Theory of long time peptide dynamics: test of various reduced descriptions and role of internal variables

Yi Hu, Graham R. Fleming 1, Karl F. Freed 1
Department of Chemistry and the James Franck Institute, the University of Chicago, Chicago, IL 60637, USA

and

Angelo Prisco
Centro studi Chimico-Fisici di Macromolecole Sintetiche e Naturali, Consiglio Nazionale delle Ricerche, 16132 Genova, Italy

Received 25 July 1991

We extend our recent theory of the long time dynamics for peptide molecules in solution to include side chain motions, interresidue interactions, the influence of solvent motions, and the relaxation correlation functions for arbitrary vector directions in the molecule. The theory represents a reduced statistical mechanical description of the system, in which several models are introduced to probe the relevant slow variables governing the long time peptide dynamics. The solution of the dynamical equations requires only the rather trivial diagonalization of matrices, but the complicated portion involves the computation of equilibrium averages for the bond vector correlation and hydrodynamic matrices. The latter are evaluated here using molecular dynamics simulations (including the solvent molecules) that are performed with the CHARMm package. The three models considered here take as relevant slow variables (a) the positions of the Cα-carbons (the virtual bond model), (b) the positions of all backbone bonds (the backbone bonds model), and (c) both the positions in (b) and those along the side group chains. Ignoring memory functions and using an additivity model for group friction coefficients, calculations with all three models are performed of the orientational relaxation correlation functions and correlation times for a variety of local bonds and directions in the peptide. Comparisons are made between the predictions of the three models (and with experiment) to assess the relative importance of retaining in the reduced description various categories of degrees of freedom and the importance of the neglected memory functions.

1. Introduction

The development of a description and an understanding of the internal motions of peptides and proteins remains an extremely challenging problem. Given the availability of protein interaction potentials [1] it is in principle possible to run molecular dynamics (MD) calculations to describe this dynamics. However, when interest centers on the biologically interesting longer time scale dynamics, it becomes totally impractical to, for instance, calculate every femtosecond the coordinates of each atom in the system (including the solvent molecules) for a period of many nanoseconds or longer. In addition, the information generated in this fashion would be so overwhelming as to be practically useless.

It is clear that some kind of reduced description is essential to provide experience and information which will enable predictive dynamical models to be developed. Of particular interest is the construction of theories describing the long time protein dynamics, say over time scales that exceed the duration of typical MD trajectories by a factor of several thousand. One statistical mechanical approach to the problem is associated with finding, at an acceptable level of accuracy, which degrees of freedom must be included explicitly in a reduced description and which can be considered part of the heat bath [2].

In an earlier paper [3] we developed and tested a particularly simple reduced description for the dynamics of a polypeptide in aqueous solution. An essential aspect of the theory is the fact that the treat-

---

1 To whom correspondence should be addressed.
The dynamical theory uses a hydrodynamic model (for a review, see ref. [4]) to describe the solvent friction, and one of the input matrices is the equilibrium average of the hydrodynamic interaction matrix. The other input matrix is a structural matrix of bond vector equilibrium correlations. Our previous calculations consider a simple model for the peptide which is represented as a sequence of virtual bonds connecting residues with individual friction coefficients, but with no internal structure (side chains). The equilibrium averages of the hydrodynamic and structural matrices are evaluated using the ECEPP [5] empirical potential function and neglecting interactions between residues. The calculations predict very interesting qualitative dependence of the protein long time dynamics on the residue sequence, and comparisons with experimental data [3] on fluorescence depolarization from tryptophan in peptides are quite encouraging. These last features of our previous rather primitive virtual bond model have stimulated us to lift several of the previous assumptions and to test their significance. This paper describes these new theoretical developments and focuses on how the internal peptide degrees of freedom affect the long time peptide dynamics. The new work employs a MD simulation of a hexapeptide (the ACTH fragment 5-10 [6], see fig. 1) in a model water solvent to provide the equilibrium averages required by our dynamical theory and to test models of varying levels of sophistication.

In particular, the theory is extended to the treatment of any number of bonds in the system as primary dynamical variables. The most complete model uses all the bonds (including those in the side chains but excluding extraneous rigid ones in aromatic rings) in the dynamical treatment, while an intermediate model retains only those bonds lying along the chain backbone. The simple virtual bond model is also employed for comparison. All inter-residue interactions are now included as the bond vector correlation and hydrodynamic interaction matrices may be computed directly from the MD simulation.

Comparisons with experimental fluorescence depolarization data [6] require the consideration of the orientational correlation function for a dipole that is oriented along the \( ^1L_\alpha \) tryptophan transition moment. The previous virtual bond model is forced to approximate this dipole orientation as lying along one of the virtual bonds, while the present full bond model has bond 36 in fig. 1 pointing essentially along the tryptophan \( ^1L_\alpha \) vector direction, thereby facilitating a more direct comparison with experiment.

