# Molecular Dynamics of Conformational Substates for a Simplified Protein Model

Helmut Grubmüller and Paul Tavan
Institut für Medizinische Optik, Theoretische Biophysik
Ludwig-Maximilians-Universität München
Theresienstraße 37, D-80333 München, Germany

email: Helmut.Grubmueller@physik.uni-muenchen.de

**Abstract**

Extended molecular dynamics simulations covering a total of $0.232\,\mu$s have been carried out on a simplified protein model. Despite its simplified structure, that model exhibits properties similar to those of more realistic protein models. In particular, the model was found to undergo transitions between conformational substates at a time scale of several hundred picoseconds. The computed trajectories turned out to be sufficiently long as to permit a statistical analysis of that conformational dynamics.

To check whether effective descriptions neglecting memory effects can reproduce the observed conformational dynamics, two stochastic models were studied: a one-dimensional Langevin effective potential model derived by elimination of sub-picosecond dynamical processes could not describe the observed conformational transition rates. In contrast, a simple Markov model describing the transitions between but neglecting dynamical processes within conformational substates reproduced the observed distribution of first passage times. These findings suggest, that protein dynamics generally does not exhibit memory effects at time scales above a few hundred picoseconds, but confirms the existence of memory effects at a picosecond time scale.

# 1 Introduction

Knowledge about detailed atomic structures of biological macromolecules has been rapidly accumulated in recent years (see, e.g., Refs. 1,2). That progress opens the chance to acquire an understanding of macromolecular biological function in terms of basic physical and chemical notions. Many aspects particularly of protein function are known to be connected to dynamical processes within these macromolecules.[3–5] Therefore, adequate descriptions of that molecular dynamics (MD) are required and represent essential clues in the attempt to derive function from structure. Due to the structural complexity of proteins and a corresponding lack of well-founded coarse-grained effective models for the dynamics, the method of MD-simulation[6,7] currently is the only approach, to which some reliability can be assigned. That method conceives a macromolecule as a classical many-body system of 'atoms' and describes the quantum-mechanical forces like the chemical binding forces, which are caused by the electronic degrees of freedom, by a semi-empirical force field. Accordingly, the molecular dynamics is simulated by integration of the Newtonian equations of motion.

The enormous computational task associated with MD-simulation of biological macromolecules entails an upper limit to the time scale of dynamical processes accessible by this method: the MD-simulation of one nanosecond ($10^{-9}$ s) of an average-sized system consisting of 30,000 atoms requires roughly $2 \cdot 10^{16}$ floating point operations if all long-range interactions are taken into account. About 200 days of CPU-time on a 1 GFLOPS[a]-supercomputer are necessary to perform that task. Hence, the limit of accessible time scales set by current computer technology is in the nanosecond range.

However, many biochemical processes occur at time scales, which are by six to twelve orders of magnitude larger than that limit: typical ligand binding reactions as well as quaternary rearrangements occur in the range $10^{-3}$ to $10^{-1}$ seconds; protein aggregation and protein folding processes require up to $10^3$ seconds.[8] Admittedly, enormous efforts have been spent to increase the computational performance of MD including efficient implementations of MD-codes on vector-machines (e.g. Ref. 9) or, more recently, on parallel computers.[10–12] However, assuming that such efforts generate an increase of processing capabilities at about the rate of $10^3$ every ten years one is forced to the conclusion that computer technology will not allow MD-descriptions of many important biochemical processes before the year 2030 (cf. also Ref. 13).

At present, a *reduction* of the amount of computation involved in the description of protein dynamics is the prerequisite to further extend the range of accessible biochemical processes. Accordingly, various techniques have been developed and employed among which three main approaches can be distinguished:

---

[a] $10^9$ **FL**oating point **O**perations **P**er **S**econd

(a) *Mathematical and numerical methods* attempt to reduce the amount of necessary computation essentially *without* any modification of the physical description. These methods include higher order integration algorithms,[14] the use of generalized internal coordinates,[15,16] symplectic integration algorithms,[14,17] fast multipole methods,[18,19] variable time step methods,[20,21] and various multiple time step methods.[20,22–26]

(b) Proper *approximations* modify the molecular model employed in MD-simulations in a way that enables a reduction of the computational task. Here, care has to be taken to ensure, that the approximations do not too seriously alter the physics of the macromolecular dynamics.[27] Examples are the neglect of long-range interactions by use of a "cut-off" function,[23] the suppression of fast degrees of freedom,[28] and the so-called "mass-tensor" molecular dynamics.[29]

(c) *Effective models* are designed to replace the original MD-model. They rest on a classification of 'relevant' vs. 'irrelevant' system properties for a given dynamical process. These models reduce the explicit description to the relevant properties and assume that the action of the irrelevant properties can be implicitly taken into account by renormalized interactions or other quantities representing statistical averages. Stochastic models, in particular, are based on the assumption, that the detailed dynamics of fast degrees of freedom, such as bond- or bond angle vibrations, is not essential for protein structure and function. Successful applications of stochastic descriptions like Monte Carlo simulation,[30,31] Langevin dynamics,[32] generalized Langevin dynamics with memory friction (e.g. Ref. 33) or the use of statistical potentials,[34] support that assumption.

