

Long-Time Dynamics of Met-Enkephalin: Comparison of Theory with Brownian Dynamics Simulations

Konstantin S. Kostov* and Karl F. Freed[#]

James Franck Institute and the Departments of *Biochemistry and Molecular Biology and [#]Chemistry, The University of Chicago, Chicago, Illinois 60637 USA

ABSTRACT A recent theory for the long time dynamics of flexible chain molecules is applied for the first time to a peptide of biological importance, the neurotransmitter met-enkephalin. The dynamics of met-enkephalin is considerably more complicated than that of the previously studied glycine oligomers; met-enkephalin contains the interesting motions of phenyl groups and of side chains relative to the backbone, motions that are present in general flexible peptides. The theory extends the generalized Rouse (GR) model used to study the dynamics of polymers by providing a systematic procedure for including the contributions from the memory function matrices neglected in the GR theory. The new method describes the dynamics by time correlation functions instead of individual trajectories. These correlation functions are analytically expressed in terms of a set of equilibrium averages and the eigenvalues and eigenfunctions of the diffusion operator. The predictions of the theory are compared with Brownian dynamics (BD) simulations, so that both theory and simulation use identical potential functions and solvent models. The theory thus contains no adjustable parameters. Inclusion of the memory function contributions profoundly affects the dynamics. The theory produces very good agreement with the BD simulations for the global motions of met-enkephalin. It also correctly predicts the long-time relaxation rate for local motions.

INTRODUCTION

The dynamics of proteins and other biomolecules are essential for their biological function (Subbiah, 1996). The theoretical investigation of this dynamics is very challenging because biologically relevant motions occur on a wide variety of time scales—from tens of picoseconds for fluctuations around equilibrium positions to seconds for protein folding. For example, the widely used molecular dynamics (MD) simulations (Brooks et al., 1988) are currently limited to the time domain of nanoseconds at most. Realistic MD simulations of biomolecules in aqueous solution involve tens of thousands of atoms and become prohibitively expensive for longer times. Despite the continual rapid increase in computer power and technical advances in simulation algorithms (Zhang and Schlick, 1994; Figueirido et al., 1997), MD simulations reaching biologically relevant times cannot be expected in the near future. In addition, such simulations would generate an overwhelming amount of information, the analysis of which would present new problems of its own. These limitations emphasize the need for theoretical advances permitting the description of the long-time dynamics in biological systems by statistical mechanical methods.

We have been developing a theory for predicting and characterizing the *long-time* dynamics of peptides and proteins under conditions where one or a few configurations are

insufficient for describing the range of conformational states accessed during the dynamics. There are numerous instances when such flexibility is important for biological function. For example, although most proteins in their native state are compact globular objects, many proteins have rather flexible portions that may have physiological significance. A recently discovered example is the structure of the prion protein PrP^C, which contains a 98-residue amino-terminal segment that exhibits features of a flexible random coil (Riek et al., 1997). This flexible coil may enable structural transitions to PrP^{Sc} aggregates. Even more striking is the absence of a stable secondary and tertiary structure in the native state of the protein p21 (involved in cell cycle control) when it is not bound to its ligand(s) (Kriwacki et al., 1996). Denatured proteins, on the other hand, are totally flexible and may roam through an enormous number of different conformations during their dynamical motion and during the process of protein folding. Recent experiments indicate that a compact protein state is achieved only late in the folding process (Sosnick et al., 1996; Smith and Dobson, 1996). Likewise, small peptides are generally not characterized by single native structures, and the accessibility of several different low-energy conformations may contribute to their ability to bind to several different receptors and thereby exhibit several different physiological functions. An example with biomedical significance is the amyloid β -peptides involved in the plaque formation occurring with Alzheimer's disease. These peptides do not have a dominant conformation in solution and instead preferentially adopt random-coil, α -helix, or β -sheet (the latter promotes plaque formation) structures, depending on the temperature, solvent conditions, and pH (Barrow and Zagorski, 1991; Kirshenbaum and Daggett, 1995). These examples, as well as many others (see Discussion), demon-

Received for publication 11 May 1998 and in final form 28 August 1998.

Address reprint requests to Dr. Konstantin S. Kostov, Department of Chemistry, University of Chicago, 5735 S. Ellis Avenue, Room 117, Chicago, IL 60637. Tel.: 773-702-7052; Fax: 773-702-0805; E-mail: kostov@rainbow.uchicago.edu.

© 1999 by the Biophysical Society

0006-3495/99/01/149/15 \$2.00

strate the need for understanding the dynamics of flexible biomolecules and also show that the currently available experimental and theoretical methods are insufficient to provide such an understanding. Therefore, the goal of our theory is to remedy this deficiency by 1) developing a general method for characterizing the motion of flexible biomolecules and 2) predicting this motion on the long time scales that are pertinent to biological functions.

As noted above, MD simulations are well developed for following the detailed trajectory of a biological system for short times and modest system sizes. However, when applied to a flexible chain molecule that may traverse many different conformational states during a single trajectory and that may exhibit very disparate sequences of conformational states between different trajectories, the MD methods are clearly limited in their descriptive powers. A statistical treatment of the dynamics therefore appears more appropriate. Statistical mechanics provides a complementary method to MD simulations for probing long-time dynamics. Rather than following a large number of individual trajectories (or a very long single trajectory) and averaging over their (its) behavior, the alternative method, from the outset, studies the dynamics as averaged over a whole equilibrium ensemble of systems. Such a statistical description is provided by various time correlation functions, which may be obtained, in principle, from averaging one long or several shorter MD trajectories but which emerge naturally from a statistical mechanical theory. Examples include the $P_2(t)$ time correlation functions measured by NMR and fluorescence depolarization methods. However, these methods provide information only on the time scale for local bond and transition dipole reorientations, respectively. Thus, they do not directly probe the slower more global relative motions of different groups, e.g., the two domains forming the scissors motion in immunoglobulins. Hence, we recognize the need for assessing which correlation functions can characterize the slow large-scale global motion of flexible denatured proteins, the relative motions of protein domains, and the flexible motions of peptides, as well as the need for developing a theory to predict these motions. A predictive analytical theory is distinguished from MD simulations, which merely implement Newton's equations numerically. Because our goal lies in describing motions on time scales that are long compared with those accessible to MD simulations, it is clear that a predictive theory must contain fundamental conceptual advances in the statistical description of long-time dynamics.

Our theory extends to flexible chain molecules the projection operator techniques developed by Mori and Zwanzig (Mori, 1965; Zwanzig, 1961), which provide a formally exact reduced description of the dynamics of many-body systems. Zwanzig and Mori show that by identifying a suitable set of slow variables the original full Liouville equation of motion can be rewritten in the form of a generalized Langevin equation (GLE), which contains explic-

itly only the small set of N slow variables $\mathbf{r}(t)$,

$$\frac{d\mathbf{r}(t)}{dt} = \Lambda_0 \mathbf{r}(t) - \int_0^t d\tau \Phi(\tau) \mathbf{r}(t - \tau) + \mathbf{f}_r(t), \quad (1)$$

with $\mathbf{f}_r(t)$ the random forces, Λ_0 a frequency matrix, and $\Phi(\tau)$ the memory function matrix. The apparent simplicity of the GLE is deceiving; the complexity of the many-body problem is hidden in the properties of the memory functions and random forces. Nevertheless, when a clear time scale separation exists between the slow and fast degrees of freedom, the solution of the GLE can be simplified considerably by invoking a Markovian approximation for the memory functions, i.e., that the memory functions decay very rapidly in time and therefore $\Phi(\tau) \propto \Gamma \delta(\tau)$, where Γ is a relaxation matrix, and $\delta(t)$ is the Dirac delta function. This has allowed the successful application of the GLE in many areas of condensed-matter physics, such as the dynamics of simple fluids.

The application of the GLE to the description of the dynamics of flexible chain molecules is complicated by the fact that no clear separation of time scales exists for these systems. The lack of a time scale separation, combined with the connectivity and flexibility of the polymer chain, leads to a whole set of challenging problems. The first problem is the choice of the slow variables themselves. A second problem is the treatment of polymer-solvent interactions, a problem that can be tackled by introducing a hydrodynamic model for the solvent (Metiu et al., 1977) and, perhaps, by including an explicit shell of solvent molecules in the set of slow variables. The hydrodynamic model describes the influence of the solvent through hydrodynamic interactions and through friction coefficients for atoms or groups of atoms in the polymer chain. The final challenge arises because the GLE describing polymer dynamics is inherently multidimensional, and therefore the memory functions are $N \times N$ matrices with poorly understood properties. The theoretical treatment of these matrices is very complicated as they contain both intermolecular and intramolecular contributions.

The first approach to these problems is due to Zwanzig and Bixon (Zwanzig, 1974; Bixon and Zwanzig, 1978). They begin from the diffusion equation with hydrodynamic interactions and project onto the slow variables chosen as the linear set of bead positions in the chain. With this choice of slow variables and a Gaussian bead-spring model with preaveraging of the hydrodynamic interactions the memory function terms become zero and the GLE reduces to the optimized or generalized Rouse-Zimm (GRZ) equations of motion. The GRZ equations extend the classical Rouse-Zimm theory for polymer dynamics by including non-nearest-neighbor interactions. Perico and co-workers (Perico, 1989; Guenza and Perico, 1993) have demonstrated that the non-nearest-neighbor contributions very substantially influence both the local and global dynamics of synthetic polymers. Moreover, they show that the time correlation func-

tions describing the dynamics can be derived exactly within the framework of the GRZ model and that the evaluation of these dynamical quantities requires as input only equilibrium information and bead friction coefficients (Perico and Guenza, 1985). However, for any realistic interaction potentials beyond the simple bead-spring model, the memory function terms no longer vanish, and due to the enormous difficulties of their calculation, these terms have been customarily neglected in the earlier treatments.