We continue to use a hydrodynamic model of the solvent friction since the solvent dynamics is expected to be rather fast compared to the 200 ps time scale protein dynamics of interest to us. Given this hydrodynamical solvent model, the protein molecule dynamics is Markovian, but is still much too complicated. It is therefore relevant to inquire whether some protein degrees of freedom can be treated as “relevant” slow dynamical variables that evolve under the influence of a new bath that now consists of the remaining “fast” internal protein degrees of freedom in addition to the solvent [2]. This reduction procedure is formally exact provided that the correct memory functions are available or suitably approximated. However, it is rare that memory functions are computed from first principles, and it is extremely
difficult to introduce model forms for the many memory functions that would be required in treating the protein dynamics (for general arguments see appendix A in ref. [7]). Thus, we continue to follow the standard practice of neglecting memory functions in the three models considered here. When the memory function is neglected, the viability of partitioning the protein degrees of freedom into "active" and "bath" parts depends on the existence of a time scale separation in their characteristic motions. Our three models enable us to progressively include more and more protein degrees of freedom into the space of active variables. The simplest model has 5 virtual bonds, the intermediate has 15 backbone bonds, and the most complete model has all 36 bonds, including those in side chains. The three models are illustrated in fig. 1. As we pass from the simplest to the most complete model, more degrees of freedom are included into the theory and hence also included are portions of the memory functions that are neglected in the simpler models. Thus, comparisons between the models enable us to discern some aspects of the importance of the neglected memory functions. This approach also enables us to test for the appropriate choice of slow dynamical variables for describing the long-time behavior of the peptide motion. In addition, the model of additive solvent friction coefficients turns out to depend on the number of degrees of freedom retained in the model, and a characterization of this effect is provided here.

Our goals in this paper are manyfold. The theory is generalized here to describe the dynamics of floppy peptides in solution, and significant further developments are obviously required to treat the dynamics of proteins with both floppy portions and higher-order structures. Thus, the present work is an initial step in the direction of a theory for long-time protein dynamics. While comparisons with experiments are limited both by the quality of the potentials and the length of the simulation run, the comparisons of different reduced descriptions are unaffected by the potentials and simulation length. The MD approach provides a self-consistent algorithm for producing the three different models with varying degrees of freedom. The comparisons of these models provide essential information that is necessary to pass onto the next stages of treating peptides with higher-order structures.

The next section briefly describes the general dynamical theory and the three models used, while the following section discusses the MD simulation. Computed orientational correlation functions and correlation times are described and compared between the three models. Appendix A provides the extension of our previous dynamical theory to enable treatment of branched chains, bonds of unequal lengths, and relaxation of orientational correlations for arbitrary directions, while appendix B generalizes a theory of Zwanzig and Ailawadi [8] to provide error estimates on the computation of equilibrium averages from trajectories of finite duration.

2. Dynamical model

A detailed account of our simplest dynamical model is presented in ref. [3], so only a brief discussion is given here. Extensions of this simplest model are introduced to provide a more detailed and realistic description of the protein dynamics. The present section summarizes the theory for a linear polymer chain with all bonds of equal root-mean-square length, while appendix A documents the generalizations required to describe the side chain motion and bonds of unequal lengths.

The polypeptide chain is modelled as a sequence of $n-1$ bonds joining $n$ beads at the spatial positions $R_i$. Each bead has a friction coefficient $\zeta$. Following standard projection operator methods [2], the time evolution of the bead coordinates obeys a generalized Langevin equation. Implicit bath degrees of freedom and projected out protein degrees of freedom contribute to the dynamical equation through an additional time-dependent friction, i.e., the memory function, which is ignored. Hydrodynamic interactions are taken to describe the hydrodynamic transmission of friction forces through the solvent and are treated using Kirkwood's preaveraging approximation in which the Oseen tensor (see below) is averaged with the exact equilibrium distribution function. The resultant theory displays a form which is a generalization of the classic Rouse-Zimm theory of polymer dynamics [9-12],

