Theory for the nonequilibrium dynamics of flexible chain molecules: Relaxation to equilibrium of pentadecane from an all-*trans* conformation

Wilfred H. Tang
James Franck Institute and the Department of Chemistry, University of Chicago, Chicago, Illinois 60637

Konstantin S. Kostov
James Franck Institute and the Department of Biochemistry and Molecular Biology, University of Chicago, Chicago, Illinois 60637

Karl F. Freed
James Franck Institute and the Department of Chemistry, University of Chicago, Chicago, Illinois 60637

We extend to nonequilibrium processes our recent theory for the long time dynamics of flexible chain molecules. While the previous theory describes the equilibrium motions for any bond or interatomic separation in (bio)polymers by time correlation functions, the present extension of the theory enables the prediction of the nonequilibrium relaxation that occurs in processes, such as T-jump experiments, where there are sudden transitions between, for example, different equilibrium states. As a test of the theory, we consider the ''unfolding'' of pentadecane when it is transported from a constrained all-*trans* conformation to a random-coil state at thermal equilibrium. The time evolution of the mean-square end-to-end distance $\langle R_{end}^2(t) \rangle_{noneq}$ after release of the constraint is computed both from the theory and from Brownian dynamics (BD) simulations. The lack of time translational symmetry for nonequilibrium processes requires that the BD simulations of the relaxation of $\langle R_{end}^2(t) \rangle_{noneq}$ be computed from an average over a *huge number* of independent trajectories, rather than over successive configurations from a single trajectory, which may be used to generate equilibrium time correlation functions. Adequate convergence ensues for the nonequilibrium simulations only after averaging 9000 trajectories, each of 0.8 ns duration. In contrast, the theory requires only equilibrium averages for the initial and final states, which may be readily obtained from a *few* Brownian dynamics trajectories. Therefore, the new method produces enormous savings in computer time. Moreover, since both theory and simulations use identical potentials and solvent models, the theory contains no adjustable parameters. The predictions of the theory for the relaxation of $\langle R_{end}^2(t) \rangle_{noneq}$ agree very well with the BD simulations. This work is a starting point for the application of the new method to nonequilibrium processes with biological importance such as the helix-coil transition and protein folding. © *1998 American Institute of Physics.* [S0021-9606(98)50720-9]

I. INTRODUCTION

The nonequilibrium dynamics of polymers is important for understanding a variety of phenomena. For instance, the dynamics of protein folding/unfolding is a fundamental biological process. Our understanding of this process, however, is severely limited by a shortage of adequate theoretical methods for studying the protein *dynamics* with realistic molecular models. Experimentally, it is difficult to achieve sufficiently short time resolution to follow the full dynamics of protein folding/unfolding. Conversely, the main theoretical technique, molecular dynamics, cannot reach time scales long enough to probe many of the interesting events due to the tremendous amount of computation required.

We have been developing a theory for the long time dynamics of flexible polymers and polypeptides in solution. Our mode coupling theory[1–5] is based on a significant extension of the generalized Rouse–Zimm theory[6–15] to include contributions to the dynamics from the memory function (''internal friction'') terms that are customarily ignored. Prior papers demonstrate that the mode coupling theory pro-

duces an excellent representation of the long time equilibrium time correlation functions for both alkanes and short polypeptides. The present paper describes the extension of the mode coupling theory to treat the nonequilibrium dynamics occurring when a system in equilibrium (or in a constrained equilibrium) is suddenly transported to a different equilibrium (perhaps constrained) state, such as in T-jump experiments.

We apply the theory to describe the relaxation of pentadecane from nonequilibrium initial conditions to equilibrium at a temperature of 300 K. The pentadecane chain begins in its maximally stretched conformation, in which all the dihedral angles are *trans*. The molecule subsequently relaxes to equilibrium, where there is a distribution of dihedral angles due to thermal fluctuations. The direct simulation of this nonequilibrium process requires orders of magnitude more computer time than is necessary for simulating equilibrium time correlation functions. The complexity of the nonequilibrium simulations is the reason pentadecane is taken to begin in the all-*trans* configuration, rather than using an initial thermal distribution at a temperature lower than 300 K,

8736

where a huge number of additional nonequilibrium trajectories would be necessary to average over the thermal distribution of initial conditions. In contrast, the treatment of such a T-jump process by the theory merely doubles the more limited computational effort necessary to describe the dynamics beginning from the all-*trans* pentadecane. Studying the ''unfolding'' dynamics of this relatively simple system provides insight into how to apply the mode coupling theory to more complicated biological systems, such as the unwinding of an α-helix or the unfolding of a protein.

A crucial step in the theory is the selection of a suitable basis set. Due to symmetry, the basis functions for describing nonequilibrium dynamics must differ from those previously used for calculating equilibrium dipole time correlation functions. However, the same principles apply independent of the dynamical observables of interest. For equilibrium dipole time correlation functions, a simple first-order basis set produces reasonable estimates[1–5] of alkane dynamics, and a more sophisticated sorting method yields excellent approximations while maintaining a manageable basis set size.[4,5] The present paper extends our mode coupling procedure to determine the appropriate basis sets for describing the long time portions of the nonequilibrium dynamics. As in previous studies, we test the effectiveness of various basis sets by comparing the predictions of the theory with ''exact'' results from Brownian dynamics simulations. Because all the equilibrium inputs required by the theory are taken from the Brownian dynamics simulations, there are *no* free parameters in this comparison.