More recently, the GRZ approach has been applied to peptides containing a single tryptophane residue, glucagon, the hormone ACTH, and their fragments (Chen et al., 1987; Hu et al., 1990, 1991, 1995). The equilibrium averages required as inputs to the dynamical theory are obtained from realistic potentials, either with rotational isomeric state calculations, using the ECEPP potentials developed by Scheraga and co-workers (Li and Scheraga, 1987) or from MD simulations with the CHARMM package. The theoretical predictions are in general qualitative, and in many cases even quantitative, agreement with experimental measurements of the decay of fluorescence anisotropy of the tryptophane residue. Although this agreement is very encouraging, the neglect of the memory functions in this work, combined with the other approximations introduced in the GLE that are described above, implies that the agreement may possibly result from a fortuitous cancellation of errors. For example, it is possible to assess indirectly the significance of the memory function terms by extending in a hierarchical fashion the set of slow variables from the set of virtual bonds in the peptide to include all backbone bonds and finally the side chain bonds. Such an expansion of the slow variable set indeed shows that the memory functions exert a significant influence on the dynamics of the ACTH(6-10) fragment and that the best agreement with experiment is obtained only when all backbone and side chain bonds are included in the set of slow variables. Hence, it is obvious that additional work is required to test the significance of the individual approximations to the GLE and thereby to instill confidence in the predictive capabilities of the GRZ model and its extensions.

The complexity of these problems is best approached by separately testing each approximation in a manner that is independent of the other approximations involved. The largest and least studied of these approximations is the neglect of the memory functions in the GRZ equations. Our recent work (Perico et al., 1993; Tang et al., 1995; Kostov and Freed, 1997) has designed an approach that evaluates only the contributions of the memory functions to the dynamics predicted by the GRZ equations. The point of departure is the diffusion equation with hydrodynamic interactions neglected. The approach of Zwanzig and Bixon applied to such a diffusion equation produces a generalized Rouse (GR) model. This diffusion equation is solved alternatively in two ways: 1) numerically by Brownian dynamics (BD) simulations and 2) analytically with a theory described below. Both the theory and the simulations use the same friction coefficients and identical, realistic potential func-

tions, and hence the theory contains no adjustable parameters. These features, as well as the absence of an approximate hydrodynamic model for the solvent in both theory and simulations, allow the unambiguous testing of the significance of the memory functions on the dynamics.

The dynamics of the polymer are described by time correlation functions (which can be obtained from the theory for arbitrary long times). The theory expresses the time correlation functions in terms of a set of equilibrium averages and the eigenvalues and eigenfunctions of the adjoint of the diffusion operator. The eigenvalues are found by expanding the eigenfunctions in a suitable basis set and by solving the resulting generalized eigenvalue problem. We have developed a procedure for the choice of the basis functions and have shown that this choice corresponds to including the contributions from the previously neglected memory functions (Tang et al., 1995; Kostov and Freed, 1997). The calculations proceed in two steps. In the first step, the necessary equilibrium averages are obtained from the BD simulations. In the second step, the generalized eigenvalue problem is solved, and the correlation functions obtained from the theory are compared with those computed from the BD simulations. The simulations are considered to provide the numerically exact correlation functions as the simulated trajectories are quite long and therefore the statistical error is small.

The selection of a suitable basis set is the most important step of both projection operator methods and our theory. If a basis is chosen as the individual atomic positions (first-order basis), the theory reduces to the generalized Rouse (GR) theory, which completely neglects the memory functions. Adding more functions in the basis set includes contributions from the memory function matrix and therefore produces a more accurate description of the internal dynamics. In the spirit of mode-coupling theory for the treatment of long-time critical dynamics (Kawazaki, 1970; Kadanoff and Swift, 1968; Kapral et al., 1976), the additional basis functions have been taken as products of the normal modes (GR modes) obtained from the first-order basis set. Thus, we refer to the theory as the mode-coupling theory, which is a more descriptive name than the initially used term matrix expansion method.

The size of the mode-coupling basis sets, however, increases dramatically with system size, as the number of possible products of the GR modes is virtually unlimited even for small peptides. For example, the automatic inclusion of all possible trilinear and pentilinear products, used initially for alkanes, yields basis sets the size of which increases as N^3 and N^5 , respectively, where N is the number of atoms. This explosive growth in the number of basis functions requires the introduction of a selection procedure that includes in the basis set only those modes that contribute most to the long-time dynamics. Such a procedure was described and tested in a previous paper studying the dynamics of glycine oligomers (Kostov and Freed, 1997). A large number of high-order products of the normal modes is sorted according to their first-order relaxation times, and

only a small subset consisting of the slowest decaying terms is included in the basis set. The use of such a selection procedure is motivated by physical considerations; the slowest decaying modes are anticipated to provide the most significant contributions to the long-time dynamics. A basis set selected in this manner has been demonstrated to yield quantitative agreement with the BD simulations for all global and many local bonds in triglycine and octaglycine. This basis set scales nearly linearly with system size and leads to a considerable reduction in the number of basis functions needed to provide an accurate description of the long-time internal dynamics of peptides. The new sorting procedure therefore allows the application of the theory to the long-time dynamics of larger, biologically relevant peptides.

To further test and refine the theory, in this paper we consider the internal dynamics of the neurotransmitter met-enkephalin, which is displayed in Fig. 1. Met-enkephalin is chosen to provide a very stringent test of the theory as its flexibility is combined with a considerable degree of rigidity due to its small size and the presence of two aromatic rings. Moreover, the dynamics contains the interesting motions of phenyl groups and of side chains relative to the backbone, motions that are present in general flexible peptides. In addition, met-enkephalin has been well studied both experimentally and theoretically because of its biological importance (Li and Scheraga, 1987; Vasquez et al., 1994; Montcalm et al., 1994; van der Spoel and Berendsen, 1997; Benham and Deber, 1984; Anteunis et al., 1977).

We first briefly review the basic theory and then present the details of the Brownian dynamics simulations employed for calculating the time correlation functions and the equilibrium averages and describe the sorting procedure used to

select the basis set. The results are presented and discussed in the last two sections, which provide illustrations for both global and local motions, with the latter including various phenyl group motions. The Discussion considers the biological relevance of our computations and of peptide flexibility.

THEORY

The basic theory is described in detail elsewhere (Perico et al., 1993; Chang and Freed, 1993; Tang et al., 1995). The following summary introduces necessary notation and concepts. Let $\mathbf{r}(t)$ represent the $3N$ Cartesian coordinates of the N peptide atoms $\mathbf{r}(t) = \{\mathbf{r}_1(t), \mathbf{r}_2(t), \dots, \mathbf{r}_N(t)\}$. The starting point is the diffusion equation with hydrodynamic interactions neglected:

$$\frac{\partial \Psi(\mathbf{r}, t | \mathbf{r}_0)}{\partial t} = \mathcal{D} \Psi(\mathbf{r}, t | \mathbf{r}_0), \quad (2)$$

$$\mathcal{D} = \sum_{i=1}^N \frac{k_B T}{\zeta_i} [\nabla_{\mathbf{r}_i}^2 + \beta \nabla_{\mathbf{r}_i} \cdot \nabla_{\mathbf{r}_i} U(\mathbf{r}, t)],$$

where $\Psi(\mathbf{r}, t | \mathbf{r}_0)$ is the probability density that the peptide configuration is \mathbf{r} at time t given that it is \mathbf{r}_0 at $t = 0$, $U(\mathbf{r}, t)$ is the potential function described in the next section, ζ_i is the friction coefficient for atom i , k_B is Boltzmann's constant, T is the absolute temperature, and $\beta = 1/(k_B T)$. The equilibrium solution of Eq. 2 is the Boltzmann distribution $\Psi_{\text{eq}}(\mathbf{r}) = Z^{-1} \exp(-\beta U)$, where Z is the partition function $\int d\mathbf{r} \exp(-\beta U)$. The time evolution of any dynamical

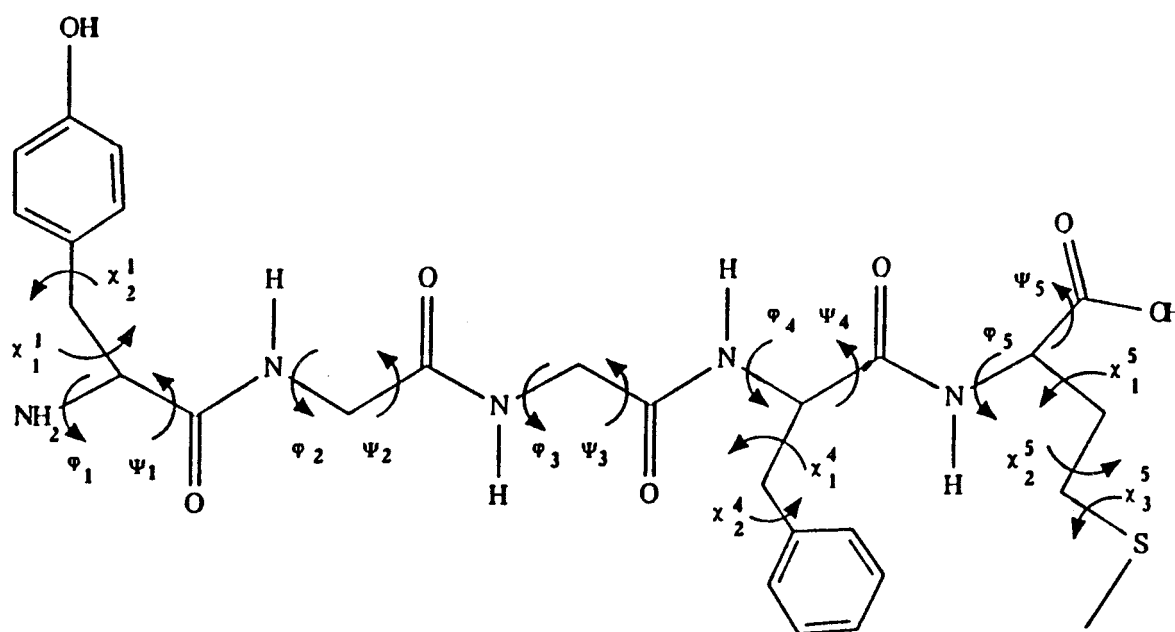


FIGURE 1 Structure of met-enkephalin (Tyr-Gly-Gly-Phe-Met). Arrows indicate the dihedral angles that may exhibit rotations.

observable $g(t)$ is given by

$$\frac{\partial g(t)}{\partial t} = \mathcal{D}^\dagger g(t), \quad (3)$$

where \mathcal{D}^\dagger is the adjoint of \mathcal{D} :

$$\mathcal{D}^\dagger = \sum_{i=1}^N \frac{k_B T}{\zeta_i} [\nabla_i^2 - \beta \nabla_i U(\mathbf{r}, t) \cdot \nabla_i]. \quad (4)$$