Certainly, this classification is not mutually exclusive which becomes obvious by considering the neglect of the long-range part of the Coulomb interaction as an example: that method equally can be regarded as an approximation and as an effective model implicitly accounting for shielding effects caused by atomic polarizabilities.

Most of the above methods have been designed for a wide class of many-body systems and, therefore, represent *general purpose* methods. Application of these methods to proteins typically speeds up MD-simulations by about one order of magnitude. However, in view of the desire to increase accessible time spans by six to twelve orders of magnitude, the efficiency gains achieved by these general purpose methods represent only a moderate success.

Major efficiency gains can be expected if computational methods and effective models are developed which more specifically take advantage of structural and dynamical properties *particular to proteins*. That expectation rests on the emerging notion that proteins actually possess unique properties which distinguish these many-body systems from others.[35] In particular, a clear-cut identification of irrelevant degrees of freedom, explicit consideration of which usually is computationally demanding, should allow considerable efficiency gains by development of coarse-grained effective models, which are adjusted to the particular dynamical and structural properties of proteins.

From the above discussion we conclude, that a proper characterization of protein dynamics is a prerequisite for the development of efficient protein dynamics descriptions. We note that a separation of relevant observables from irrelevant ones is also required for an application oriented *evaluation* of a given MD-method. Usually, such an evaluation is based on a comparison of certain quantities computed from test-simulations carried out with the given MD-method, with corresponding quantities obtained from simulations employing a reference method, which is assumed to provide more accurate results (cf. Refs. 36, 37). However, for an *application oriented* evaluation, the quality of the given method should be evaluated solely with respect to its ability to describe *relevant* properties accurately. These ideas are discussed and exemplified in detail in a forthcoming paper[38] as well as in.[26]

In the present paper we focus on the question, how knowledge on the very special dynamical properties of proteins can be acquired. Below we will motivate the hypothesis, that studies of the dynamics of simplified protein models are well suited to contribute to such knowledge, provided the dynamical properties of the protein model can be shown to be sufficiently similar to those of more realistic protein models.

Contrary to less complex many-body systems the dynamics of proteins appears to involve a hierarchy of time scales.[39,40] The high-frequency dynamics of protein models has been examined in detail by means of MD-simulation[41,42] as well as by normal mode analysis,[43] whereas knowledge about the low-frequency dynamics is sparse. However, many quantities which are important to protein function are defined only at slow time scales well above 100 picoseconds: mean first passage times for transitions between conformational substates,[44,4] which are considered as elementary steps for ‘functionally important motions’,[45] fall in that region. Computations of corresponding transition rates, e.g. by transition state or activated dynamics,[5,46] or of other relevant quantities, like free energies, typically require large statistical ensembles and, therefore, show slow convergence. Accordingly, studies of infrequent conformational motions have been possible only for small polypeptides.[47–50] Typically, the time scale covered by available sampling techniques like umbrella sampling[51] or various force-bias methods[31,52] is too short as to provide an ensemble large enough for accurate results. Hence a characterization of protein dynamics is required especially in the low-frequency region.

At the first glance, this requirement appears to entail a vicious circle which impedes the development of efficient protein dynamics descriptions: On the one hand, due to the structural complexity of the system as well as due to the lack of experimental data, studies of dynamical properties in the low-frequency region have to rely on extended MD-simulations. On the other hand, it is difficult to carry out such simulations unless sufficiently efficient protein dynamics descriptions have been developed.

In view of that problem we note that insights into slow phenomena of protein dynamics, such as the folding process, have been provided by studies of small, oversimplified protein models, such as lattice models[53] or 'bead'-models.[54] Of course, the simplified structure of such model systems requires a careful interpretation of results in order to provide information on properties of real proteins. But at the same time their simplicity entails the key advantage of such systems which is to permit extended simulations covering time spans several orders of magnitude larger than those accessible to simulations on more realistic protein models. Hence, analysis of the dynamics of a simplified protein model by means of extended MD-simulations should enable insights into dynamical properties of proteins and, therefore, should contribute to the development of more efficient protein dynamics descriptions.