Section II presents the general theory for describing the nonequilibrium dynamics, while the Brownan dynamics (BD) simulation methods and united atom alkane potentials are described in the subsequent two sections. Theory and simulations are compared in Sec. V with *no* adjustable parameters. The final section discusses the potential application of the theory to more complex systems, such as proteins.

II. NONEQUILIBRIUM AVERAGES FROM THE THEORY

We consider a polymer consisting of N beads. Denote the $3N$ Cartesian coordinates of the polymer beads by $\mathbf{r}(t) = \{r_1(t), r_2(t), \ldots, r_{3N}(t)\} = \{x_1(t), y_1(t), z_1(t), x_2(t), y_2(t), z_2(t), \ldots, x_N(t), y_N(t), z_N(t)\}$. The dynamics of this model polymer are assumed to be governed by the Smoluchowski equation,[16–18]

$$\frac{\partial P(\mathbf{r},t|\mathbf{r_0},0)}{\partial t} = \mathscr{D}(\mathbf{r})P(\mathbf{r},t|\mathbf{r_0},0), \qquad (1)$$

$$\mathscr{D}(\mathbf{r}) = \sum_{i=1}^{3N} \left(\frac{k_B T}{\zeta} \frac{\partial^2}{\partial r_i^2} + \frac{1}{\zeta} \frac{\partial}{\partial r_i} \frac{\partial U(\mathbf{r})}{\partial r_i} \right), \qquad (2)$$

where $P(\mathbf{r},t|\mathbf{r_0},0)$ denotes the probability that the coordinates of the N beads at time t are $\mathbf{r}$, given that the coordinates of the N beads at time 0 are $\mathbf{r_0}$, k_B is Boltzmann's constant, T is the absolute temperature, $U(\mathbf{r})$ is the potential energy, and ζ is the bead friction coefficient. The only explicit degrees of freedom in these equations are the positions

$\mathbf{r}$ of the polymer beads. The solvent affects the motion of the polymer through implicit frictional and stochastic forces. The frictional force is taken to be large enough that the motion lies is in the overdamped regime—that is, inertial effects may be neglected. We also neglect hydrodynamic interactions between different beads, but the general theory below may readily be applied including hydrodynamic interactions in Eq. (2). The hydrodynamic interactions are omitted here to reduce the computational labor but have been treated previously within the generalized Rouse–Zimm theory.[14] The simplifying nature of these assumptions is useful for specifying a model for which it is possible to make rigorous, parameter-free comparisons between theory and ''exact'' results from computer simulations of Eq. (1). Moreover, this particular model contains many of the important physical features without being overly complicated. After understanding the application of the theory to this relatively simple system, the theory can be generalized to treat more realistic and complex systems. For example, the equilibrium theory has recently been applied to describe the dynamics in alkane melts where the dynamics is governed by the full Liouville equation.[5]

We are interested in describing the relaxation of a polymer from nonequilibrium starting conditions to equilibrium. In the example illustrated below, the nonequilibrium initial condition is the all-*trans* conformation of pentadecane, while in the final state, the pentadecane chain presents a large distribution of conformations. Denote by $\langle f(t)\rangle_{\text{noneq}}$ the expected value of some quantity $f(t)$ for a system that starts from nonequilibrium initial conditions at time $t=0$. The nonequilibrium average $\langle f(t)\rangle_{\text{noneq}}$ can be expressed in terms of probability distributions as

$$\langle f(t)\rangle_{\text{noneq}} = \int d\mathbf{r_0} P(\mathbf{r_0},0) \int d\mathbf{r} f(\mathbf{r}) P(\mathbf{r},t|\mathbf{r_0},0), \qquad (3)$$

where $P(\mathbf{r_0},0)$ denotes the probability that the coordinates of the N polymer beads at time 0 are $\mathbf{r_0}$, $P(\mathbf{r},t|\mathbf{r_0},0)$ is the formal solution to Eq. (1),

$$P(\mathbf{r},t|\mathbf{r_0},0) = e^{t\mathscr{D}(\mathbf{r})}P(\mathbf{r},0|\mathbf{r_0},0) = e^{t\mathscr{D}(\mathbf{r})}\delta(\mathbf{r}-\mathbf{r_0}), \qquad (4)$$

and $\delta(\mathbf{r}-\mathbf{r_0})$ is the Dirac delta function. Thus, Eq. (3) may be rewritten as

$$\langle f(t)\rangle_{\text{noneq}} = \int d\mathbf{r_0} P(\mathbf{r_0},0) \int d\mathbf{r} f(\mathbf{r}) e^{t\mathscr{D}(\mathbf{r})} \delta(\mathbf{r}-\mathbf{r_0})$$

$$= \int d\mathbf{r_0} P(\mathbf{r_0},0) \int d\mathbf{r} \delta(\mathbf{r}-\mathbf{r_0}) e^{t\mathscr{L}(\mathbf{r})} f(\mathbf{r})$$

$$= \int d\mathbf{r_0} P(\mathbf{r_0},0) e^{t\mathscr{L}(\mathbf{r_0})} f(\mathbf{r_0}), \qquad (5)$$

where $\mathscr{L}$ is the adjoint of $\mathscr{D}$,

$$\mathscr{L}(\mathbf{r}) = \mathscr{D}^\dagger(\mathbf{r}) = \sum_{j=1}^{3N} \left(\frac{k_B T}{\zeta} \frac{\partial^2}{\partial r_j^2} - \frac{1}{\zeta} \frac{\partial U(\mathbf{r})}{\partial r_j} \frac{\partial}{\partial r_j} \right). \qquad (6)$$