$$\frac{\partial R_i(t)}{\partial t} + \sigma \sum_{j=0}^{n-1} (HA)_{ij} R_j(t) = v_i(t) . \quad (1)$$
The matrix $A$ in eq. (1) is of order $n$ and is expressed in terms of the inverse $U$ of the normalized equilibrium bond correlation matrix \[ U^{-1} = \langle l_i^* l_j \rangle / l^2 \] as \[ A = M^T \begin{pmatrix} 0 & 0 \\ 0 & U \end{pmatrix} M , \] with the matrix $M$ of order $n$ given for linear chains as \[ M = \begin{pmatrix} 1 & 1 & \ldots & 1 \\ n & n & \ldots & n \\ -1 & 1 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix} . \] The generalized $M$ matrix for branched chains is described in appendix A. The bond vector $l_i$ in eq. (2) may be written in terms of the bead positions as \[ l_i = R_i - R_{i-1} \quad (i = 1, \ldots, n-1) , \] and $l$ is the average length of all the bonds included in the description. The matrix $H$ is the equilibrium average of the Oseen hydrodynamic interaction matrix \[ H_{ij} = \xi_i^{-1} \{ \delta_{ij} + \xi_i \xi_j \langle l_i/R_{ij} \rangle (1 - \delta_{ij}) \} \] with the angular brackets designating an equilibrium average and $\xi_i$ the reduced bead friction coefficient of segment $i$, \[ \xi_i = \xi_i / \zeta = n \xi_i / \sum_{j=0}^{n-1} \xi_j . \] The reduced overall friction coefficient is defined by \[ \zeta = \xi / 6 \pi n_0 l \] and the limiting value of $\zeta = 0$ yields the idealized free draining limit in which the hydrodynamic disturbance of the solvent by the bead motion is neglected. The following calculations choose the value $\zeta = 0.25$ which is appropriate for ordinary polymers in ideal solutions [4], while ref. [3] studies the implications of other choices of $\zeta$. The Gaussian random velocity fluctuations $v_i^c$ have the correlation function \[ \langle v_i^c(t) v_j^c(t') \rangle = (6 k_B T / \zeta) H_{ij} \delta(t-t') , \] and the constant $\sigma$ in eq. (1) is \[ \sigma = 3 k_B T / l^2 \zeta = k_B T / 2 \pi n_0 l^2 \zeta . \] The solution of the generalized Rouse-Zimm model, eq. (1), permits the evaluation for each bond $l_i$ of a rotational correlation function $P_l^z(t)$ which is defined as \[ P_l^z(t) = \frac{1}{2} \langle [l_i(t) \cdot l_i(0)]^2 / l_i^2(t) \cdot l_i^2(0) \rangle - \frac{1}{2} . \] This $P_l^z(t)$ can be expressed in terms of the fundamental time correlation function $M'(t)$ \[ M'(0) = \langle l_i(t) \cdot l_i(0) \rangle / \langle l_i^2 \rangle . \] Transforming eq. (1) to normal coordinates $\{ \zeta_n \}$ provides the fundamental time correlation function $M'(t)$ in the form \[ M'(t) = \sum_{\alpha=1}^{n-1} (Q_{\alpha\alpha} - Q_{\alpha-1\alpha})^2 \mu_\alpha^{-1} \exp(-\sigma \lambda_\alpha t) , \] where $Q$ and $\{ \lambda_\alpha \}$ are, respectively, the matrix of eigenvectors and eigenvalues of the product matrix $HA$, while $\mu_\alpha^{-1}$ is proportional to the mean-square length of the normal mode $\zeta_n$ \[ \langle \zeta_n^2 \rangle = l^2 \mu_\alpha^{-1} . \] We emphasize the fact that the current theory completely reduces the description of the protein dynamics to the evaluation of the equilibrium averages of the $H$ and $A$ matrices and the simple determination of the eigenvalues and eigenvectors of $HA$. The computation of the basic correlation functions (10) and (11) then merely involves substitution into the analytical formulas (12) and (14) below, respectively. The second-order time correlation function then follows exactly from eq. (1) as \[ P_l^z(t) = 1 - 3 [x^2 - x^3 (\pi / 2) \{ 1 - (2 / \pi) \arctan x \} ] , \] where \[ x = [1 - (M'(t))^2]^{1/2} / M'(t) . \]
Experiments are able to measure the rotational correlation time $T'_2$ of a particular bond, and this is related to the function $P'_2(t)$ by

$$T'_2 = \int_0^\infty P'_2(t) \, dt .$$

(16)

The accuracy of our dynamical theory depends on the severity of the two basic approximations, namely, neglecting the memory function in the generalized Langevin equation to obtain eq. (1) and preaveraging the hydrodynamic interactions in eq. (6) [10]. The latter approximation has been studied extensively for synthetic polymers where the errors due to preaveraging are remarkably small [14]. Thus, we focus here on errors associated with the common neglect of memory functions. An increasingly detailed description of the protein dynamics is illustrated by reference to fig. 1 for the polypeptide ACTH fragment (5–10). The simplest approximation coincides with the model used in ref. [3] where the $\{R_{i+1} - R_i\}$ involve only the virtual bonds, imaginary links joining successive $\alpha$-carbon atoms of residues $i+1$ and $i$ (5 in total). The second model takes the $\{R_{i+1} - R_i\}$ to include all backbone bonds (15 in total), while the most complete model has all bonds, including those in side chains (36 in total, as displayed in fig. 1). Backbone bond numbers are not explicitly noted to avoid clutter in the figure. Because of the rigidity of aromatic rings, the all-bond model retains only sufficient bonds to specify the orientation of these rings. Thus, the all-bond model permits all the dihedral transitions that are possible in the ACTH(5–10) fragment. Some of these motions have a relaxation time considerably longer than the experimental tryptophan residue orientational correlation time, reflecting the unimportance in the model of these dihedral transitions on this time scale. The $H$ and $A$ matrices in the three models are evaluated directly by explicit MD simulations for the ACTH(5–10) fragment. Computations are then made of the $T'_2$ and $P'_2(t)$ for all $i$ with each of the three models to determine how the correlation times are influenced by the progressive neglect of more and more protein degrees of freedom.

3. Molecular dynamics simulation

The MD simulations for ACTH(5–10) plus 945 solvent water molecules have been performed using CHARMM, version 19 [11]. The computations begin by generating an all-trans initial configuration of ACTH(5–10) from the CHARMM residue topology and parameter files. ACTH(5–10) with this configuration is then placed at the center of a cube containing TIP3 model [15] water molecules, whose oxygen atoms are not permitted to lie within 2.6 Å of any heavy atom in ACTH to avoid overlap or close contact between solute and water molecules. The cube has an edge length of 31.1032 Å, and the whole system contains $945 \times 3 + 110 = 2945$ atoms. The energy of the system is then minimized to release energetically unfavorable contacts between atoms or highly strained dihedral angles which might have been introduced when constructing the starting configuration. This is followed by a 30 ps period during which the system is equilibrated to establish total energy conservation. During this process, the atom velocities (and hence the total kinetic energy) are periodically scaled to maintain a constant temperature. The final MD simulation then takes the equilibrated structure as its initial point and freely evolves to generate a trajectory for each atom in the system. The stretching of hydrogen containing bonds is suppressed using the SHAKE algorithm, enabling an integration step size of 1 fs to be used. Periodic boundary conditions are employed, and a cutoff radius of 8 Å is used in calculating both electrostatic and van der Waals energies. The nonbonded and hydrogen bonded interaction lists are updated periodically.