The present paper exemplifies that approach by considering a 'minimal model' for proteins, which is described in the following Section. Despite its simplified structure, MD-simulations carried out on this protein model reveal dynamical properties similar to those computed from MD-simulations of more realistic, complex protein models or to those obtained from experiments (Section 3). In particular, the dynamics of the protein model was found to exhibit conformational transitions at a time scale of several hundred picoseconds. Such conformational dynamics appears to be ubiquitous in protein dynamics.[55,56] As we will argue, these similarities support the assumption that results obtained by our extended MD-simulations of the simplified protein model actually provide information about the low-frequency dynamics of real proteins.

In Section 4 we analyze whether memory-effects are present in the dynamics of our model. Memory-effects of dynamic quantities are well known to exist in proteins at short time scales up to the picosecond range.[33] Here they give rise to non-vanishing autocorrelation functions of atomic positions or velocities.[57] Proper consideration of these short time correlations is essential for stochastic descriptions of fast degrees of freedom.[58] Accordingly, the development of coarse-grained effective descriptions of slow degrees of freedom requires knowledge about time scales of correlations. To contribute to such knowledge, we address the question, to what extent memory effects show up in the low-frequency conformational dynamics of proteins.

## 2 Design of a simplified protein model

To enable studies of low-frequency protein dynamics by extended MD-simulations of a *simplified* protein model, the design of that model has to meet two main requirements: (a) the model should exhibit structural and dynamical properties similar to those of real proteins, in particular, it should enable a study of conformational transitions; (b) the model should exhibit as few degrees of freedom as possible, and, accordingly, should be structurally simple. Thus, the protein model should represent a 'minimal' model. Such model can not contain *all* features of proteins, nor can it model the dynamics of a *specific* protein. Therefore, we included only those structural elements, the combination of which a priori seemed to be essential for low-frequency dynamical properties.

For the design of the protein model we have chosen a two step procedure: first, we defined a 'primary' structure, consisting of 100 residues, and a force field. Second, we simulated a 'folding' process of that model in order to obtain a stable tertiary structure.

To optimally meet the above conflicting objectives, we decided to neglect the internal structure of the residues and to describe the polypeptide by a chain of 100 van der Waals spheres, which are linearly connected via interactions resembling chemical bonds. The employed force field included bond stretch, bond angle, van der Waals and Coulomb interactions[b]:

$$E = E_b + E_\Theta + E_{\mathrm{vdW}} + E_{\mathrm{el}}\,,$$

where the energy contributions are defined as in Ref. 23. The particle masses and force parameters were those of $CH_2$ 'extended atoms' and associated single bonds, as defined in the CHARMm force field.[23]

With the above definitions, the model describes a 100-alkane ('hektane') rather than a protein. Therefore, additional properties which can mimic the low-frequency behavior of proteins have to be included. In order to identify such properties we note that a characteristic feature of proteins is their ability to fold into and to maintain a unique tertiary structure[c] in native environment.[59] That property is a prerequisite for their specific biochemical function. The tertiary structure is determined by specific interactions of particular amino acid side groups, e.g. by disulfide bonds or by H-bonds as well as by less specific long-range interactions like Coulomb or hydrophilic and hydrophobic interactions. The latter type of interaction, in particular the hydrophobic force, is known to dominantly contribute to the stability of folded proteins.[60,61] We therefore decided to add heterogeneous long-range interactions to our protein model.

[b] We did not include an angle torsion potential.

[c] More accurately, that is a set of mutually similar conformational substates.

*Primary structure:*

For that purpose we defined an artificial, heterogeneous 'primary structure' by assigning different partial charges to the 100 van der Waals spheres of the model. Figure 1 shows the chosen charge distribution (bold, wavy curve) along the stretched chain. As can be seen, that distribution divides the protein model into five parts, three of which carry a positive charge, while the remaining two are charged negatively. The inset of Fig. 1 shows the detailed structure of the model.