Equation 3 can be formally solved for $g(t)$:

$$g(t) = \exp(\mathcal{D}^\dagger t) g(0) \quad (5)$$

The application of projection operator techniques to the diffusion Eq. 2 produces the GLE of Eq. 1 (Zwanzig, 1974; Bixon and Zwanzig, 1978). The diffusion equation, however, can be solved alternatively by considering the eigenvalue problem of the operator \mathcal{D}^\dagger :

$$\mathcal{D}^\dagger \psi_n = -\lambda_n \psi_n, \quad (6)$$

where λ_n and ψ_n are the complete set of eigenvalues and eigenfunctions, respectively, of \mathcal{D}^\dagger . Using Eq. 6 and expanding $\Psi(\mathbf{r}, t)/\Psi_{\text{eq}}(\mathbf{r})$ in the complete set of eigenfunctions ψ_n any time correlation function $\langle f(0)g(t) \rangle$ can be expressed as

$$\langle f(0)g(t) \rangle = \langle f(0)\exp(\mathcal{D}^\dagger t)g(0) \rangle = \sum_n \langle f\psi_n | \langle \psi_n g \rangle \exp(-\lambda_n t), \quad (7)$$

where $\langle \cdot \cdot \cdot \rangle$ denotes the equilibrium average:

$$\langle \cdot \cdot \cdot \rangle = Z^{-1} \int d\mathbf{r} \exp[-\beta U(\mathbf{r})] (\cdot \cdot \cdot) \quad (8)$$

As the eigenvalue equation cannot be solved analytically for the complicated potentials describing the interactions in biomolecules, we approximate the exact solution using a basis set expansion:

$$\psi_n \approx \sum_{i=1}^M C_{in} \phi_i. \quad (9)$$

The C_{in} and λ_n are then determined from the generalized eigenvalue problem:

$$\mathbf{FC} = \mathbf{SC}\lambda, \quad (10)$$

where $S_{ij} = \langle \phi_i \phi_j \rangle$ is the positive definite metric matrix, λ is the diagonal matrix of (approximate) eigenvalues $\lambda_1, \lambda_2, \dots$, and

$$F_{ij} = \langle \phi_i \mathcal{D}^\dagger \phi_j \rangle = - \sum_{k=1}^{3N} \frac{k_B T}{\zeta_k} \left\langle \frac{\partial \phi_i}{\partial r_k} \frac{\partial \phi_j}{\partial r_k} \right\rangle, \quad (11)$$

where the last step follows from successive integrations by parts. The normalization condition is $\mathbf{C}^T \mathbf{SC} = \mathbf{I}$, where \mathbf{I} is

the unit matrix. The time correlation function then becomes

$$\langle f(0)g(t) \rangle = \sum_n \exp(-\lambda_n t) \left(\sum_{i=1}^M C_{in} \langle f | \phi_i \rangle \right) \left(\sum_{j=1}^M C_{jn} \langle \phi_j | g \rangle \right). \quad (12)$$

It is important to note that the theory contains no adjustable parameters; it uses the same potentials and friction coefficients as the Brownian dynamics simulations to be described in the next section. The only approximation lies in the use of a finite basis set in Eq. 9.

BROWNIAN DYNAMICS SIMULATIONS

The diffusion equation is solved numerically by considering the solutions to the equivalent Langevin equation,

$$\frac{d\mathbf{r}_i(t)}{dt} = \frac{\mathbf{F}_i(t)}{\zeta_i} + \chi_i(t), \quad (13)$$

where $\mathbf{F}_i = -\partial U(\mathbf{r})/\partial \mathbf{r}_i$ is the force acting on the i th atom, and $\chi_i(t)$ is a Gaussian random variable with a zero mean. The fluctuation-dissipation theorem provides the variance of $\chi_i(t)$ as

$$\langle \chi_i(t) \chi_j(t') \rangle = 2 \frac{k_B T}{\zeta_i} \delta_{ij} \delta(t - t'),$$

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 (van Gunsteren and Berendsen, 1982):

$$\mathbf{r}_i(n+1) = \mathbf{r}_i(n) + \frac{\mathbf{F}_i(n)}{\zeta_i} \Delta t + \frac{\mathbf{F}'_i(n)}{2\zeta_i} (\Delta t)^2 + \chi_{in}, \quad (14)$$

where Δt is the time step, $F'_i(n) = [F_i(n) - F_i(n-1)]/\Delta t$ 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 $2(k_B T/\zeta_i)\Delta t$.

The simulations and theory use the standard GROMOS87 (van Gunsteren and Berendsen, 1987) force field, which includes the following interactions:

$$U(\mathbf{r}) = U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}} + U_{\text{improper}} + U_{\text{LJ}} + U_{\text{Coulomb}}. \quad (15)$$

The local interactions are modeled by harmonic bond, angle, and improper dihedral angle terms and by specified dihedral angles. The long-range interactions include standard Lennard-Jones and electrostatic potentials. Details about the interaction potentials and the potential parameters are given in the GROMOS manual (van Gunsteren and Berendsen, 1987) and in our earlier paper (Kostov and Freed, 1997).

The dynamics of met-enkephalin at temperature $T = 300$ K is studied in a solvent modeled as a structureless continuum with a viscosity η of 0.84 cp. The sole effect of the solvent is in exerting a viscous damping force and a random force on each atom of the peptide. Each peptide atom is included explicitly, with the exception of the CH_3 , CH_2 , and CH groups, which are treated as united atom units. The atomic friction coefficients are calculated using Stokes' law as $\zeta_i = 6\pi\eta R_i$, where $\{R_i\}$ are the van der Waals radii. The simulations are performed with a time step of 2 fs. The time correlation functions and equilibrium averages are calculated from 18 trajectories starting from random initial conformations. Each 15-ns data collection trajectory is preceded by energy minimization and a 3-ns equilibration. No cutoff is used for the nonbonded interactions. The total kinetic and potential energies remain constant within small fluctuations, a feature indicating the stability of the simulations.

Choice of basis set

The selection of a basis set is the most important part of the theory as it represents the only approximation that is introduced to solve the diffusion equation. We consider polynomial basis functions of the general form $\phi_i, \phi_i\phi_j, \phi_i\phi_j\phi_k, \phi_i\phi_j\phi_k\phi_m, \dots$, where i, j, k, m, \dots are integers between 1 and $N - 1$, and $\{\phi_i\}$ can be any suitable set of functions. If all such possible functions are included in the basis set, the basis set would be complete, and the theory would produce the exact time correlation functions. However, it is impossible to use such a basis set as it contains an infinite number of functions. Therefore, it is essential to select subsets of the complete basis set that produce accurate results. The choice of basis has been extensively discussed in our previous work (Tang et al., 1995; Kostov and Freed, 1997), and therefore only a brief summary is given here for the sake of completeness.

First-order basis

The simplest basis set, which we call first order, is the set of bond vectors

$$\{\mathbf{I}_{ij}\}, (i = 1, \dots, N_b)$$

connecting bonded neighboring atoms. Only two bonds are included from each of the aromatic rings as this is sufficient to describe the orientation of the rings. Chang and Freed have shown (Chang and Freed, 1993) that the $\langle \mathbf{I}_i(t) \cdot \mathbf{I}_i(0) \rangle$ positional correlation functions computed with this basis set are identical to those obtained from the generalized Rouse model with neglected memory functions. Hence, this basis set has a central importance in establishing the connection between the GR model obtained by projection operator methods and the eigenfunction expansion of the diffusion equation, as well as the correspondence between the corrections to the first-order basis set (described below) with

the neglected memory functions in the GR model. In addition, this basis contains a small number of basis functions equal to the number of bonds in the peptide and is therefore convenient to use. The theory with this first-order basis, however, represents the polymer chain as being too flexible, and hence the predicted dynamics are too fast. A more accurate approximation to the correlation functions can be obtained by adding more functions to the first-order basis set.

Mode-coupling basis

Alternative basis sets can be constructed by using physical arguments in the spirit of the mode-coupling theories proposed to describe long-time critical dynamics (Kawazaki, 1970; Kadanoff and Swift, 1968). Define the normal modes, or generalized Rouse modes, as the set of first-order eigenvectors $\{\Psi_i^{(1)}\}$ (with $\{\lambda_i\}$ the corresponding eigenvalues),

$$\Psi_i^{(1)} = \sum_j C_{ij} \mathbf{I}_j. \quad (16)$$

Ordering the eigenvalues as $\lambda_1 < \lambda_2 < \dots < \lambda_n$, it follows that $e^{-\lambda_1 t} > e^{-\lambda_2 t} > \dots > e^{-\lambda_n t}$. Therefore, at sufficiently long times, the slowest modes must provide the dominant contributions to the correlation functions as calculated from Eq. 12 (provided that the amplitudes are not vanishingly small). If λ_1 is much smaller than the next slowest eigenvalue λ_2 , the next slowest mode that can be constructed has a first-order relaxation rate of $2\lambda_1$, which corresponds to the approximate relaxation rate of the bilinear product $\Psi_1^{(1)} \cdot \Psi_1^{(1)}$. Thus, the first-order basis set is expanded by adding new basis functions constructed from products of the slowest decaying normal modes obtained from the first-order basis. Symmetry considerations, discussed in detail by Tang et al. (1995), demonstrate that only odd products of the normal modes contribute to the $\langle \mathbf{I}_i(t) \cdot \mathbf{I}_i(0) \rangle$ time correlation functions studied in the present paper.

Simply including in the basis set all possible trilinear (second-order basis set) or pentilinear (third-order basis set) products of the GR modes produces an explosive growth of the number of basis functions, scaling as N^3 and N^5 , respectively. For the simpler alkanes, it is possible to obtain excellent agreement with the simulations by using only a small subset of the full set of trilinear and pentilinear products. In fact, for the alkanes the dominant correction to the first-order GR theory is provided by including only one trilinear product of the slowest normal mode, and therefore the required basis sets scales linearly with system size. However, peptides have much more complicated potential functions and dynamical behavior than alkanes. Consequently, more basis functions are required to capture the complex correlations in the motions of different portions of the peptide. Our computations for oligoglycines indicate that even the inclusion of all possible trilinear and pentilinear products is inadequate to produce good agreement with the BD simulations. Blindly adding to the basis set yet

higher-order products of the GR modes quickly renders the problem computationally infeasible. In a previous paper (Kostov and Freed, 1997), we show that this explosive growth in the number of basis functions can be avoided by using a simple selection procedure that includes in the basis set only those modes that contribute most to the long-time dynamics.