A single 90 ps trajectory has been generated, with the coordinates, velocities, and energies saved every 20 fs. The simulation has been performed on an Ardent TITAN computer (which is no longer available to us). Each 6 ps of trajectory requires approximately 70 h of CPU time. Obviously, the CPU intensive nature of this MD simulation precludes the possibility of averaging over several trajectories, extending the simulation to longer times, or considering larger proteins (with more water molecules required). These facts emphasize the need for a theory of long-time protein dynamics.

The calculation of the $H$ and $A$ matrices is straightforward given the definitions in eqs. (3) and (6).
They are determined by replacing the equilibrium averages by time averages along the MD trajectory. Errors associated with this procedure are discussed in the next section and in appendix B. The unavailability of experimental friction coefficients for fragments of amino acids has led us to use Stokes' law \( \zeta_i = 6\pi\eta_i r_i \), with \( \eta_i \) the solvent viscosity and \( r_i \) the hydrodynamic radius of the group. The van der Waals increment method [16] is used to estimate friction coefficients for all three models in a fashion that permits comparisons to be made between computations for the three models. The friction coefficients for the most complete (all-bond) model are calculated using Stokes' law with \( r_i \) the van der Waals radius for the group of atoms comprising the segment, e.g., a C-H unit. The aromatic rings are apportioned between pairs of segments for the units associated with bonds 21–22, 25–26 and 35–36. The segment friction coefficients in the other two models are computed by summing the all-bond model friction coefficients over the constituent subunits of that segment. This model of additivity of group friction coefficients omits the influence of hydrodynamic interactions within the group, an approximation that is discussed further in the next section.

Our use of MD simulations to evaluate the equilibrium average \( H \) and \( A \) matrices represents a significant advance beyond our previous work in several respects. First of all the CHARMM potentials differ from the ECEPP ones. The current simulations explicitly include the water molecules which move and rearrange during the MD run, while the water interactions are only contained within the ECEPP potentials as an approximate potential of mean force. The prior work [3] evaluates the \( A \) matrix by averaging only over rotations about the backbone bonds, with side groups frozen in a preferred orientation, inter-residue interactions are ignored, and the \( H \) matrix is approximated using a simple Gaussian model (often employed for polymers). The current treatment, on the other hand, includes averaging over side group motions and inter-residue interactions, and the \( H \) matrix of eq. (6) is determined by directly evaluating the average \( \langle 1/R_{ij} \rangle \) from the MD simulation. It was our original intention to run an additional MD simulation with inter-residue interactions suppressed in order to test the Flory theory of their relative unimportance in random coil polypeptides [17]. However, the enormous computational cost prevented us from testing this longstanding approximation.

### 4. Results and discussion

The ensemble averages \( \langle l_i^*l_j \rangle \) and \( \langle 1/R_{ij} \rangle \) are calculated from a single trajectory of finite duration \( T \), and therefore there are errors associated with this finite averaging. The error in evaluating the average \( \langle A \rangle_T \) of a dynamic variable \( A(t) \) from the finite trajectory in place of the equilibrium value \( \bar{A} \) is shown in appendix B to be

\[
\sigma^2 = c \left[ x^2 \exp \left( -\frac{1}{x} \right) + x - x^2 \right],
\]

where \( x = \tau/T \) with \( \tau \) the relaxation rate of the correlation function \( \langle \delta A(0)\delta A(t) \rangle \) for \( \delta A(t) = A(t) - \bar{A} \), and \( c \) is defined in appendix B. The quantity \( c \) is computed as typically on the order of \( 10^{-2} \)–\( 10^{-1} \) for \( H \) matrix elements and \( 10^{-4} \)–\( 10^{-3} \) for \( A \) matrix elements. Eqs. (12)–(16) display \( T_1^* \) as a complicated function of the eigenvalues and eigenvectors of the \( HA \) matrix. Therefore, it is nontrivial to evaluate the error in \( T_1^* \) from the uncertainty in the \( H \) and \( A \) matrices, especially when each matrix element has a different variation. However, it is obvious that the error \( \sigma^2 \) decreases when the average is generated from a longer time trajectory and/or for a faster decaying dynamical process. To further assess the error involved in finite time averaging, the full 90 ps trajectory is split into three equal segments and separate calculations are made using each of the 30 ps trajectories.