The formal expression in Eq. (5) is rendered computationally tractable by introducing the eigenfunctions Ψ_n and eigenvalues λ_n of $\mathscr{L}$,

$$\mathscr{L}\Psi_n = -\lambda_n \Psi_n, \quad n=1,2,\ldots . \tag{7}$$

Because $e^{-U/(k_B T)}\mathscr{L}$ is self-adjoint, the natural inner product is defined by

$$(\chi_i,\chi_j) = \frac{\int d\mathbf{r}\,\exp[-U(\mathbf{r})/(k_B T)]\chi_i\chi_j}{\int d\mathbf{r}\,\exp[-U(\mathbf{r})/(k_B T)]} \tag{8}$$

and coincides with the definition of the *equilibrium* average,

$$\langle \chi_i\chi_j\rangle = \frac{\int d\mathbf{r}\,\exp[-U(\mathbf{r})/(k_B T)]\chi_i\chi_j}{\int d\mathbf{r}\,\exp[-U(\mathbf{r})/(k_B T)]}. \tag{9}$$

With this choice of inner product, a complete, orthonormal set $\{\Psi_n\}$ of eigenfunctions can, in principle, be found. Thus, the $f(\mathbf{r_0})$ of Eq. (5) can be expanded in terms of the eigenfunctions $\{\Psi_n\}$ as

$$f(\mathbf{r_0}) = \sum_n (f,\Psi_n)\Psi_n(\mathbf{r_0}) = \sum_n \langle f\Psi_n\rangle \Psi_n(\mathbf{r_0}). \tag{10}$$

Substituting Eq. (10) into Eq. (5) produces

$$\langle f(t)\rangle_{\text{noneq}} = \int d\mathbf{r_0} P(\mathbf{r_0},0) e^{t\mathscr{L}(\mathbf{r_0})} \sum_n \langle f\Psi_n\rangle \Psi_n(\mathbf{r_0})$$

$$= \sum_n e^{-t\lambda_n}\langle f\Psi_n\rangle \int d\mathbf{r_0} P(\mathbf{r_0},0)\Psi_n(\mathbf{r_0}). \tag{11}$$

Since the eigenvalue Eq. (7) cannot be solved analytically for any realistic potential, the solutions are approximated using a basis set expansion,

$$\Psi_n \approx \sum_{i=1}^{M} C_{in}\phi_i. \tag{12}$$

The coefficients C_{in} are therefore determined by solving the matrix eigenvalue problem,[19]

$$\mathbf{FC}=\mathbf{SC\Lambda}, \tag{13}$$

subject to the normalization constraint,

$$\mathbf{C}^T\mathbf{SC}=\mathbf{I}, \tag{14}$$

where $\mathbf{I}$ is the unit matrix,

$$S_{ij}=\langle \phi_i\phi_j\rangle, \tag{15}$$

$$F_{ij}=\langle \phi_i\mathscr{L}\phi_j\rangle = \sum_{k=1}^{3N} -\frac{k_B T}{\zeta}\left\langle \frac{\partial\phi_i}{\partial r_k}\frac{\partial\phi_j}{\partial r_k}\right\rangle, \tag{16}$$

and $\mathbf{\Lambda}$ is the diagonal matrix of (approximate) eigenvalues $\lambda_1',\lambda_2',\ldots,\lambda_M'$. Substituting the C_{in} and λ_n' obtained from the matrix eigenvalue problem into Eq. (11) produces the final equation for $\langle f(t)\rangle_{\text{noneq}}$ as

$$\langle f(t)\rangle_{\text{noneq}} \approx \sum_{n=1}^{M}\left(\sum_{i=1}^{M}C_{in}\langle f\phi_i\rangle\right)$$

$$\times\left(\sum_{j=1}^{M}C_{jn}\int d\mathbf{r_0}P(\mathbf{r_0},0)\phi_j(\mathbf{r_0})\right)e^{-\lambda_n't}. \tag{17}$$

The nonequilibrium average in Eq. (17) is evaluated using as input only equilibrium information ($\langle f\phi_i\rangle$, $\mathbf{F}$, and $\mathbf{S}$) at the final temperature T and similar averages ($\int d\mathbf{r_0}P(\mathbf{r_0},0)\phi_j(\mathbf{r_0})$) for the initial condition. In our case, the initial condition is the all-*trans* conformation of pentadecane, and ''averaging'' over the initial condition simply involves evaluating ϕ_j for the all-*trans* conformation. However, Eq. (17) can be used for more general nonequilibrium starting conditions, such as an initial temperature T_{init} (or pressure) different from the final temperature T, in which case $\int d\mathbf{r_0}P(\mathbf{r_0},0)\phi_j(\mathbf{r_0})$ is calculated by averaging ϕ_j over the Boltzmann distribution corresponding to T_{init}. In fact, the initial condition could even correspond to a state (equilibrium or constrained) of the polymer in a different solvent.