The relaxation rates of the products of the GR modes are determined in lowest order by the sums of their corresponding eigenvalues. To include in the basis set only the slowest decaying products of GR modes, a large number of possible sums of eigenvalues are selected for sorting as follows:

$$\left\{ \begin{array}{l} \lambda_i + \lambda_j + \lambda_k \quad [i, j, k = 1, \dots, Q; j \geq k] \\ \lambda_i + \lambda_j + \lambda_k + \lambda_m + \lambda_n \\ \quad [i = 1, \dots, P; \text{all possible } \{j, k\}, \{m, n\} \\ \quad \text{with } j, k, m, n = 1, \dots, P] \\ \lambda_i + \lambda_j + \lambda_k + \lambda_m + \lambda_n + \lambda_p + \lambda_q \\ \quad [i = 1, \dots, S; \text{all possible } \{j, k\}, \{m, n\}, \{p, q\} \\ \quad \text{with } j, k, m, n, p, q = 1, \dots, S] \\ \dots \end{array} \right\}.$$

A suitably modified standard index sorting routine is used (Press et al., 1992). The sorting procedure is very fast; it scales as $N \log N$ and adds negligible overhead to the calculations. This allows the evaluation of the first-order relaxation times for a very large number of normal mode products; in principle Q, P, S, \dots can be chosen to be equal to N_b , the number of basis functions in the first-order basis set. However, we find that choosing $Q, P = 12$ and $S, \dots = 4$ is sufficient to capture the slowest decaying modes with largest contributions to the long-time dynamics of met-enkephalin. We include in the sorting procedure sums of up to 13 eigenvalues, corresponding to basis functions containing as many as 13th-order polynomials of the normal modes, although it is computationally feasible to include in the sorting procedure terms of even greater than 13th order. The basis functions corresponding to the smallest 1000–1400 terms are retained in the basis set. These functions are constructed as

$$\left\{ \begin{array}{cccc} \Psi_i^{(1)} & (\Psi_j^{(1)} \cdot \Psi_k^{(1)}) & & \\ \Psi_i^{(1)} & (\Psi_j^{(1)} \cdot \Psi_k^{(1)}) & (\Psi_m^{(1)} \cdot \Psi_n^{(1)}) & \\ \Psi_i^{(1)} & (\Psi_j^{(1)} \cdot \Psi_k^{(1)}) & (\Psi_m^{(1)} \cdot \Psi_n^{(1)}) & (\Psi_p^{(1)} \cdot \Psi_q^{(1)}) \\ \dots & & & \end{array} \right\},$$

where the indices i, j, k, \dots now refer to the specific set of modes selected in the basis set. Thus, a basis set can, in principle, contain normal mode products to any order. We term such a basis the mode-coupling basis set.

This simple sorting procedure provides a general automated rule for the selection of the slow variables. The relevant modes that must be included in the basis set are determined by the differences in magnitude of roughly the 10 smallest first-order eigenvalues. These differences are in turn determined by the time scales and coupling between various collective slow motions. For example, the lowest

eigenvalue corresponds to the overall rotational motion of the polymer chain, whereas the next lowest eigenvalues correspond to collective motions on successively smaller distance scales and faster time scales. Therefore, the selected modes, in general, vary for different molecules. The rotational motion of alkanes is not strongly coupled to the other collective motions, and hence only a few basis functions, containing the slowest first-order normal mode, provide the dominant contribution to the long-time dynamics. In contrast, peptides do not exhibit such time-scale separation, and the overall rotation of the peptide is more strongly coupled with other cooperative motions through the mode-coupling terms. Inclusion of a large number of these mode-coupling terms in the basis set is therefore necessary. (We use ~ 1200 basis functions for met-enkephalin).

RESULTS

We compute the normalized interatom distance autocorrelation functions,

$$C_{ij}(t) = \frac{\langle \mathbf{l}_{ij}(t) \cdot \mathbf{l}_{ij}(0) \rangle}{\langle \mathbf{l}_{ij}(0) \cdot \mathbf{l}_{ij}(0) \rangle}, \quad (17)$$

where $\mathbf{l}_{ij} = \mathbf{r}_i - \mathbf{r}_j$, and \mathbf{r}_i and \mathbf{r}_j are the position vectors for arbitrary atoms i and j in the peptide. Calculations for alkanes (Tang et al., 1995) show that similar trends are obtained for the computationally more involved $P_2(t)$ correlation functions. The correlation functions $C_{ij}(t)$ are evaluated directly from the BD simulations using a time average over 18 trajectories and are displayed with solid lines in Figs. 2–6. The figures also present the same correlation functions as computed from the theory using the first-order basis set and the mode-coupling basis set generated with the sorting procedure described in the previous section. We do not present the computations for the second- and third-order basis sets as we have demonstrated previously that these basis sets 1) include a large number of functions that are inconsequential to the long-time dynamics and 2) omit contributions from other important terms. Thus, the essential contributions from these basis sets are fully incorporated by the sorting procedure.

Global motions

The global motions can be divided into several types: the first group involves only backbone atoms, the second group includes motions of the side chains relative to the backbone, whereas the third encompasses the relative motions of two side chains. Fig. 2 displays a representative illustration of the general trends for the dynamics of long-range virtual bonds in the backbone of met-enkephalin. Fig. 2 *A* compares the autocorrelation function for the virtual bond connecting the C_1^α and C_4^α atoms as determined from the simulations, the first-order basis, and the sorting procedure with 1169 and 1345 basis functions. The theory with the first-order basis set produces poor agreement with the simula-

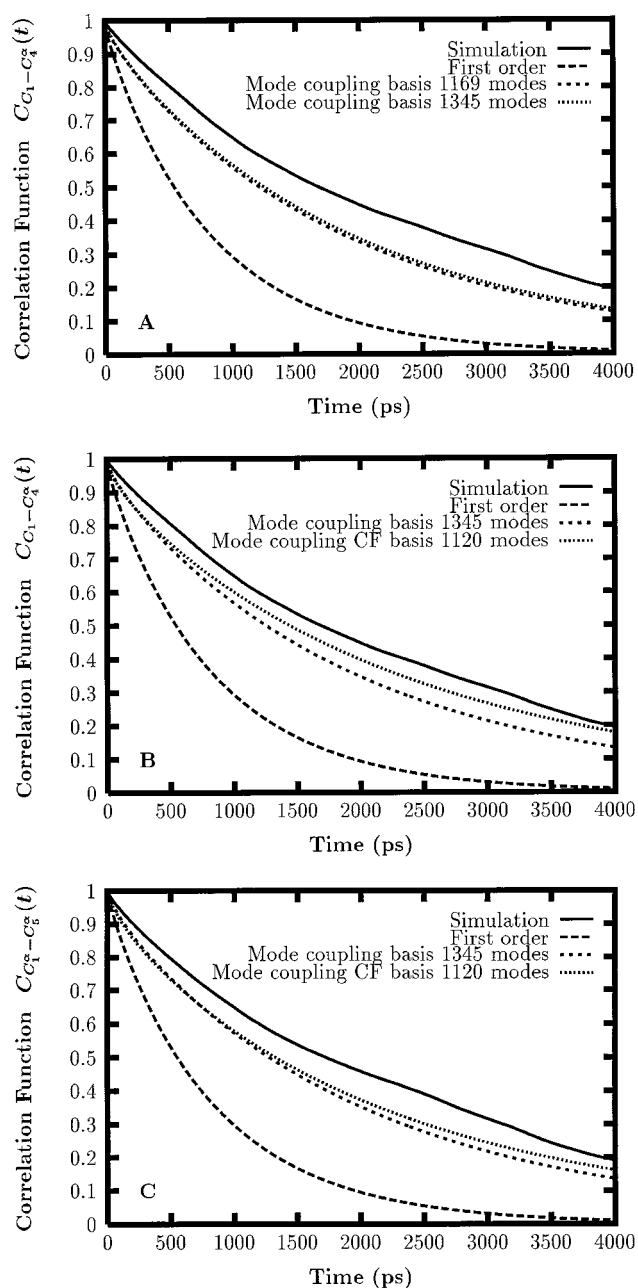


FIGURE 2 Comparison of correlation functions for long-range motions in met-enkephalin as calculated directly from the BD simulations (solid line) and the first-order theory (long dashed line). (A) The C_1-C_4 virtual bond; comparison of the mode-coupling theory with the 1169 slowest modes (short dashed line) and the mode-coupling theory with the 1345 slowest modes (dotted line). (B) The C_1-C_4 virtual bond; comparison of the mode-coupling theory without centers of friction and the 1345 slowest modes (short dashed line) and the CF mode-coupling theory with the 1120 slowest modes (dotted line). (C) The end-to-end C_1-C_5 virtual bond. The line pattern is the same as in B.

tions. This is expected as the first-order basis set corresponds to the GR model, which neglects the memory functions, thereby underestimating the local stiffness and predicting a too-fast dynamics. Fig. 2 A demonstrates that the basis set selected by using the sorting procedure pro-

duces very good agreement with the simulations. The quite small difference between the predictions of the theory with 1169 and 1345 basis functions indicates that convergence is practically reached and that adding more basis functions beyond 1345 would not change the results substantially. The results in Fig. 2 A, as well as those for many other correlation functions indicate that the theory reliably and consistently predicts the relaxation time scales for the long- and mid-range motions in the peptide. However, as discussed in the following subsection, the agreement between theory and simulations can further be improved by using a modified basis set that treats the rigid portions of the peptide more accurately.

Centers of friction (CF) basis set

A large portion of met-enkephalin consists of rigid units: the two aromatic rings as well as the peptide bonds. The local dynamics of these rigid units are the most difficult for the theory to describe (see the following subsection). In addition, the presence of rigid groups influences the global dynamics.