#### 4.1 Influence of degrees of freedom

Figs. 2, 3, and 4 display the bond correlation times \( T_1^* \) which are obtained from the 30 ps and 90 ps trajectories for the virtual bond, backbone bond and all-bond calculations, respectively. The \( T_1^* \) in fig. 2 from the full 90 ps trajectory is roughly the average of that from the 30 ps portions, but figs. 3 and 4 show much greater divergences. While the differences among the three 30 ps trajectories with each model exhibit the anticipated fluctuations, the differences between the \( T_1^* \) from 30 ps and 90 ps trajectories become larger in passing from the virtual bond, to the backbone bonds.
calculations. Although the $\langle I_i \cdot I_j \rangle$ and $\langle 1/R_{ij} \rangle$ matrices emerging from the 90 ps trajectory are the averages of those from the three 30 ps trajectories, the dynamics are rather nonlinear functions of the eigenvalues and eigenvectors of the $HA$ matrix of eq. (1). Thus, there is no simple relation between the $T_i^v(90\text{ps})$ and $T_i(30\text{ps})$. The fact that the 90 ps trajectory calculation yields $T_i^v(90\text{ps})$ with similar time scales as $T_i^v(30\text{ps})$ for the virtual bond model suggests that statistical equilibrium is established for this simplified model. Perhaps this is due to the fact that each virtual bond unit represents the sum of a number of random variables, i.e., one virtual bond segment contains a few flexible bonds (three in the backbone and several in side groups). This contrasts with the two more detailed models where the $T_i^v(30\text{ps})$ fail to converge to the $T_i^v(90\text{ps})$. These two models contain more “internal” degrees of freedom; thus, a longer trajectory appears to be required to achieve an equilibrium average. Calculating the dynamics of these two models using an average over too short a trajectory simply exaggerates the protein stiffness. Figs. 3 and 4, on the other hand, demonstrate better agreement between the $T_i^v(30\text{ps})$ and $T_i^v(90\text{ps})$ for those bonds with a relaxation time that is shorter than 80 ps. This arises because an equilibrium is more readily established for processes with
fast dissipation, and the error estimate (17) is improved for smaller \( \tau \). Unfortunately, longer MD trajectories become computationally prohibitive even for the small protein fragment considered here.

Fig. 5 presents the \( T_2^s \)'s for each bond as computed using the three models and the full 90 ps trajectory. The bond numbering is as presented in fig. 1. Each virtual bond unit contains three backbone bonds and the associated side group. Because of the existence of relative motion among the backbone bonds, the backbone bonds relax faster than the virtual bonds whose units contain three backbone bonds. Bonds 1–15 in the backbone and all-bond models refer to identical bonds and thus allow direct comparison between the two models. In passing from the all-bond to the backbone model, there is a substantial increase in the relaxation rates for the rotational correlation of most backbone bonds. Several competing factors produce this general increase, including the neglect of side group degrees of freedom in one model and a change in effective friction between the two models. This is more clearly demonstrated in fig. 6 where rotational relaxation times for the virtual bonds 1–5, determined with the backbone bond and all bond models (in doing so, we use the same \( H \) and \( A \) matrices as in the backbone or all bond model to construct the dynamics for these virtual bonds, see appendix A), are presented along with the corresponding times from the virtual bond model. The backbone bond model produces virtual bond relaxation that follows the same trends as a function of bond number as in the virtual bond model, although the relaxation becomes noticeably faster. On the other hand, the virtual bond dynamics in the all-bond model becomes slower and also displays more structure due to explicit inclusion in the model of the side group motions that strongly couple to the relaxation of the virtual bond orientational correlations.

The sawtooth pattern, exhibited by the backbone bond relaxation times as a function of bond number, may easily be understood in terms of rigidity of the peptide bonds which have partial double bond character. Each of these rigid bonds (bonds numbered 2, 5, 8, 11, and 14) has connections with two rather flexible bonds whose motions affect the orientation of the peptide bonds. The dynamics of these rigid bonds are therefore faster than that of the flexible neighboring bonds which are connected to one rigid peptide bond and another flexible bond. This behavior persists when side chain motions are included directly into the model; their influence on the \( \langle I_1 \cdot I_2 \rangle \) and \( \langle 1/R_{ij} \rangle \) averages is contained in the other models. In addition, there is no obvious time scale separation between the motion of the backbone bonds (bond number 1–15) and side chain bonds (bond number 16–36) as would otherwise be implied when taking backbone bonds as the only slow dynamic

![Fig. 5](image-url)  
Fig. 5. The \( T_2^s \) for all three models as calculated using the full 90 ps trajectory. The solid curve is for the virtual bond (5 in total) model, the dotted curve is for the backbone bond (15 in total) model, and the dashed curve is for the all-bond (backbone bonds: 1–15, side chain bonds: 16–36) model. Refer to fig. 1 for detailed bond numbering.

![Fig. 6](image-url)  
Fig. 6. The \( T_2^s \) for virtual bonds 1–5 as calculated from the three models. The line pattern is the same as in fig. 5.
variables. This clearly demonstrates that contributions from side chains to the backbone bond dynamics cannot be simply described by taking the side chains as part of the bath along with the solvent molecules \[18\]. The side group motions must either be included explicitly into the dynamic model on an equal footing with the backbone bonds, or they must be incorporated implicitly into memory functions. However, the former procedure is considerably simpler to accomplish. We should note that the all-bond model contains as dynamical variables all the bonds with the slightest degree of flexibility. Extended models would include mode-coupling nonlinear powers \[7\] of these bond vectors or, equivalently, memory functions. In passing, for example, from the all-bond model to the backbone model, side chain degrees of freedom are omitted. Because these degrees of freedom move on the same time scales as the backbone bonds, omitting them effectively reduces the overall friction forces acting on backbone bonds. These friction forces would appear in the memory functions if they were retained. Thus, the neglect of side group degrees of freedom tends to speed up the backbone bond dynamics. Another factor that changes between the two models is the overall friction, and this, as discussed in the next subsection, works in the opposite direction to the neglect of degrees of freedom.