III. BROWNIAN DYNAMICS SIMULATIONS

The BD computer simulations are more conveniently pursued by rewriting the Smoluchowski equation (1) as a set of equivalent Langevin equations,

$$\frac{dr_i(t)}{dt} = -\frac{1}{\zeta}\frac{\partial U[\mathbf{r}(t)]}{\partial r_i(t)}+X_i^*(t), \quad i=1,2,\ldots,3N, \tag{18}$$

where $X_i^*(t)$ is a Gaussian random variable with the properties,

$$\langle X_i^*(t)\rangle = 0,$$

$$\langle X_i^*(t)r_j(t')\rangle = 0 \quad \text{for } t>t', \tag{19}$$

$$\langle X_i^*(t)X_j^*(t')\rangle = 2\frac{k_B T}{\zeta}\delta_{ij}\delta(t-t'),$$

where δ_{ij} is the Kronecker delta function and $\delta(t-t')$ is the Dirac delta function. The Brownian dynamics simulations are performed using the algorithm of van Gunsteren and Berendsen,[20,21]

$$r_i(n+1)=r_i(n)+\frac{F_i(n)}{\zeta}\Delta t+\frac{1}{2}\frac{F_i'(n)}{\zeta}(\Delta t)^2+\chi_{in},$$

$$i=1,\ldots,3N, \tag{20}$$

where Δt is the time step, $F_i(n)=-\partial U(\mathbf{r})/\partial r_i$ is the force acting on the ith coordinate, $F_i'(n)=[F_i(n)-F_i(n-1)]/\Delta t$ is the finite difference approximation to the time derivative of the force, and χ_{in} is a random positional displacement taken from a Gaussian distribution with zero mean and variance $2(k_B T/\zeta)\Delta t$.

The calculation of the nonequilibrium averages $\langle f(t)\rangle_{\text{noneq}}$ from the BD simulations uses a series of trajectories run at a temperature of 300 K starting with pentadecane in the all-*trans* conformation where the molecule lies

in the $x-y$ plane with the long axis parallel to the x axis and with the first three beads having the coordinates $(0.00, 0.87, 0.00$ Å$)$, $(1.26, 0.00, 0.00$ Å$)$, and $(2.52, 0.87, 0.00$ Å$)$. Because all of the trajectories differ greatly from one another, obtaining good averages requires many trajectories, and the simulations use 9000 trajectories, each of 800 ps duration.

The theory of Sec. III calculates the equilibrium averages $\langle \cdots \rangle$ from a single 400 ns trajectory run at a temperature of 300 K. The system is allowed to equilibrate for 20 ns, after which data are collected for 380 ns.

IV. POLYMER MODEL

Pentadecane is modeled as a string of 15 CH_2 or CH_3 united atom groups. The potential U is, except for one modification, that of the GROMOS package.[22] This potential has the general form

$$U = U_{bond} + U_{angle} + U_{dihedral} + U_{nonbond}. \tag{21}$$

The bond length potential is

$$U_{bond} = \sum_{i=1}^{14} \frac{1}{2} K_b (l_i - l_0)^2, \tag{22}$$

where l_i is the length of bond i, $l_0 = 1.53$ Å is the GROMOS equilibrium bond length for alkanes, and K_b is the bond length force constant. We use the force constant $K_b = 160$ kcal mol^{-1}Å^{-2}, which is smaller than K_b in GROMOS by a factor of 5. This change does not significantly affect the long-time dynamics but allows using a larger Brownian dynamics time step ($\Delta t = 5$ fs). The bond angle potential is

$$U_{angle} = \sum_{i=1}^{13} \frac{1}{2} K_\theta (\theta_i - \theta_0)^2, \tag{23}$$

where θ_i is the ith bond angle, $\theta_0 = 111^0$ is the GROMOS equilibrium bond angle for alkanes, and $K_\theta = 110$ kcal mol^{-1} rad^{-2} is the bond angle force constant. The dihedral angle potential is

$$U_{dihedral} = \sum_{i=1}^{12} K_\phi (1 + \cos 3\phi_i), \tag{24}$$

where ϕ_i is the ith dihedral angle, and $K_\phi = 1.4$ kcal/mol. The nonbonded potential $U_{nonbond}$ uses a 6-12 Lennard-Jones potential to describe the interaction between two nonbonded united atoms. More details are given in Ref. 22.

The solvent is modeled as a structureless continuum that exerts a viscous damping force on the pentadecane chain. The friction coefficient is calculated using Stokes' law $\zeta = 6 \pi \eta R$, where the viscosity η is taken to be 1 cp and the hydrodynamic radius of a CH_2 or CH_3 united atom group is taken to be 1.5Å.

V. RESULTS

The only approximation in the theory is associated with the use of a finite basis set in Eq. (12). Thus, we begin by studying the effectiveness of various basis sets for estimating the time evolution of the mean-square end-to-end distance $R_{end}^2 = (x_{15} - x_1)^2 + (y_{15} - y_1)^2 + (z_{15} - z_1)^2$. In contrast to the previously studied equilibrium time correlation functions of dynamical quantities that are odd in the position coordinates,[1-5] the mean-square end-to-end distance is an even function of the position coordinates. Therefore, symmetry considerations dictate the use of different (even-power, as opposed to odd-power) basis functions in the mode-coupling theory calculation of $\langle R_{end}^2(t) \rangle_{noneq}$. The scalar nature of $R_{end}^2(t)$ implies that a general mode coupling basis may be constructed from the basis functions[23]

$$1,$$

$$\mathbf{l_i} \cdot \mathbf{l_j} \quad [\forall \text{ (distinct) pairs } \{i,j\}, \quad i,j = 1, \dots, 14],$$

$$(\mathbf{l_i} \cdot \mathbf{l_j})(\mathbf{l_k} \cdot \mathbf{l_m}) \quad [\forall \text{ (distinct) pairs of (distinct) pairs}$$

