{6N} \Gamma \delta(\underline{\vec{\Phi}}^* - \underline{\vec{\Phi}}') \exp(-\beta H_0). \tag{6.5}$$

Some of the dependence on $\underline{\vec{\Pi}}$ that would have been in Q cancels with a factor in $\exp(-\beta H_0)$ via a statistical argument [19], while the remainder was embedded in W. With this, H_0 as it appears in (6.4) is given by

$$H_0 = K + U(\underline{\vec{\Phi}}^*, \underline{\vec{\Phi}}, \vec{r}_1, \ldots, \vec{r}_N), \tag{6.6}$$

and W is a residual probability factor that depends on $\underline{t} = \{t_1, t_2, \ldots\}$, U is the N-atom potential energy expressed in terms of the order parameters and residual dependence on the atomic position, and K is $p_1^2/2m_1 + \cdots + p_N^2/2m_N$. We term ρ_0 the "nanocanonical" probability density.

Appendix 2: The Basis Functions

Optimum choice of basis functions leads to order parameters that are best suited for the system and phenomenon of interest. For example, the DNA of bacteria and mitochondria is organized as a closed loop. In that case, $u_k(\tau)$ is best chosen to be a periodic function of τ. Whatever basis functions are used, they should fit the following criteria: (1) they should be orthonormal, and (2) they should not include a weighing function so that all points along the macromolecule can be treated in an equivalent manner. This suggests that, among common orthogonal polynomials, Legendre polynomials would be the most suitable. However, the orthogonality conditions of the latter involve an integral instead of a discreet sum as is needed in our case. Using the Gram-Schmidt process, we modify the Legendre polynomials to generate a set of discreetly orthonormal polynomials; i.e. the set of basis functions is generated such that (2.4) is satisfied.

To demonstrate the use of these basis functions take the peptide antibiotic Alamethicin as an example. Alamethicin consists of 20 amino acids and is known to assume a helical conformation [31]. To create Fig. 1, the positions of the backbone atoms are used as a starting point. These are taken from a pdb file for one structural conformation of Alamethicin. The center of mass of each monomer is calculated (the black dots in Fig. 1). For this structure, 19 polynomials are needed for an accurate representation. Applying (2.5), we obtain the set of order parameters, and (2.1) is then used to plot the curve. Since the algorithm used to generate the basis functions yields the values of u_k at the integer values of τ only, we use a least-square fitting method in order to find the coefficients that define each polynomial.

Performing the suggested simulation algorithm described in Sect. 5 would yield the fluctuating dynamics of the macromolecule. Note, different values of the order parameters give different structures over the time coarse of evolution while the basis functions are generated only once at the beginning. While there is 57 order parameters in this description, the number of atomistic degrees of freedom is 867. Since the CPU requirement increases roughly with the square of the number of degrees of freedom, our approach has great advantage.

Acknowledgements Discussions with T. Keyes (Boston University) and J. Schmidt (Stanford University) encouraged us to extend our all-atom multiscale analysis to the macromolecular dynamics problem. Support of the U.S. Department of Energy, Office of Science, and the Indiana University College of Arts and Sciences is greatly appreciated.