It is interesting to compare the aromatic side group orientational relaxation rates as described by the different models. Bonds 21, 25, and 35 reside on aromatic side groups and are closer to the backbone than are bonds 22, 26, and 36, respectively (see fig. 1). Calculations with the all-bonds model (see fig. 5) show the former bonds relax faster than the latter. Bond 36 resides in indole moiety of the tryptophan residue and points roughly along the tryptophan \(1L_a\) transition dipole moment. Fluorescence anisotropy measurement of the tryptophan in ACTH(5-10) monitors the rotational relaxation of the \(1L_a\) dipole moment, and experiments find a relaxation time of 250±10 ps \[6\]. This is to be compared with our calculation of 240 ps for bond 36 in the all-bond model. The two reduced models represent the indole ring motion as the motion of the two neighboring virtual bonds (numbers 4 and 5) and backbone bonds (numbers 12 and 13), and the average (for the pairs of bonds) times computed within these models are 202 ps and 195 ps, respectively. Our earlier study \[3\] used the virtual bond model and substantially underestimates the relaxation time as 100 ps. However, this previous work employs different potential functions (ECEPP), ignores interactions between residues, introduces a Gaussian approximation for evaluating the \(H\) matrix, and uses slightly different friction coefficients. Therefore, it seems that as long as a proper choice is available for the potential function and for calculating the \(HA\) matrix, the virtual bond motions provide a reasonable description of the side chain motions. However, the averages for virtual bonds 4 and 5 and for backbone bonds 12 and 13 involve averages for pairs of bonds with quite disparate correlation times. Thus, the agreement with experiment for the two simpler models may be purely coincidental, and more study is required. Of course, it would be desirable to extend the MD simulation such that the bond relaxation rates can be calculated directly from the MD trajectory and compared with the model calculations to test their self-consistency. Unfortunately, this procedure would require a trajectory of at least 2 to 3 times longer than the typical bond rotational time, i.e., about 200–300 days on a TITAN or 10 days on a Cray XMP computer. The treatment of the dynamics of the full ACTH molecule would require one order of magnitude more computer time to account for the longer relaxation times as well as at least another two orders of magnitude because of the considerably larger size of the system. These computer time estimates underscore the need for theories, such as ours, of long-time protein dynamics.

4.2. Time dependence of the correlation functions

Figs. 7, 8 and 9, respectively, present the time evolution of rotational relaxation functions \(P_2(t)\) for the bonds that are used to describe the tryptophan \(1L_a\) dipole relaxation within the three models, namely, the virtual bonds 4 and 5, backbone bonds 12 and 13, and side chain bonds 35 and 36. Double exponential fits to some of these \(P_2(t)\) functions are given in table 1. As the model includes more detail, the long components of the decay become shortened. Paralleling this trend is a large decrease of the amplitude of the short component, although the rates themselves are not changed much. The shortening of the long component arises because the clumping together of sev-
eral beads into an effective unit leads to an overestimation of the overall friction when using the approximation of additive friction coefficients. As an example, consider the case of a Gaussian chain with $n$ beads of friction coefficient $\zeta$. The total friction coefficient from the Rouse–Zimm model is proportional to $n^{1/2}$, but for the clumped single effective unit in the additive friction coefficient approximation (the Rouse model) it is proportional to $n$ [4]. The clumped case, therefore, clearly gives too much friction for long chains. Hence, relaxation in the clumped model is slowed down relative to the unclumped model. The hydrodynamic interactions, which are
absent in the clumped model, tend to decrease the friction by introducing correlations in the disturbance of the fluid flow by the individual portions of the polymer chain. One measure of the change in friction forces between the models is provided by the translational friction coefficients as evaluated using the three models. The translational friction coefficient $f$ is evaluated as $f = n \zeta \nu_0^{-1}$, where the average segment friction coefficient is defined in eq. (7) and $\nu_0$ is the lowest (translational mode) eigenvalue of the $H$ matrix. Using the approximation $Q_{35,1}^{-1} = n^{-1/2}$ and $\zeta = 0.25$ yields the expression

$$\nu_0 \approx \left( \frac{1}{n} \sum_{i} r_i \right) \left( \frac{1}{n} \sum_{i} \frac{1}{r_i} \right) + \frac{\zeta}{n} \sum_{i,j} \langle l/R_{ij} \rangle (1-\delta_{ij}),$$

whereupon we find $f = 0.59 \times 10^{-7}, 0.36 \times 10^{-7}$, and $0.33 \times 10^{-7}$ g sec$^{-1}$ for the virtual bond, backbone bond, and all-bond models, respectively.

The decreased relative weight of the short component of $P_3(t)$ in table 1 with increased degrees of freedom in the model is related to the neglect of memory functions in the less detailed models. The memory function generally has a shorter time decay than the correlation functions and describes the time-dependent friction caused by neglected degrees of freedom. Thus, introducing extra degrees of freedom effectively adds the omitted memory of the simpler models and produces the added short-time friction in the more detailed models. Only when there is a clear separation of time scales between the correlation times for the bath degrees of freedom and those for the system degrees of freedom, can the influence of the bath be approximated by a constant friction (Markovian approximation). Otherwise, omitting memory functions or explicit treatment of the slow bath degrees of freedom may lead to a serious distortion of the short-time dynamics, although the longtime behavior may remain intact. However, a complete description would not lead to perfect exponential decays for the protein dynamics (c.f. eqs. (12), (14), and (15)). The fact that the fast dynamics in virtual bond 5 is not seen to have a more exponential decay is partly associated with the extra flexibility of bonds near chain ends.