$$\{\{i,j\}, \{k,m\}\}, \quad i,j,k,m = 1, \dots, 14],$$

$$(\mathbf{l_i} \cdot \mathbf{l_j})(\mathbf{l_k} \cdot \mathbf{l_m})(\mathbf{l_n} \cdot \mathbf{l_p}) \quad [\forall \text{ (distinct) triples of (distinct) pairs,}$$

$$\{\{i,j\}, \{k,m\}, \{n,p\}\}, \quad i,j,k,m,n,p = 1, \dots, 14], \text{ etc.,}$$

where $\mathbf{l_i}$ denotes the ith pentadecane bond vector with components $l_{ix} = x_{i+1} - x_i$, $l_{iy} = y_{i+1} - y_i$, and $l_{iz} = z_{i+1} - z_i$, for $i = 1, \dots, 14$. The ordering of pairs (and triples, etc.) is unimportant; for example, the pair $\{1,2\}$ is equivalent to $\{2,1\}$ since $\mathbf{l_1} \cdot \mathbf{l_2} = \mathbf{l_2} \cdot \mathbf{l_1}$, and hence only one of them is included in the basis set. The constant basis function 1, which produces a vanishing eigenvalue in Eq. (13), accounts for the nonzero value of $\langle R_{end}^2(t) \rangle_{noneq}$ at infinite time.

The mode coupling theory is first applied using the relatively simple basis set,

$$\{1, \mathbf{l_i} \cdot \mathbf{l_j} \quad [\forall \text{ (distinct) pairs } \{i,j\},$$

$$i,j = 1, \dots, 14]\}, \quad \text{(basis set I)}.$$

Figure 1 demonstrates that basis set I provides a reasonable estimate for the time evolution of $\langle R_{end}^2(t) \rangle_{noneq}$ whose numerically ''exact'' dynamics is represented by the BD simulation. A more accurate approximation of $\langle R_{end}^2(t) \rangle_{noneq}$ requires expanding basis set I to include more basis functions of the appropriate symmetry, such as the tetralinear products $(\mathbf{l_i} \cdot \mathbf{l_j})(\mathbf{l_k} \cdot \mathbf{l_m})$. However, the total number of tetralinear products scales as N^4, where N is the number of monomers. Thus, inclusion of all tetralinear functions makes the basis set unmanageably large even for the relatively small ''polymer'' pentadecane. Clearly, all the tetralinear products cannot generally be retained in the basis set, nor are all these basis functions relevant to the long time dynamics. Those tetralinear products that contribute the most to the long time dynamics are determined using a selection technique similar to that introduced previously for computing equilibrium time correlation functions.[3-5]

Based on prior treatments of equilibrium time correlation functions, it proves convenient to work with basis functions constructed as linear combinations of individual bond
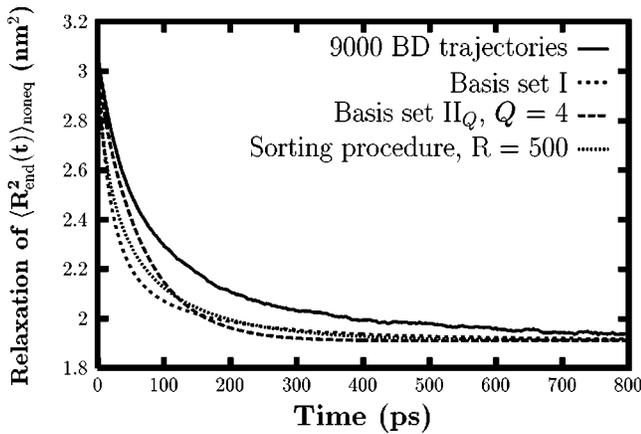
FIG. 1. The relaxation of the mean-square end-to-end distance $\langle R_{\text{end}}^2(t) \rangle_{\text{noneq}}$ of pentadecane to equilibrium from the all-*trans* conformation. The solid line represents the ''exact'' result calculated from the average of 9000 non-equilibrium BD simulations, while the various dashed lines represent the results of mode coupling theory calculations using different basis sets, as described in the text.

vectors that provide a first approximation to the collective chain motions. An approximate set of eigenfunctions, denoted as the first-order GR (generalized Rouse) modes of $\mathscr{L}$ may be obtained from Eqs. (12) to (16) with a basis set consisting of all linear functions in the bond vectors,