As the rigid units experience highly collective motions, a natural approach to treat the dynamics of these units is to combine their atoms in one group by including in the basis set only the center of friction of the group. For example, grouping the CO-NH atoms in a peptide bond gives the center of friction as

$$\mathbf{R}_{cf}^{pep} = \frac{\zeta_C \mathbf{r}_C + \zeta_O \mathbf{r}_O + \zeta_N \mathbf{r}_N + \zeta_H \mathbf{r}_H}{\zeta_C + \zeta_O + \zeta_N + \zeta_H}. \quad (18)$$

The centers of friction (\mathbf{R}_{cf}^{ring}) of the aromatic rings are defined similarly to include all atoms in the ring along with the coplanar C_β united atom groups.

The bonds between neighboring atoms within the peptide groups and the aromatic rings are replaced in the first-order basis set by the bonds $C_\alpha^i - \mathbf{R}_{cf}^{pep,i}$ and $\mathbf{R}_{cf}^{pep,i} - C_\alpha^{i+1}$ for each peptide group and by a bond between the C_α atom and \mathbf{R}_{cf}^{ring} for the aromatic amino acids. (We find that the best results are obtained when including in the basis the centers of friction of both the peptide bonds and the aromatic rings.) This reduces the total number of first-order basis functions to 20 from the original first-order basis with 39 functions. The mode-coupling basis set is then constructed as outlined above using this modified first-order basis. Spatial resolution is not lost by using centers of friction as Eq. 12 can still be used to compute the dynamics of any bond within a rigid group, even if this bond is not included explicitly in the first-order basis. Basis sets employing the centers of friction will be referred to as CF basis sets both in first-order as well as for the mode-coupling basis sets.

Fig. 2 B displays (for the same bond presented in Fig. 2 A) a comparison of the correlation functions obtained when the bonds in the rigid units are included explicitly in the first-order basis and when these units are represented by their centers of friction. Clearly, grouping together the at-

oms of the rigid units yields better agreement between theory and simulations. Figure 2 C presents the calculations for the end-to-end $C_1-C_5^\alpha$ bond, which exhibit a similar trend.

The improvement in the theoretical predictions by using centers of friction is more pronounced for correlation functions corresponding to global virtual bonds between the backbone and the side chains or for side chain-side chain virtual bonds. The examples in Figs. 3 and 4 reaffirm the conclusion that the theory can predict quite accurately the relaxation times for a wide variety of long- and mid-range motions in met-enkephalin. Fig. 3 presents the calculations for two virtual bonds between the backbone and the ends of the Phe (Fig. 3 A) and Met (Fig. 3 B) side chains. Fig. 3 A emphasizes again that using the CF basis set yields a substantial improvement whereas Fig. 4 A displays the correlation functions for a virtual bond connecting the two aromatic rings, where the contributions from the CF basis are most pronounced. On the other hand, the correlation function in Fig. 4 B for the relative motions of the Phe and Met side chains incurs negligible improvement from using the

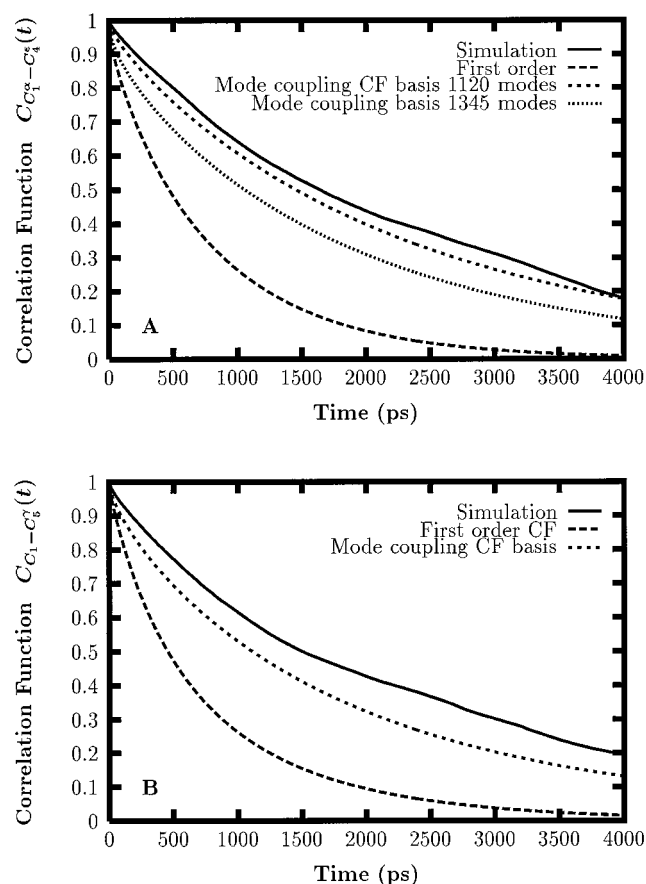


FIGURE 3 Comparison of the correlation functions in met-enkephalin as calculated from the simulations (solid line), the first-order theory (long dashed line), the CF mode-coupling theory with 1120 modes (short dashed line), and the mode-coupling theory without centers of friction and 1345 modes (dotted line) for (A) the $C_1^\alpha-C_4^\alpha$ backbone-Phe bond and (B) the $C_1-C_3^\alpha$ backbone-Met bond.

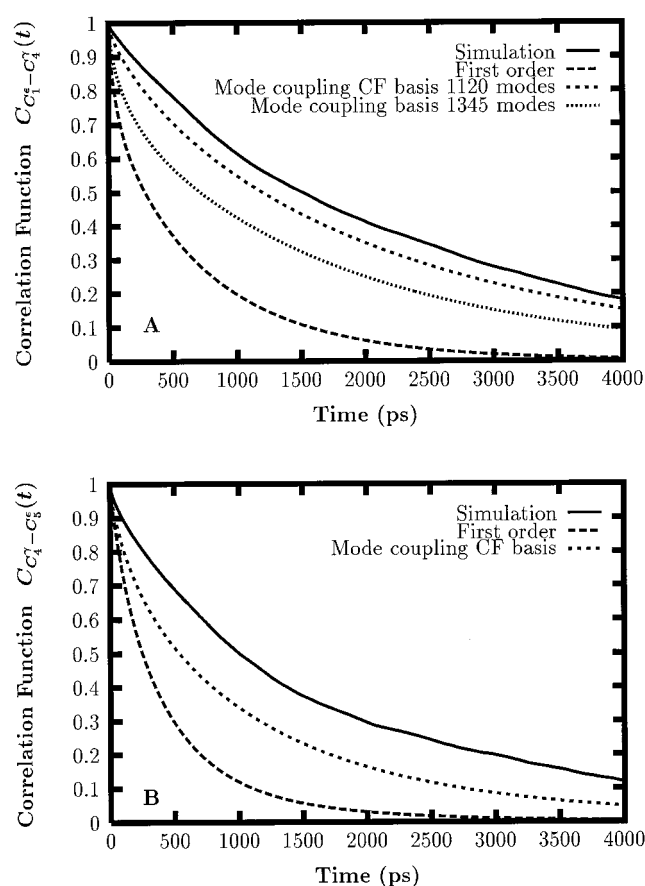


FIGURE 4 Comparison of the correlation functions in met-enkephalin for (A) the $C_1^\epsilon-C_4^\gamma$ Tyr-Phe virtual bond and (B) the $C_4^\gamma-C_5^\epsilon$ Phe-Met bond. The line pattern is the same as in Fig. 3.

CF basis. Although the CF basis set leads to improved theoretical predictions for many bonds, in some cases, as for example the bonds in the Met side chain, the CF basis yields identical correlation functions to those from the original mode-coupling basis set without the centers of friction grouping. As this lack of improvement usually appears for bonds that are not part of the rigid units, these results can generally be explained by the absence of additional coupling between the motions of the rigid units and these particular bonds.

Two advantages accrue from using a centers of friction basis set for the rigid groups. 1) The number of basis functions in the first-order basis is significantly reduced, thereby allowing the treatment of larger polypeptides. Although the twofold reduction from 39 to 20 functions is inconsequential for met-enkephalin as the number of the mode-coupling basis functions is far greater than the size of the first-order basis, this size reduction may become important for proteins where the first-order basis is already large. 2) The theoretical predictions are improved for the correlation functions of a large number of bonds, in particular for bonds involving the rigid units. Therefore, hereafter we present only calculations with the CF basis set using 1120 mode-coupling terms.

Local Motions

Previous work for glycine oligomers (Kostov and Freed, 1997, 1998) demonstrates that the theory provides a good representation for the long-time decay of many local correlation functions. The present work for met-enkephalin confirms these trends, as depicted in Fig. 5 *A* for the $N_3-C_3^\alpha$ bond. However, the accurate prediction of the local dynamics for some of the local bonds represents a challenge for the theory. This is due, in part, to the fact that the theory is developed to treat the dynamics of flexible polymers, whereas the local rigidity of the peptide chain is manifested to a greater extent for the local than for the global motions. In addition, our method is directed at the long-time dynamics, whereas many of the local bonds (such as N—H bonds) exhibit fast relaxation. Thus, computations for a significant number of local bonds appear to exhibit a considerable discrepancy between the theory and BD simulations. The differences in the relaxation times, as obtained from integrals of the correlations functions ($\tau = \int_0^\infty dt C(t)$), vary from a factor of 2—3 for backbone bonds that are part of the peptide group to a factor of 5 or more for the side chain

bonds involving the aromatic rings. Such a discrepancy, although to a lesser extent, is contained in previous computations for the oligoglycines. Its origins are better evidenced by semi-log plots of the correlation functions as presented in Fig. 5 *B* for the $C_1-N_2^\alpha$ bond. Fig. 5 *B* clearly demonstrates that the theory, indeed, correctly predicts the long-time relaxation rates of the correlation functions (i.e., the slopes of the theoretical curves agree with the simulations) and that the discrepancy between theory and BD simulations for the correlation functions is solely due to an initial very fast decay in the theory at short times of up to a few hundred picoseconds. This initial fast decay in the theory is simply explained by noting that the selection of the slowest modes by the sorting procedure emphasizes the long-time dynamics at the expense of omitted modes responsible for accurately depicting the shorter-time dynamics.