Eqs. (12), (14), and (15) demonstrate that the time evolution of $P_3(t)$ is determined not only by the eigenvalues and eigenvectors of the product matrix $H A$, but also by the bond position in the protein. A certain bond $i$ may have the quantity $(Q_{i,a} - Q_{i,a-1})$ enter with a sufficiently large weight for a faster nor-

Table 1

<table>
<thead>
<tr>
<th>Bond Type</th>
<th>Virtual bond 4</th>
<th>Backbone bond 12</th>
<th>Side chain bond 35</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>model 1</td>
<td>model 2</td>
<td>model 3</td>
</tr>
<tr>
<td>$a_1$</td>
<td>0.59</td>
<td>0.60</td>
<td>0.22</td>
</tr>
<tr>
<td>$\tau_1$</td>
<td>62</td>
<td>41</td>
<td>53</td>
</tr>
<tr>
<td>$a_2$</td>
<td>0.37</td>
<td>0.36</td>
<td>0.74</td>
</tr>
<tr>
<td>$\tau_2$</td>
<td>572</td>
<td>443</td>
<td>338</td>
</tr>
</tbody>
</table>

* Fit is to $P_3(t) = a_1 \exp(-t/\tau_1) + a_2 \exp(-t/\tau_2)$, $\tau_1$ and $\tau_2$ are in ps. The sum $a_1 + a_2$ is slightly less than unity because of the imperfect double exponential fit to $P_3(t)$.

b) Model 1: virtual bonds; model 2: backbone bonds; and model 3: all bonds.
The computed range of normal mode relaxation times for the all-bond model spans about 5 orders of magnitude with the fastest one being 13 fs, whereas the fastest normal mode in the virtual bond model, and therefore the fastest possible bond relaxation rate in this model, is only about 4 ps. The fastest modes of the all-bond model are very crude representations of molecular vibrations; memory functions or alternative formulations are required to describe properly such fast motions. The longer time scale motions describe dihedral transitions of lower probability. Two of these normal modes have relaxation times in excess of the tryptophan orientational correlation time. Our theory is designed for the much more difficult long-time dynamics. Despite the inherent differences between the three models of reduced description studied here, the general similarity of relaxation time scales between the three models suggests the relative unimportance of memory functions for the longer-time dynamics, especially for the most complete all-bond model which is basically as simple to apply as the crude virtual bond model. Of course, a more complete assessment will require appropriate computation of these memory functions from the MD trajectory, a rather nontrivial problem.

5. Concluding remarks

While the 90 ps MD simulation is too short to provide highly accurate equilibrium averages, especially for the H matrix, our comparison between the three different models is independent of any limitations due to a short MD trajectory. In effect, the 90 ps trajectory produces equilibrium averages for these models that may correspond to potentials differing slightly from those in the CHARMM package; they are quite realistic for proteins and may be no more deficient in this regard than the approximate CHARMM potentials themselves. Because the 90 ps MD simulation does not exhibit dihedral transitions other than those involving local fluctuations, a 800 ps Brownian dynamics simulation has been performed using CHARMM (and omitting the solvent). This trajectory has three large-scale dihedral transitions. Since introduction of the water molecules tends to deepen potential wells [19], it follows that the time scales for these large dihedral motions are significantly longer and, hence, considerably exceed the tryptophan orientational correlation time. These features are reflected also in the all-bond model which has two normal mode relaxation times that exceed 1 ns. Thus the shortness of the 90 ps MD trajectory may not be as severe a limitation in the accuracy of computing equilibrium averages and thus of comparisons with experiment as is naively apparent.

Our investigation of the role of internal degrees of freedom in peptide dynamics leads us to the following conclusions:

1) The side group motion is on the same time scale as that of the backbone. The motions of the side chains couple to that of the backbone through the memory function and the overall friction. Thus, in general, side chain motions cannot be ignored or simply included in the bath without the use of a memory function. However, the neglect of the memory function mainly influences the short-time dynamics. Its influence on the overall relaxation rate is partially compensated by the change in the overall friction resulting from the additivity model for the group friction coefficients.

2) The agreement between the calculated relaxation time for the tryptophan fluorescence anisotropy in the all-bond model and the experimental value suggests that, for slow dynamics, our extended ORZ model is realistic provided accurate potential functions and a model with sufficient degrees of freedom are used. Further detailed comparisons of theory and experiment will be valuable.

3) In order to extend the theory to arbitrary peptides and allow predictions to be made, a data bank of the appropriate equilibrium properties must be accumulated. Monte Carlo techniques might appear to provide a more powerful method of obtaining the input H and A matrices for the calculations, but similar computer times are required [20]. This probably arises because shorter Monte Carlo time steps are necessary to avoid rejection of a large fraction of trial moves.

---

81 This feature is exhibited by computations with a model containing Gaussian side group motions (see ref. [18]).
Acknowledgement

We thank Martin Karplus and Ron Elber for providing us with the CHARMM package. We also thank Stuart Rice for helpful discussions. This work is supported, in part, by grants from the National Science Foundation.