$$\{l_{ix} \;\; [i=1,\ldots,14], \quad l_{iy} \;\; [i=1,\ldots,14],$$

$$l_{iz} \;\; [i=1,\ldots,14]\}.$$

The first-order GR modes are thus linear combinations of the bond vector components. Note that the linear basis functions and the corresponding first-order GR modes do not contribute to the expression in Eq. (17) for $\langle R_{\text{end}}^2(t) \rangle_{\text{noneq}}$ due to symmetry considerations. Products of first-order GR modes, however, may be used as providing a first approximation to $\langle R_{\text{end}}^2(t) \rangle_{\text{noneq}}$. Due to the invariance of $\mathscr{L}$ under interchange of the Cartesian directions x, y, and z, the first-order GR modes need only be computed for one direction, say x, and the GR modes in the other two directions follow trivially by symmetry. Let $\Psi_{1x}^{(1)}, \Psi_{2x}^{(1)}, \ldots, \Psi_{14x}^{(1)}$ be the first-order GR modes obtained from solving Eqs. (12) to (16), and let their corresponding (approximate) eigenvalues $\lambda_1^{(1)}, \lambda_2^{(1)}, \ldots, \lambda_{14}^{(1)}$ be ordered in increasing magnitude. Thus, the first eigenfunction $\Psi_{1x}^{(1)}$ provides the first-order estimate to the GR eigenfunction with the slowest decay, called the slowest mode, $\Psi_{2x}^{(1)}$ provides the first-order estimate to the GR eigenfunction of the next slowest decaying mode, and so on. Let $\Psi_{1y}^{(1)}, \Psi_{2y}^{(1)}, \ldots, \Psi_{14y}^{(1)}$ be the GR modes obtained by replacing all the x's in $\Psi_{1x}^{(1)}, \Psi_{2x}^{(1)}, \ldots, \Psi_{(N-1)x}^{(1)}$ with y's, and let $\Psi_{1z}^{(1)}, \Psi_{2z}^{(1)}, \ldots, \Psi_{14z}^{(1)}$ be the corresponding GR modes in z. It is convenient to introduce the notation $\Psi_i^{(1)} \; [i=1,2,\ldots,14]$ to represent the vector $(\Psi_{ix}^{(1)}, \Psi_{iy}^{(1)}, \Psi_{iz}^{(1)})$. We expect that the slowest decaying first-order eigenfunctions contain the most information about the long-time dynamics. Thus, rather than including all possible tetralinear products in the basis set, we only include

tetralinear products constructed from the slowest decaying GR modes. This produces the basis set which contains the functions[24]

$$\{1, \Psi_i^{(1)} \cdot \Psi_j^{(1)} \;\; [\forall \;\; (\text{distinct}) \;\; \text{pairs} \;\; \{i,j\}, i,j=1,\ldots,14],$$

$$(\Psi_i^{(1)} \cdot \Psi_j^{(1)})(\Psi_k^{(1)} \cdot \Psi_m^{(1)})$$

$$[\forall \;\; (\text{distinct}) \;\; \text{pairs of} \;\; (\text{distinct}) \;\; \text{pairs}$$

$$\{\{i,j\},\{k,m\}\}, i,j=1,\ldots,Q]\},$$

$$(\text{basis set II}_Q),$$

where $0 \leq Q \leq 14$. The $Q=0$ case reduces basis set $\text{II}_{Q=0}$ to basis set I, while using $Q=14$ is equivalent to including all possible tetralinear products in the basis set. Note that the size of basis set II_Q scales roughly as $N^2 + Q^4$. Hence, basis set II_Q is significantly smaller than the full basis set containing all possible tetralinear products if $Q \ll N$. Figure 1 demonstrates that basis set II_Q with $Q=4$ produces an improvement in the long time dynamics predicted by the theory, though for short times below 100 ps, basis set I is in better agreement with the simulations. Increasing Q beyond 4 does not produce any significant improvement. Greater accuracy of the theory therefore requires the inclusion of hexalinear and higher order products in the basis set. However, the size of the basis set grows extremely rapidly as higher order functions are included, even when a procedure similar to that in basis set II_Q is used. In order to select those higher order functions that contribute the most to the long time dynamics, we now test an alternative sorting procedure first described in Ref. 4 for treating the long time dynamics of small flexible peptides.

The sorting procedure orders basis functions according to the sum of the (approximate) eigenvalues of the GR modes in the tetralinear and higher order (even) products of GR modes. For example, $\lambda_i^{(1)} + \lambda_j^{(1)} + \lambda_k^{(1)} + \lambda_m^{(1)}$ represents the first order relaxation rate for the tetralinear product $(\Psi_i^{(1)} \cdot \Psi_j^{(1)})(\Psi_k^{(1)} \cdot \Psi_m^{(1)})$. Since we expect that the functions most relevant to the long time dynamics are composed of the slowest-decaying GR modes (smallest approximate eigenvalues), the basis set retains the products with the smallest eigenvalue sums. While first-order eigenvalues may not provide an optimal estimate for the relative importance of different products of GR modes, previous experience suggests that this method is sufficient for obtaining excellent equilibrium time correlation functions.[4,5] The first-order eigenvalue sorting procedure produces the basis set which is represented as

$$\{1, \Psi_i^{(1)} \cdot \Psi_j^{(1)} \;\; [\forall \;\; (\text{distinct}) \;\; \text{pairs} \;\; \{i,j\}, \quad i,j=1,\ldots,14],$$

higher order GR mode products (up to 14 products) with the R smallest eigenvalue sums$\}$,

$$(\text{basis set III}_R).$$

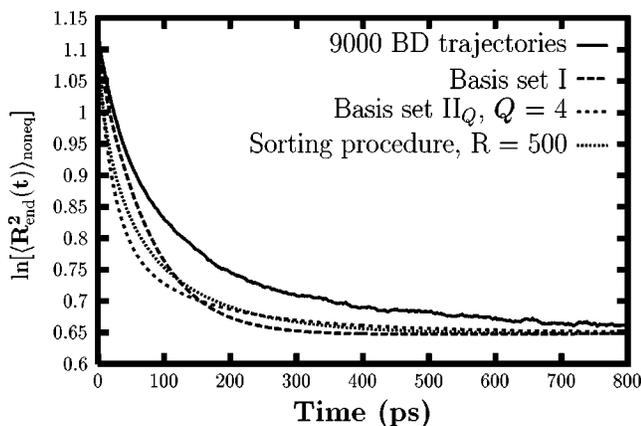Figure 1 indicates that basis set III_R with $R=500$ produces further improvement in the theory. Increasing R beyond 500