More significantly, the largest discrepancies between theory and simulations are seen for bonds that are part of rigid units, such as the peptide bond or the phenyl side groups. The strongly correlated motions exhibited by such rigid units are difficult to capture with the polynomial basis set. Yet, even in these cases, the long-time portions of the theoretical correlation functions correctly predict the relative relaxation times for the different motions, e.g., the motions of the aromatic rings in different directions. Correlation functions for a set of bonds in the phenyl rings are displayed in Fig. 6, which reaffirms the assertion that agreement between theory and simulations is obtained for the long-time relaxation rates.

The short-time deficiency of the theory is natural and is not of great concern as the short time scales are readily accessible by molecular dynamics simulations. Nevertheless, we have tried to improve the results at short times by selecting alternative basis sets. One possible approach is to note that the motion of a particular local bond of interest is most likely to be correlated to the motions of its neighboring bonds. We can therefore form a basis set focused on describing the dynamics of a specific bond by including in the basis set products of the GR modes (describing the global motions) together with products of the bond vectors for the chosen bond and its neighbors (describing the local correlations),

$$\left\{ \begin{array}{l} \Psi_i^{(1)}(\Psi_j^{(1)} \cdot \Psi_k^{(1)}) \\ \Psi_i^{(1)}(\Psi_j^{(1)} \cdot \Psi_k^{(1)}) \cdot (\Psi_m^{(1)} \cdot \Psi_n^{(1)}) \\ \Psi_i^{(1)}(\Psi_j^{(1)} \cdot \Psi_k^{(1)}) \cdot (\Psi_m^{(1)} \cdot \Psi_n^{(1)}) (\Psi_p^{(1)} \cdot \Psi_q^{(1)}) \\ \vdots \\ \mathbf{l}_\alpha(\mathbf{l}_\beta \cdot \mathbf{l}_\gamma) \quad \forall \alpha, \beta, \gamma \\ \mathbf{l}_\alpha(\mathbf{l}_\beta \cdot \mathbf{l}_\gamma)(\mathbf{l}_\delta \cdot \mathbf{l}_\epsilon) \quad \forall \alpha, \beta, \gamma, \delta, \epsilon \end{array} \right\},$$

where the Greek indices refer to the selected bond and its neighbors. The addition of the neighboring bond correlations indeed produces some improvement in the correlation function for the selected bond at short times. This improvement, however, is rather small and does not alter substantially the agreement with the BD simulations.

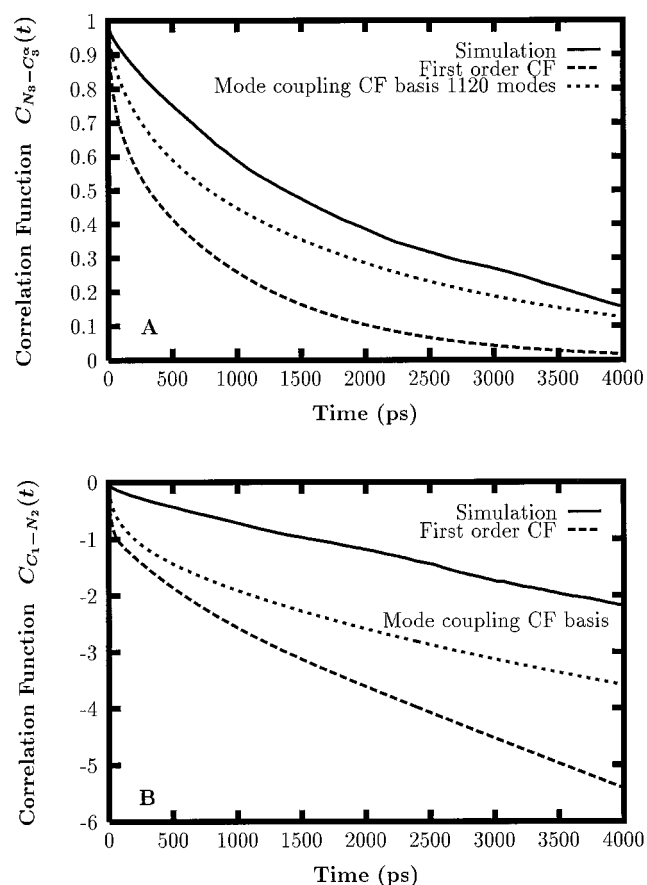


FIGURE 5 Comparison of the correlation functions in met-enkephalin as calculated from the simulations (solid line), the first-order CF theory (long dashed line), and the CF mode-coupling theory (1120 modes) (short dashed line) for (A) the $N_3-C_3^\alpha$ (ϕ_3) bond. (B) Semi-log plot of the C_1-N_2 bond.

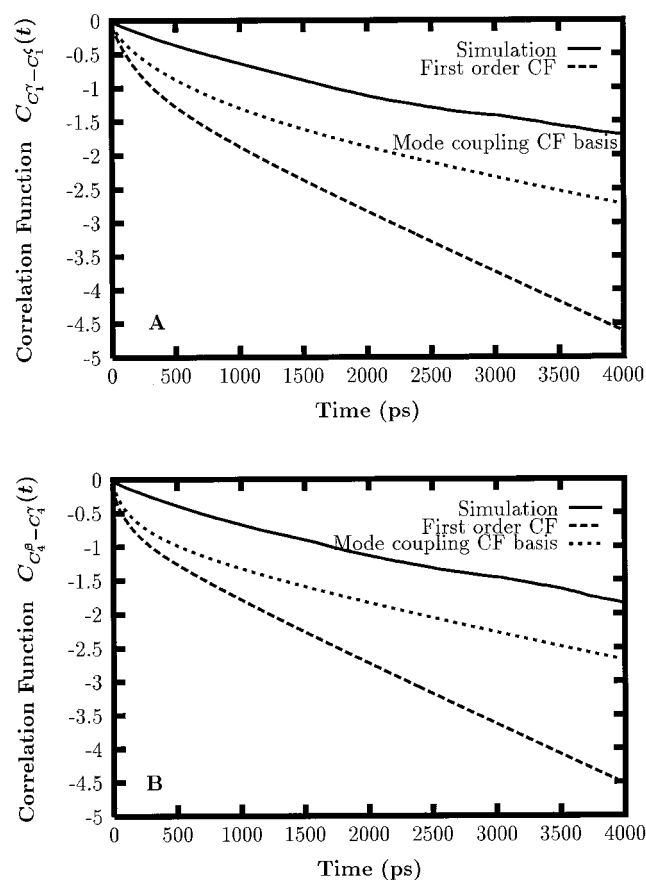


FIGURE 6 Semi-log plot comparison of the correlation functions in met-enkephalin for (A) the $C_1^\gamma-C_1^\delta$ bond in the Tyr ring and (B) a C—C bond in the Phe ring. The line pattern is the same as in Fig. 5.

A similar approach, termed the maximal correlation mode-coupling approximation (MCA), has been recently proposed by Perico and Pratolongo (1997). They study as model systems the dynamics of the freely jointed chain, broken rods, and rods and show that the dominant corrections to the optimized Rouse-Zimm dynamics (a first-order theory) are contained in a small number of product functions involving only the bonds having maximal correlation with the bond of interest. Their MCA basis set adds to the first-order basis all diagonal trilinear products of the form

$$l_{\alpha\alpha}(\mathbf{l}_\alpha \cdot \mathbf{l}_\alpha); \alpha = 1, \dots, N_b,$$

and thus is twice as large as the first-order basis set. A variation of this basis set for treating the dynamics of a given local bond l_α involves including all trilinear nondiagonal terms containing that bond. Perico and Pratolongo show that for the rods this extended basis set improves the MCA predictions by only a few percent. Although the MCA basis contains the dominant corrections to the first-order basis set for the freely jointed chain and rods, our calculations demonstrate that for met-enkephalin this MCA basis improves the first-order dynamics to a lesser extent than achieved by adding to the first-order basis the same number of normal mode products selected with the sorting proce-

dure. In addition, when applied to more complicated peptides and polymers, the MCA approach has the drawback that a different basis set must be constructed for every bond for which the dynamics is studied. Therefore, even though the dimensionality of the basis set is reduced, the generalized eigenvalue problem must be solved many times if the dynamics are of interest for a large number of bonds. In contrast, our mode-coupling basis set provides a universal description of the dynamics of all global and local bonds in terms of a single set of basis functions (and eigenvectors).

DISCUSSION

The global and backbone dynamics of met-enkephalin exhibit many similarities when compared with the previously studied dynamics of oligoglycines (Kostov and Freed, 1997). The relaxation time scales in met-enkephalin are longer by $\sim 50\%$ as compared with octaglycine. Although quantitative agreement of the theory with simulations is obtained for the oligoglycines, the calculations for met-enkephalin display a small discrepancy with the simulations. Semi-log plots indicate that this discrepancy is almost entirely due to an incorrect initial fast decay in the theory. This discrepancy is specific to bonds in rigid groups and might arise from the particular structure of met-enkephalin, which has a large number of rigid units relative to the size of the peptide. Future studies are necessary to determine whether this discrepancy is present for peptides with other lengths and sequences. Met-enkephalin has many interesting motions of phenyl groups and of the side chains relative to the backbone that are absent in the simpler oligoglycines.

Additional insight into the internal motions of met-enkephalin may be obtained by comparing the slopes of the correlation functions at long times. The slopes for various correlation functions differ, frequently by as much as 50%. This indicates that the overall rotation of the molecule, expected to provide the dominant long-time contribution, does not solely determine the very long-time dynamics of both the global and the local bonds. Rather, the overall rotational motion is coupled to other motions on shorter length scales in a manner that is specific for each bond. The local motions in met-enkephalin are, therefore, not only determined by their local environment, but also strongly coupled to more global motions on larger scales. In contrast, traditional interpretations of NMR relaxation data, as for example the model of Lipari and Szabo (1982), frequently rest on the assumption that the local motions in polypeptides are decoupled from the global rotational motion of the molecule. Although this assumption may be appropriate for proteins, the present computations show that it is not valid for small peptides, and additional studies of this assumption are warranted for very flexible portions of proteins. Our theoretical prediction of the coupling between global and local motions accords with NMR experiments for triglycine (Daragan and Mayo, 1994).