Appendix A. Extension of theory to branched chains and to bonds of unequal average lengths

The MD trajectory allows a direct calculation for the equilibrium average bond vector correlation function

$$\overline{U}_{ij} = \langle l_i \cdot l_j \rangle, \quad i,j = 1, \ldots, n-1,$$

which for \( i=j \) becomes the mean-square length of bond \( l_i \) (that may be a virtual bond, backbone bond, or side chain bond). \( n \) is again the total number of atoms or groups of atoms whose motion are described by the Langevin model of eq. (1). In order to extend the treatment of ref. [3] to situations where the \( \langle l_i^2 \rangle \) are unequal, it is convenient to define an average mean-square bond length as

$$l^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} \overline{U}_{ii}^{-1}.$$

The matrix \( \mathbf{U} \) of eq. (2) may then be written as

$$\mathbf{U}_{ij}^{-1} = \frac{1}{l^2} \overline{U}_{ij}^{-1}.$$

While the construction of the \( \mathbf{H} \) and \( \mathbf{A} \) matrices strictly follows from their definitions (3) and (6) with the new meaning of \( l^2 \) (and \( l \)) given by eq. (20), the position-to-bond transformation matrix \( \mathbf{M} \) now assumes a new form to describe the branched side chains,

$$M_{ii} = n^{-1}, \quad i = 1, \ldots, n;$$

$$M_{ij} = 1, \quad i = 2, \ldots, n;$$

$$M_{ij} = -1, \quad \text{for } i > j \text{ and if the atoms or groups of atoms numbered } i \text{ and } j \text{ are directly connected};$$

$$M_{ij} = 0, \quad \text{otherwise}.$$

These modifications permit evaluation of the dynamic quantities \( P_a^2(t) \) and \( T_a^2 \) for each bond in the backbone bond and all-bond models described in fig. 1.

Figs. 6 and 7 compare relaxation times and correlation functions for the virtual bonds in the backbone bond and all-bond models with those in the virtual bond model. Our general dynamical theory enables the exact solution of eq. (1) for the relaxation of any vector \( l^a \) joining any pair of atoms in the protein as follows: the Langevin equation (1) is first transformed to normal coordinates \( \{\xi_a\} \) and then is solved for the dynamics of these normal coordinates. Any vector \( l^a = R_{i_2} - R_{i_1} \) between two atoms \( i_1 \) and \( i_2 \) may be expressed in terms of normal coordinates as

$$l^a = \sum_{a=1}^{n-1} (Q_{1,a} - Q_{2,a})^2 \xi_a.$$

Following the procedure of ref. [11], define the fundamental time correlation function of the vector \( l^a \) as

$$M_\ell(t) = \langle l^a(t) \cdot l^a(0) \rangle / \langle |l^a|^2 \rangle.$$

It follows directly from eq. (23) and the exact solution of eq. (1) for the normal dynamics that

$$M_\ell(t) = \sum_{a=1}^{n-1} (Q_{1,a} - Q_{2,a})^2 \xi_a^{-1} \times \left( \sum_{\Gamma_a} \mathbf{U}_{\Gamma_a}^{-1} \right)^{-1} \exp(-\sigma \lambda_a t).$$

with \( \Gamma_a \) the set of atom (or group) indices spanning the bonds whose vector sum gives \( l^a \). Finally, \( P_a^2(t) \) and \( T_a^2 \) are calculated from eqs. (14) and (16) using the definition in eq. (24) for \( M_\ell(t) \).

Appendix B. Estimation of errors involved in computing equilibrium averages using molecular dynamics trajectories of finite duration

The following derivation is patterned after one by Zwanzig and Ailawadi [8] for errors in time correlation functions.

Consider a dynamic variable \( A(t) \). Its average value over a trajectory of finite duration \( T \) is defined as
\[ A_T = \frac{1}{T^2} \int_0^T dt A(t). \]  

(26)

The error in using \( A_T \) in place of its equilibrium average \( \bar{A} \) is defined as usual through

\[ \langle \Delta_T^2 \rangle = \langle (A_T - \bar{A})^2 \rangle = \langle A_T^2 \rangle - \bar{A}^2 \]

\[ = \frac{1}{T^2} \int_0^T dt \int_0^T dt' \langle A(t')A(t'') \rangle - \bar{A}^2 \]

\[ = \frac{1}{T^2} \int_0^T dt' \int_0^{T-t'} ds \langle A(t')A(t'+s) \rangle - \bar{A}^2 \]

\[ = \frac{2}{T^2} \int_0^T ds \langle A(0)A(s) \rangle (T-s) - \bar{A}^2. \]  

(27)

Define the correlation function \( C(s) = \langle (A(0) - \bar{A})(A(s) - \bar{A}) \rangle = \langle A(0)A(s) \rangle - \bar{A}^2 \) and insert this into eq. (27) to give

\[ \langle \Delta_T^2 \rangle = \frac{2}{T^2} \int_0^T ds [C(s) + \bar{A}^2] (T-s) - \bar{A}^2 \]

\[ = \frac{2}{T^2} \int_0^T ds C(s) (T-s). \]  

(28)

Assume that the correlation function \( C(s) \) decays exponentially with an average relaxation time of \( \tau \). Therefore, the standard deviation in eq. (28) becomes

\[ \sigma^2 = \frac{\langle \Delta_T^2 \rangle}{\bar{A}^2} \]

\[ = 2 \left[ \frac{\langle A^2 \rangle - \bar{A}^2}{\bar{A}^2} \right] \left[ x^2 \exp \left( -\frac{1}{x} \right) + x - x^2 \right], \]  

(29)

where \( x = \tau/T \) and \( \langle A^2 \rangle \) is the equilibrium average of \( A^2 \).

References


W.F. van Gunsteren and M. Karplus, Biochemistry 21 (1982) 2259;


R.A. Engh, PhD thesis, the University of Chicago, 1987;