FIG. 2. Same as Fig. 1, except that the $\langle R_{end}^2(t) \rangle_{noneq}$ axis has a logarithmic scale.

results in numerical instabilities in the calculation, though previous experience indicates that using longer equilibrium simulations rectifies the problem.[4,5]

The discrepancy in the ''infinite'' time R_{end}^2 in Fig. 1 is due to the difference between the ''infinite'' time (i.e., equilibrium) $\langle R_{end}^2(t) \rangle_{noneq}$ from the 9000 nonequilibrium BD simulations and the $\langle R_{end}^2 \rangle$ from the 400 ns equilibrium BD simulation. This difference arises from statistical fluctuations in calculating averages, and the magnitude of these fluctuations can be estimated by a simple calculation. The variation in R_{end}^2 is comparable to the size of R_{end}^2 itself—about 2 nm^2. In a BD simulation, the ''memory'' of previous values of R_{end}^2 is about 100 ps (see Fig. 1), so a 400 ns BD simulation yields about 4000 independent samples of R_{end}^2. Because R_{end}^2 is *not* normally distributed, the central limit theorem is not strictly applicable.[25] Nevertheless, the theorem can provide an estimate for the order of magnitude of the fluctuations in calculating the average $\langle R_{end}^2 \rangle$ as $2 \, \text{nm}^2/\sqrt{4000} \approx 0.03 \, \text{nm}^2$. This value is consistent with the discrepancy in Fig. 1 and is also consistent with the variation in the equilibrium average $\langle R_{end}^2 \rangle$ between two different 400 ns BD simulations (data not shown).

Figure 2 demonstrates that the relaxation of R_{end}^2 to equilibrium is highly nonexponential at long times. This is in contrast to equilibrium dipole time correlation functions, where the exponential decay at long times is predicted very accurately by the mode coupling theory.[4] The nonexponential behavior of $\langle R_{end}^2(t) \rangle_{noneq}$ at long times may account for the difficulty in obtaining highly accurate calculations from the theory. Nevertheless, it is encouraging that the mode coupling theory gives reasonable estimates for $\langle R_{end}^2(t) \rangle_{noneq}$, and, furthermore, correctly predicts that the relaxation at long times is nonexponential.

VI. DISCUSSION

The computational effort required to calculate $\langle R_{end}^2(t) \rangle_{noneq}$ directly from Brownian dynamics simulations is enormous and requires orders of magnitude more computer time than is necessary for simulating equilibrium time correlation functions. The absence of time translational symmetry during a nonequilibrium relaxation precludes the com-

putation of this process from a single Brownian dynamics trajectory. Instead, an ensemble average must be performed over a *huge* number of nonequilibrium trajectories. For example, 9000 trajectories are necessary to obtain with sufficiently small noise the relaxation to equilibrium of $\langle R_{end}^2(t) \rangle_{noneq}$ for pentadecane from the all-*trans* conformation. The accumulation of these nonequilibrium trajectories requires more than one month on an SGI Power Challenge. Moreover, this simulation is far simpler than that needed for most nonequilibrium processes. The direct simulation of a T-jump process would necessitate averaging over the distribution of both initial and final conditions and would require at least an order of magnitude more than the 9000 trajectories used for the simulation with constrained initial conditions. Clearly, any realistic study of the unfolding of a protein by computer simulations is currently unfeasible and will remain so even in the more distant future. In contrast, the theory requires as inputs only *equilibrium* averages which may be obtained readily from a few Brownian dynamics trajectories. Furthermore, the theory can easily handle many different types of nonequilibrium processes. Treating the relaxation of pentadecane from the all-*trans* conformation requires only equilibrium averages for the final state. To describe a T-jump process using the theory, equilibrium averages must be computed for both the initial state and the final state—merely a doubling of the computational effort.

The successful description by the theory of the ''unfolding'' of pentadecane indicates that the theory proposed here may provide a significant contribution to the study of protein folding and unfolding. Understanding how a globular protein folds requires detailed characterization of partially organized intermediates formed during the folding process. Information about such intermediates is becoming available through many complementary experimental studies involving stable partially folded states, protection from hydrogen exchange, disulfide formation in oxidative refolding, effects of amino acid substitutions, and peptide analogues of folding intermediates.[26] Nonetheless, no clear unifying view of the nature of folding mechanisms has yet emerged, and many uncertainties over fundamental issues remain.

One reason that protein folding remains an unsolved problem is that the experimental techniques used to study this process do not provide structural information with atomic detail about the multitude of folding pathways and intermediates. Instead the folding process is generally characterized experimentally by more global properties, such as the degree of compactness of the protein, the percentage of secondary structure compared to the native state, different hydrogen exchange rates of various residues, etc. The real time observation of protein folding with, for example, NMR, circular dichroism, or fluorescence techniques, also provides information about coarse-grained features of the folding intermediates and pathways with limited spatial resolution. Consequently, the folding process is frequently analyzed in terms of the free energy as a function of a heuristic one-dimensional reaction coordinate which is not directly derived from microscopic motions and thus only reflects global qualitative features of the folding process. Computer simulations provide limited information as well because only short

time scales are accessible to realistic simulations. Furthermore, the large amount of computation required by such simulations allows the study of only one (or at best a few) trajectories from which it is difficult to extract general characteristics of the folding process.