The theoretical correlation functions presented in Figs. 2–6 are typically obtained with more than 1000 basis func-

tions. Therefore, the correlation functions are in general represented by a sum of more than 1000 exponentials, although many amplitudes are vanishingly small and only a few terms contribute at the longer times. This number is substantially greater than the two or three exponentials typically used to fit experimental data. However, the correlation functions computed with the theory do not represent a fit. As stressed several times, the theory contains absolutely no adjustable parameters. The same set of eigenvalues present in the exponents for the correlation function of a particular bond are used in the expressions for the correlation functions associated with all possible bonds (real and virtual) in met-enkephalin (a total of approximately 1200 possibilities). Furthermore, this single set of eigenvalues correctly describes the (long-time) dynamics of all bonds in the peptide. The theoretical correlation functions can be fit with a sum of a few exponents, but in general each bond requires different fitting parameters. The relaxation rates obtained from such fitting no longer represent the universal relaxation rates characteristic of the molecule as a whole. On the other hand, the theory generally predicts complex nonexponential relaxations.

As discussed previously, the particular normal mode products included in the basis set are determined by the differences in the magnitudes of the first-order eigenvalues. Therefore, the selected modes, in general, vary for different molecules. As the overall rotation of small peptides strongly couples with other cooperative motions through the mode-coupling terms, inclusion of a large number of such terms in the basis set is necessary. Contrary to the situation found for alkanes, the sequence variation in more complex polypeptides and the presence of specific interactions, such as hydrogen bonds and salt bridges, are expected to require the use of different basis sets for describing peptides with different sequences. Indeed, the specific normal mode products selected for met-enkephalin do not coincide completely with those used previously for oligoglycines. This additional complication, however, is deceptive. The sorting procedure does not require the introduction of empirical fitting, nor does it increase the required computer time. On the contrary, although it is not possible to find a universal set of basis functions that is valid for all polymer molecules or amino acid sequences, the universal sorting procedure provides a simple and systematic way for selecting the slow variables relevant for the motions of any specific polymer and, in particular, polypeptide. Any bond correlations induced by specific interactions are included in the equilibrium averages through the **S** and **F** matrices. Therefore, the influence of strong correlations on the dynamics are reflected in the relaxation times predicted by the theory.

Convergence of the equilibrium averages

Of crucial importance for the accurate prediction of the long-time dynamics is the reliable calculation of the equilibrium averages that serve as input to our dynamical theory.

This requirement is especially important when our approach is extended to treat larger peptides and proteins. The present work evaluates these ensemble averages through time averages obtained from BD simulations. Thus, the time averages must be suitably converged; i.e., further sampling by extending the BD trajectories to longer times should not produce significant changes in the predicted dynamics. To ensure convergence, we include contributions from 18 trajectories, each of 15 ns duration. Calculations with subsets of only 3, 6, or 9 trajectories indicate that convergence is not reached until at least 12 trajectories are included, whereupon all the time correlation functions predicted by the theory become independent of the addition of more trajectories. Fortunately, due to the absence of explicit solvent, the BD simulations are not as computationally intensive as MD simulations and readily allow the accumulation of trajectories for a total of 270 ns, a total duration much longer than generally used for systems containing an order of magnitude more atoms.

Our theory differs substantially from other approaches to extracting information concerning long-range protein motions, such as harmonic analysis, principal component analysis (Balsera et al., 1996), and essential dynamics (Amadei et al., 1993). These methods calculate the eigenvalues and eigenfunctions of a covariance matrix, similar to the **S** matrix in our theory. The eigenvectors of this matrix are then related to displacements of the polypeptide chain due to collective motions, with magnitudes determined by the respective eigenvalues. In contrast, the eigenvalues in our theory produce the observable relaxation rates of various motions. Balsera et al. (1996) have studied the convergence of the covariance matrix with respect to the length of the trajectory. They consider the relatively large protein G-actin in explicit solvent and divide a 475-ps MD trajectory in two parts. Balsera et al. find that the eigenvalues of the covariance matrix increase with the length of the simulation and that the essential subspace changes significantly for the two consecutive time windows. An explanation for these findings is that the covariance matrix has not converged, and, therefore, Balsera et al. conclude that the principal components approach is unreliable. This result is not surprising considering the very short trajectory for a system with a complexity that dramatically exceeds the peptides that we study. A trajectory of only 475 ps is wholly inadequate to provide convergence even for small peptides, such as met-enkephalin, in the absence of explicit solvent molecules.

We carefully monitor the convergence of the equilibrium averages in the present application of our theory. Nevertheless, the evaluation of these averages is an area where significant improvement must and can be achieved to extend our theory to larger polypeptides. Recent work for alkane dynamics (Guenza and Freed, 1996) demonstrates the possibility for circumventing simulations as the means for generating the equilibrium averages by mapping the potentials onto a generalized rotational isomeric state (RIS) model. However, a similar treatment of peptide dynamics is far more complex, as it is not possible to restrict attention in

the peptide to three conformations (*trans*, *gauche*+, and *gauche*−) and as side group motions must be considered. Another possibility is to use MD simulations with potentials of mean force to compute the equilibrium averages. Such simulations would represent a huge saving in computer time. Although the predicted dynamics obtained with potentials of mean force are generally inadequate due to the absence of explicit solvent molecules, our dynamical theory requires as input only equilibrium information. Thus, it may indeed be possible to use potentials of mean force for reliably evaluating the equilibrium averages. Yet another approach is to use Monte Carlo sampling for generating the equilibrium input information. Advances in configurational-bias continuous space Monte Carlo methods have led to the development of highly efficient elementary moves for exploring the configuration space of flexible chain molecules (de Pablo et al., 1993). These advances allow the application of off-lattice Monte Carlo techniques to peptides with extended conformations.

Biological significance of the dynamics

Because of the biological importance of met-enkephalin and its close analog leu-enkephalin, their conformational behavior in solution has been the subject of numerous experimental and theoretical studies. Most of the theoretical studies are methodological papers describing optimization algorithms, but an exception is van der Spoel and Berendsen (1997), which also contains a summary of the most important experimental results. These studies yield many contradictory reports regarding the dominant conformations of the enkephalins, and presently the picture is still unclear. Although some of these studies have been performed with different solvent conditions, the only firm conclusion that can be drawn so far is that the enkephalins are very flexible and exhibit a high degree of conformational diversity. The present paper also falls in the group of methodological studies because its main aim is to test a new dynamical theory and not to identify the specific conformational states occupied by met-enkephalin.

The BD simulations do not provide a completely realistic picture of the conformational dynamics of peptides in solution. The solvent model in these simulations is certainly far too simple to generate an adequate description of the solvent's influence on the dynamics. Thus, we make no attempt to analyze the sequence of conformations traversed by met-enkephalin exhibited in the BD simulations. However, plots of interatomic distances as a function of time for any of the BD trajectories confirm the flexibility of met-enkephalin and the presence of a wide range of conformational transitions. We choose this simplest solvent model because the BD simulations provide a numerical solution to the diffusion equation, which is also the starting point of our theory. Comparison of the theory with the BD simulations then allows a self-consistent test of the theory with no adjustable parameters. We plan in the future to extend the

theory by including more realistic models for the solvent, thereby enabling comparisons of the theoretical predictions with MD simulations and with experiment. Such extensions may alter the details of the conformational dynamics, but not the essence of our general statistical mechanical approach for treating the memory function matrices. The flexibility of the enkephalins observed in both experimental and theoretical studies further emphasizes the need for a conceptual shift in the description of such motions; the long-time dynamics of flexible objects are best described not by a few conformations and the sequence of transitions between them, but by the relaxations of time correlation functions that account for the cumulative effects of many such transitions.

What may be the biological functions of flexibility and conformational diversity? The dynamical nature of proteins may be exploited by biology in many ways. One possibility is that flexibility and disorder in proteins are associated with binding diversity (Kriwacki et al., 1996; Smith and Dobson, 1996). The ability of a protein or hormone to recognize and bind to several different receptors may derive not from the complementarity of preformed protein surfaces, but rather from the ability of proteins to adopt multiple conformations. Met-enkephalin itself is a good example as it binds to several different receptors in the nervous and immune systems (Plotnikoff et al., 1986). A more dramatic example is the cell cycle regulator p21, which in solution exists in a disordered state with no appreciable secondary and tertiary structure (Kriwacki et al., 1996). The amino-terminal domain of p21, however, adopts an ordered stable conformation when bound to its substrate Cdk2. Although such a disorder-order transition has an unfavorable entropy, it allows increased specificity of the protein-protein interactions as p21 binds to multiple cyclins and cyclin-dependent kinases. The ability of p21 to adopt multiple conformations may mediate these diverse binding events.

A second category of motions is the intradomain motion in multidomain proteins with flexible linkers. This flexibility imparts motional independence to the different domains, enabling the protein to perform multiple functions. A very detailed database of such motions has been compiled by Gerstein (Subbiah, 1996).

Dynamics are biologically important even in a compact, single-domain protein with no flexible portions. A classical example is the binding of CO to myoglobin; crystallographic evidence shows that no open channel exists whereby the CO can access the buried heme. CO binding is possible due to the fluctuations of the protein, which create temporary access to the heme. Furthermore, the absence of close packing at protein surfaces allows far more mobility than in the protein core. The backbone atoms in regions not involved in secondary structure may have considerable mobility, and many side chains on the protein surfaces are known to adopt multiple conformations. The fact that many substrate-binding regions located on surfaces are found to be disordered is likely not a coincidence; it reflects the requirement for protein motions in the recognition events between protein and substrate. Flexibility not only allows a

dynamic search for the binding partner but also improves the strength of binding by maximizing the complementarity of the binding surfaces.

Finally, in addition to its functional roles, the flexibility of the polypeptide chain is important in protein folding as it enables exploring a wide range of conformations. This exploration allows the protein to select the pathway(s) with low barriers to structural transitions. The considerable degree of flexibility of partially folded states is confirmed by recent experiments (Sosnick et al., 1996; Smith and Dobson, 1996), which demonstrate that the compact and well ordered characteristics of native proteins emerge only during the final steps of the protein-folding process.