References

1. Durup, J.: Protein molecular-dynamics constrained to slow modes—theoretical approach based on a hierarchy of local modes with a set of holonomic constraints—the method and its tests on citrate synthase. J. Phys. Chem. **95**(4), 1817–1829 (1991)
2. Askar, A., Space, B., Rabitz, H.: Subspace method for long-time scale molecular-dynamics. J. Phys. Chem. **99**(19), 7330–7338 (1995)
3. Chun, H.M., Padilla, C.E., Chin, D.N., Watanabe, M.W., Karlov, V.I., Alper, H.E., Soosaar, K., Blair, K.B., Becker, O.M., Caves, L.S.D., Nagle, R., Haney, D.N., Farmer, B.L.: MBO(N)D: a multibody method for long-time molecular dynamics simulations. J. Comput. Chem. **21**(3), 159–184 (2000)
4. Elezgaray, J., Sanejouand, Y.H.: Modeling large-scale dynamics of proteins. Biopolymers **46**(7), 493–501 (1998)
5. Elezgaray, J., Sanejouand, Y.H.: Modal dynamics of proteins in water. J. Comput. Chem. **21**(14), 1274–1282 (2000)
6. Feenstra, K.A., Hess, B., Berendsen, H.J.C.: Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. J. Comput. Chem. **20**(8), 786–798 (1999)
7. Phelps, D.K., Post, C.B.: A novel basis for capsid stabilization by antiviral compounds. J. Mol. Biol. **254**(4), 544–551 (1995)
8. Phelps, D.K., Rossky, P.J., Post, C.B.: Influence of an antiviral compound on the temperature dependence of viral protein flexibility and packing: a molecular dynamics study. J. Mol. Biol. **276**(2), 331–337 (1998)
9. Reich, S.: Smoothed dynamics of highly oscillatory Hamiltonian systems. Physica D **89**(1–2), 28–42 (1995)
10. Sorensen, M.R., Voter, A.F.: Temperature-accelerated dynamics for simulation of infrequent events. J. Chem. Phys. **112**(21), 9599–9606 (2000)
11. Space, B., Rabitz, H., Askar, A.: Long-time scale molecular-dynamics subspace integration method applied to anharmonic crystals and glasses. J. Chem. Phys. **99**(11), 9070–9079 (1993)
12. Speelman, B., Brooks, B.R., Post, C.B.: Molecular dynamics simulations of human rhinovirus and an antiviral compound. Biophys. J. **80**(1), 121–129 (2001)
13. Teleman, O., Jönsson, B.: Vectorizing a general-purpose molecular-dynamics simulation program. Comput. Chem. **7**(1), 58–66 (1986)
14. Tuckerman, M.E., Berne, B.J., Rossi, A.: Molecular-dynamics algorithm for multiple time scales-systems with disparate masses. J. Chem. Phys. **94**(2), 1465–1469 (1991)
15. Tuckerman, M.E., Berne, B.J.: Molecular-dynamics in systems with multiple time scales-systems with stiff and soft degrees of freedom and with short and long-range forces. J. Chem. Phys. **95**(11), 8362–8364 (1991)
16. Tuckerman, M.E., Berne, B.J., Martyna, G.J.: Molecular-dynamics algorithm for multiple time scales-systems with long-range forces. J. Chem. Phys. **94**(10), 6811–6815 (1991)
17. Li, X., Hassan, S.A., Mehler, E.L.: Long dynamics simulations of proteins using atomistic force fields and a continuum representation of solvent effects: calculation of structural and dynamic properties. Proteins **60**(3), 464–484 (2005)
18. Shen, M., Freed, K.F.: Long time dynamics of met-enkephalin: tests of mode-coupling theory and implicit solvent models. J. Chem. Phys. **118**(11), 5143–5156 (2003)
19. Miao, Y., Ortoleva, P.: All-atom multiscaling and new ensembles for dynamical nanoparticles. J. Chem. Phys. **125**(4), 044901 (2006)
20. Miao, Y., Ortoleva, P.: Viral structural transitions: an all-atom multiscale theory. J. Chem. Phys. **125**(21), 214901 (2006)
21. Shea, J.E., Oppenheim, I.: Fokker–Planck equation and Langevin equation for one Brownian particle in a nonequilibrium bath. J. Phys. Chem. **100**(49), 19035–19042 (1996)
22. Shea, J.E., Oppenheim, I.: Fokker–Planck equation and non-linear hydrodynamic equations of a system of several Brownian particles in a non-equilibrium bath. Physica A **247**(1–4), 417–443 (1997)
23. Peters, M.H.: Fokker–Planck equation and the grand molecular friction tensor for coupled rotational and translational motions of structured Brownian particles near structured surfaces. J. Chem. Phys. **110**(1), 528–538 (1999)
24. Peters, M.H.: Fokker–Planck equation, molecular friction, and molecular dynamics for Brownian particle transport near external solid surfaces. J. Stat. Phys. **94**(3), 557–586 (1999)
25. Ortoleva, P.: Nanoparticle dynamics: a multiscale analysis of the Liouville equation. J. Phys. Chem. **109**(45), 21258–21266 (2005)
26. Peters, M.H.: The Smoluchowski diffusion equation for structural macromolecules near structured surfaces. J. Chem. Phys. **112**(12), 5488–5498 (2000)
27. Kloeden, P.E., Platen, E.: Numerical solution of Stochastic Differential Equations, 1st edn. Springer, Berlin (1992)

28. Haseltine, E.L., Rawling, J.B.: Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. J. Chem. Phys. **117**(15), 6959–6969 (2002)
29. Puchalka, J., Kierzek, A.M.: Bridging the gap between stochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks. Biophys. J. **86**(3), 1357–1372 (2004)
30. Salis, H., Kaznessis, Y.: Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. J. Chem. Phys. **122**(5), 054103 (2005)
31. Fox, Jr., R.O., Richards, F.M.: A voltage-gated ion channel model inferred from the crystal structure of alamethicin at 1.5-Å resolution. Nature **300**, 325–330 (1982)