The theory presented here potentially lacks many of the limitations of experiments and computer simulations. In principle, the theory can predict the time evolution of any interatomic distance in a polypeptide and can thus specify the average structure of the polypeptide at any time during the folding or unfolding process. However, the practical application of the theory to protein folding may emerge only after further testing and development. For example, presently it is unclear how the accuracy of the theoretical predictions scales with system size. Therefore, it is desirable to test the theory for systems larger than pentadecane. A natural choice would be the helix-coil transition in a simple polyalanine helix since the computational intensity of nonequilibrium simulations for a small helix is not significantly greater than that for pentadecane. Thus, nonequilibrium BD simulations are feasible for the unfolding of the small helical peptide, thereby again enabling no-parameter tests of the theory before applying the theory to even larger systems for which the nonequilibrium simulations are prohibitive. The recent treatment for the more complicated equilibrium dynamics of small peptides with the mode coupling theory lends support to the belief that the nonequilibrium generalization presented here is adequate for the task of describing the unfolding of a small protein.

A possible direct experimental test of the theory is provided by the work of Perkins and co-workers.[27] They measure the relaxation of a single DNA molecule from a highly stretched nonequilibrium state to equilibrium using optical tweezers to hold the DNA molecule in place and fluorescence labeling to follow the motion of the DNA molecule. The relaxation of the end-to-end vector is highly nonexponential, with a fast initial decay followed by a slower decay. The simple Rouse and Rouse–Zimm models do not fully explain the data; this is not surprising since modeling DNA as a string of beads connected by springs is clearly an oversimplification. The mode coupling theory provides a framework for treating DNA more realistically. Much work remains to be done, however, before a direct comparison can be made between the experimental data of Perkins and co-workers and the predictions of the mode coupling theory. The extension of the theory to include the solvent would be an important first step, and the treatment of hydrodynamic interactions is possible following Ref. 14. Also, the applicability of the theory to highly charged polymers, such as DNA, must be tested. Nevertheless, the predictive power of the mode coupling theory for describing the dynamics of a variety of polymers suggests that such generalizations would be fruitful.

[1] X. Y. Chang and K. F. Freed, J. Chem. Phys. **99**, 8016 (1993).
[2] X. Y. Chang and K. F. Freed, Chem. Eng. Sci. **49**, 2821 (1994).
[3] W. H. Tang, X. Y. Chang, and K. F. Freed, J. Chem. Phys. **103**, 9492 (1995).
[4] K. S. Kostov and K. F. Freed, J. Chem. Phys. **106**, 771 (1997).
[5] K. S. Kostov, K. F. Freed, E. B. Webb III, M. Mondello, and G. S. Grest, J. Chem. Phys. (in press).
[6] R. Zwanzig, J. Chem. Phys. **60**, 2717 (1974).
[7] M. Bixon and R. Zwanzig, J. Chem. Phys. **68**, 1896 (1978).
[8] A. Perico, Acc. Chem. Res. **22**, 336 (1989).
[9] A. Perico, S. Bisio, and C. Cuniberti, Macromolecules **17**, 2686 (1984).
[10] A. Perico and M. Guenza, J. Chem. Phys. **83**, 3103 (1985).
[11] A. Perico and M. Guenza, J. Chem. Phys. **84**, 510 (1986).
[12] A. Perico, J. Chem. Phys. **88**, 3996 (1988).
[13] A. Perico, F. Ganazzoli, and G. Allegra, J. Chem. Phys. **87**, 3677 (1987).
[14] Y. Hu, K. Kostov, A. Perico, S. Smithline, and K. F. Freed, J. Chem. Phys. **103**, 9091 (1995).
[15] Y. Hu, G. R. Fleming, K. F. Freed, and A. Perico, Chem. Phys. **158**, 395 (1991).
[16] M. Doi and S. F. Edwards, *The Theory of Polymer Dynamics* (Clarendon, Oxford, 1986).
[17] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (Springer, New York, 1990).
[18] H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications* (Springer, New York, 1989).
[19] E. Anderson, *et al.*, *LAPACK Users' Guide* (SIAM, Philadelphia, 1995).
[20] W. F. van Gunsteren and H. J. C. Berendsen, Mol. Phys. **45**, 637 (1982).
[21] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1987).
[22] W. F. van Gunsteren and H. J. C. Berendsen, *GROMOS Manual* (University of Groningen, Groningen, 1987).
[23] The symmetry requirements for calculating $\langle R_{end}^2(t) \rangle_{noneq}$ are the same as for some of the terms for the orientational time correlation function $P_2(t)$, as described in Ref. 3.
[24] Note that $\{\boldsymbol{\psi}_i^{(1)} \cdot \boldsymbol{\psi}_j^{(1)}$ [$\forall$ (unordered) pairs $\{i,j\}$, $i,j = 1, \ldots, 14]\}$ is equivalent to $\{\mathbf{l_i} \cdot \mathbf{l_j}$ [$\forall$ (unordered) pairs $\{i,j\}$, $i,j = 1, \ldots, 14]\}$ since both are the full bilinear sets.
[25] L. Ott, *An Introduction to Statistical Methods and Data Analysis* (Duxbury, Boston, 1984).
[26] A. D. Miranker and C. M. Dobson, Curr. Opin. Struct. Biol. **6**, 31 (1996).
[27] T. T. Perkins, S. R. Quake, D. E. Smith, and S. Chu, Science **264**, 822 (1994).