Conclusions

The enormous conformational diversity present in flexible peptides and proteins demonstrates the need to complement the traditional portrayal of their motions, in terms of a few conformations and the transitions between them, with a statistical mechanical description of the long-time dynamics by time correlation functions that account for the cumulative effects of many such transitions. The theory described in this paper naturally models the dynamics of biopolymers in terms of these correlation functions. The main focus of the present work is on the computation of the memory function matrices that describe the influence of internal frictional effects. The neglect of these memory functions is one of the central approximations of the GRZ theory. We demonstrate that the memory functions have a huge influence on the internal dynamics of met-enkephalin, a relatively complicated peptide with biological significance. In contrast to the simpler oligoglycines studied previously (Kostov and Freed, 1997), the presence of side chains, including two aromatic rings, represents a further nontrivial test of the theory and substantially increases the complexity of the dynamics. The rapid growth of basis set size, induced by including the higher-order mode-coupling products that substantially influence the long-time dynamics, is controlled by the previously proposed sorting procedure that selects from the enormous number of possible higher-order products of GR modes only the slowest decaying mode-coupling functions with the largest contributions. The resulting basis set is dramatically reduced in size and produces correlation functions in very good agreement with BD simulations for the global and local motions in met-enkephalin. The inclusion of very-high-order products of the GR modes is crucial to obtain this agreement. The computations for met-enkephalin demonstrate that modification of the basis set to include the centers of friction for the rigid groups leads to further improvement of the theoretical predictions with a concomitant reduction in basis set size. The complicated patterns of conformational transitions in met-enkephalin emphasize the fact that our mode-coupling theory produces eigenfunctions and eigenvalues describing the cumulative, cooperative influence of these local conformational transitions on the long-time dynamics.

This research is supported, in part, by ACS PRF grants 29067-AC7 and 32263-AC7.

REFERENCES

- Amadei, A., A. B. M. Linssen, and H. J. C. Berendsen. 1993. Essential dynamics of proteins. *Proteins Struct. Funct. Genet.* 17:412–425.
- Anteunis, M., A. K. Lala, C. Garbay-Jaureguiberry, and B. P. Roques. 1977. A proton magnetic resonance study of the conformation of met-enkephalin as a function of pH. *Biochemistry*. 16:1462–1466.
- Balsera, M. A., W. Wriggers, Y. Oono, and K. Schulten. 1996. Principal component analysis and long time dynamics. *J. Phys. Chem.* 100: 2567–2572.
- Barrow, C. J. and M. G. Zagorski. 1991. Solution structures of β peptide and its constituent fragments: Relation to amyloid deposition. *Science*. 253:179–182.
- Benham, B. A. and C. M. Deber. 1984. Evidence for a folded conformation of methionine- and leucine-enkephalin in a membrane environment. *J. Biol. Chem.* 259:14935–14940.
- Bixon, M. and R. Zwanzig. 1978. Optimized Rouse-Zimm theory for stiff polymers. *J. Chem. Phys.* 68: 1896–1902.
- Brooks, C. L. III, M. Karplus, and B. M. Pettitt. 1988. Proteins: A theoretical perspective of dynamics, structure and thermodynamics. *Adv. Chem. Phys.* 71: 1–259.
- Chang, X. Y. and K. F. Freed. 1993. Test of theory for long time dynamics of floppy molecules in solution using Brownian dynamics simulation of octane. *J. Chem. Phys.* 99:8016–8030.
- Chen, L. X. Q., J. W. Petrich, G. R. Fleming, and A. Perico. 1987. Pico-second fluorescence studies of polypeptide dynamics: fluorescence anisotropies and lifetimes. *Chem. Phys. Lett.* 139: 55–61.
- Daragan, V. A. and K. H. Mayo. 1994. Peptide dynamics in triglycine: coupling of internal bond rotations and overall molecular tumbling. *J. Phys. Chem.* 98:10949–10956.
- de Pablo, J. J., M. Laso, J. I. Siepmann, and U. W. Suter. 1993. Continuum-configurational-bias Monte Carlo simulations of long-chain alkanes. *Mol. Phys.* 80:55–63.
- Figureirido, F., R. M. Levy, R. Zhou, and B. J. Berne. 1997. Large scale simulation of macromolecules in solution: combining the periodic fast multipole method with multiple time step integrators. *J. Chem. Phys.* 106:9835–9849.
- Guenza, M. and K. F. Freed. 1996. Extended rotational isomeric state model for describing the long time dynamics of polymers. *J. Chem. Phys.* 105:3823–3837.
- Guenza, M. and A. Perico. 1993. Static and dynamic structure factors for star polymers in θ conditions. *Macromolecules*. 26:4196–4202.
- Hu, Y., G. R. Fleming, K. F. Freed, and A. Perico. 1991. Theory of long time peptide dynamics: test of various reduced descriptions and role of internal variables. *Chem. Phys.* 158:395–408.
- Hu, Y., K. S. Kostov, A. Perico, S. Smithline, and K. F. Freed. 1995. Extended molecular dynamics and optimized Rouse-Zimm model studies of a short peptide: various friction approximations. *J. Chem. Phys.* 103:9091–9100.
- Hu, Y., J. M. MacInnis, B. J., Cherayil, G. R. Fleming, K. F. Freed, and A. Perico. 1990. Polypeptide dynamics: experimental tests of an optimized Rouse-Zimm type model. *J. Chem. Phys.* 93:822–836.
- Kadanoff, L. P. and J. Swift. 1968. Transport coefficients near the liquid-gas critical point. *Phys. Rev.* 166:89–101.
- Kapral, R., D. Ng, and S. G. Whittington. 1976. Mode coupling description of dynamics in dilute polymer solutions. *J. Chem. Phys.* 64:539–545.
- Kawazaki, K. 1970. Kinetic equations and time correlation functions of critical fluctuations. *Ann. Phys.* 61:1–56.
- Kirshenbaum, K. and V. Daggett. 1995. pH-dependent conformations of the amyloid β (1-28) peptide fragment explored using molecular dynamics. *Biochemistry*. 34:7629–7639.
- Kostov, K. S. and K. F. Freed. 1997. Mode coupling theory for calculating the memory functions of flexible chain molecules: influence on the long time dynamics of oligoglycines. *J. Chem. Phys.* 106:771–783.

- Kostov, K. S. and K. F. Freed. 1998. Mode coupling theory for calculating the memory functions of flexible polymers: local dynamics of oligoglycines. *J. Chem. Phys.* 108:8277–8278.
- Kriwacki, R. W., L. Hengst, L. Tennant, S. I. Reed, and P. E. Wright. 1996. Structural studies of p21^{Waf1/Cip1/Sdi1} in the free and cdk-2 bound state: conformational disorder mediates binding diversity. *Proc. Natl. Acad. Sci. U.S.A.* 93:11504–11509.
- Li, Z. and H. Scheraga. 1987. Monte Carlo minimization approach to the multiple minima problem in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* 84:6611–6615.
- Lipari, G. and A. Szabo. 1982. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. I. Theory and range of validity. *J. Am. Chem. Soc.* 104:4545–4559.
- Metiu, H., D. W. Oxtoby, and K. F. Freed. 1977. Hydrodynamic theory for vibrational relaxation in liquids. *Phys. Rev. A* 15:361–371.
- Montcalm, T., W. Cui, H. Zhao, F. Guarnieri, and S. R. Wilson. 1994. Simulated annealing of met-enkephalin: low energy states and their relevance to membrane-bound, solution and solid-state conformations. *J. Mol. Struct. (Theochem)* 308: 37–51.
- Mori, H. 1965. Transport, collective motion, and Brownian motion. *Prog. Theor. Phys.* 33:423–455.
- Perico, A. 1989. Segmental relaxation in macromolecules. *Acc. Chem. Res.* 22:336–342.
- Perico, A., and M. Guenza. 1985. Viscoelastic relaxation of segment orientation in dilute polymer solutions. *J. Chem. Phys.* 83:3103–3108.
- Perico, A. and R. Prato-longo. 1997. Maximum correlation mode-coupling approach to the Smoluchowski dynamics of polymers. *Macromolecules.* 30:5958–5969.
- Perico, A., R. Prato-longo, K. F. Freed, R. W. Pastor, and A. Szabo. 1993. Positional time correlation functions for one-dimensional systems with barrier crossing: memory function corrections to the optimized Rouse-Zimm approximation. *J. Chem. Phys.* 98:564–573.
- Plotnikoff, N. P., R. E. Faith, A. J. Murgo, and R. A. Good, editors. 1986. *Enkephalins and Endorphins: Stress and the Immune System*. Plenum, New York.
- Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1992. *Numerical Recipes*. Cambridge University Press, Cambridge.
- Riek, R., S. Hornemann, G. Wider, R. Glockshuber, and K. Wüthrich. 1997. NMR characterization of the full-length recombinant murine prion protein, mPrP(23-231). *FEBS Lett.* 413:282–288.
- Smith, L. J. and C. M. Dobson. 1996. NMR and protein dynamics. *Int. J. Quant. Chem.* 59:315–332.
- Sosnick, T. R., S. Jackson, R. R. Wilk, S. W. Englander, and W. F. De-grado. 1996. The role of helix formation in the folding of a fully α -helical coiled coil. *Proteins Struct. Funct. Genet.* 24:427–432.
- Subbiah, S., editor. 1996. *Protein Motions*. Chapman & Hall, Austin, TX.
- Tang, W. H., X. Y. Chang, and K. F. Freed. 1995. Theory for long time polymer and protein dynamics: basis functions and time correlation functions. *J. Chem. Phys.* 103:9492–9501.
- van der Spoel, D. and H. J. C. Berendsen. 1997. Molecular dynamics simulations of leu-enkephalin in water and DMSO. *Biophys. J.* 72: 2032–2041.
- van Gunsteren, W. F. and H. J. C. Berendsen. 1982. Algorithms for Brownian dynamics. *Mol. Phys.* 45:637–647.
- van Gunsteren, W. F. and H. J. C. Berendsen. 1987. *GROMOS Manual*. University of Groningen, Groningen, The Netherlands.
- Vasquez, M., E. Meirovitch, and H. Meirovitch. 1994. A free energy based Monte Carlo minimization procedure for biomolecules. *J. Phys. Chem.* 98:9380–9382.
- Zhang, G. and T. Schlick. 1994. The Langevin/implicit-Euler/normal-mode scheme for molecular dynamics at large time steps. *J. Chem. Phys.* 101:4995–5012.
- Zwanzig, R. 1961. Memory effects in irreversible thermodynamics. *Phys. Rev.* 124:983–992.
- Zwanzig, R. 1974. Theoretical basis for the Rouse-Zimm model in polymer solution dynamics. *J. Chem. Phys.* 60:2717–2720.