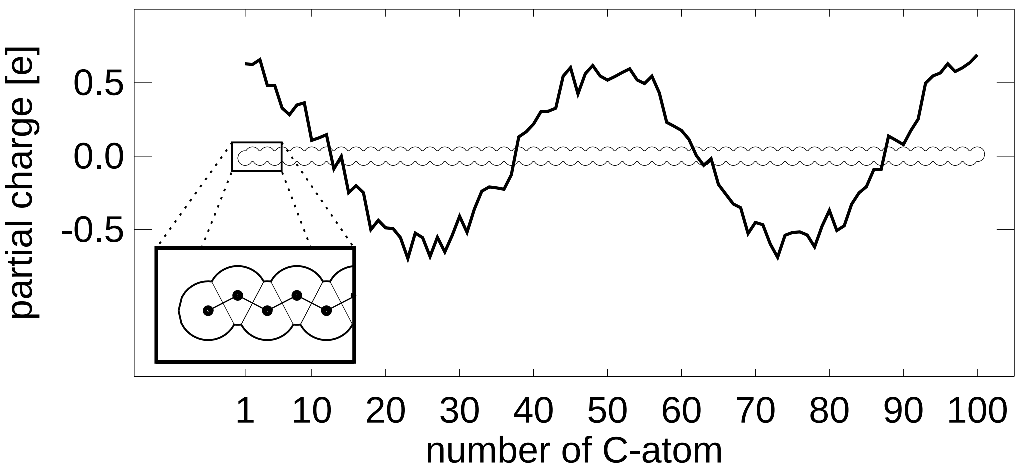


Figure 1: Protein model in a stretched, unfolded configuration, consisting of 100 $CH_2$-like 'atoms'; their partial charges are represented by the bold curve; the inset shows a zoom of the detailed structure.

The average absolute charge was chosen such that the corresponding Coulomb energy contribution becomes sufficiently large as to over-compensate entropic free energy contributions at room temperature[d] and thus to stabilize folded structures.

In that respect, the chosen Coulomb interactions mimic hydrophobic protein-solvent interactions comparable to those included into the bead-model in Ref. 54: the attraction of oppositely charged 'residues' resembles hydrophobic forces, whereas the repulsion of equally charged 'residues' models the tendency of hydrophilic groups to solvate. We note, that, as a consequence of the above interpretation, the protein model comprises a simple effective solvent model.

*Folding Process:*

The second step — the 'folding' of the protein model — was carried out

[d]We chose an average absolute charge of 0.25 e, which, accidentally, corresponds to the average value found in more realistic protein models.

by means of MD-simulation. All simulations presented in this paper have been performed *in vacuo* using the Verlet-algorithm[36] with an integration step size of one femtosecond for the integration of the Newtonian equations of motion. No 'cut-off' has been employed.

Starting from an unfolded configuration, shown in Fig. 1 and, as a ribbon-plot, in Fig. 2 (a), the protein model was allowed to freely move under the influence of bond-, van der Waals-, and Coulomb interactions. Figure 2 (b–f) shows snapshots of the structure during this initial phase of the 'folding process'. As can be seen, the compactness of the model rapidly increased and, after 10 picoseconds, came close to its final value. At this stage, the tertiary structure had not yet stabilized, and the model exhibited frequent conformational changes.

Within the first few femtoseconds of the folding process, the temperature of the model raised from 300 K to above 10,000 K. This large temperature jump is due to high conformational energy present in the initial structure, part of which quickly converted into kinetic energy. We continued this high temperature dynamics for two nanoseconds to explore configurational space. By rescaling of atomic velocities, the model was then slowly cooled down to 300 K, where it was trapped in a well-defined conformation, which remained stable during the subsequent equilibration phase of one nanosecond duration. The model was then allowed to move freely again for two nanoseconds in order to approach thermal equilibrium. This has been achieved, as is indicated by the fact that no temperature drift could be observed during that period. The resulting, folded and relaxed conformation is depicted in Fig. 2 (g). This structure was used as initial configuration for the simulations described in the following Sections.

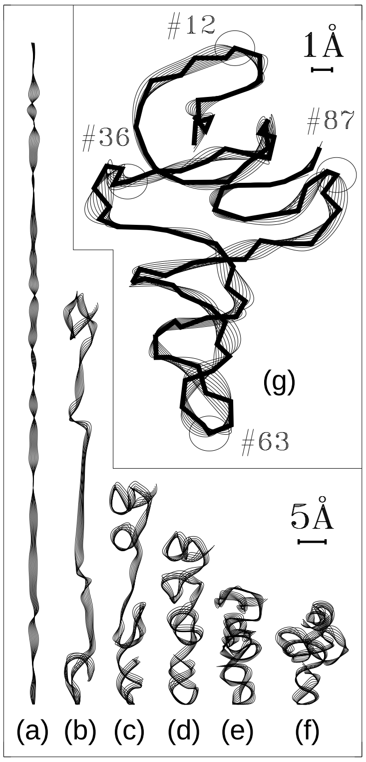


Figure 2: Snapshots of the protein structure during the simulated folding process described in the text; structures are shown as 'ribbon-plots' (a)–(f): initial, completely unfolded configuration (a); successive configurations after 3 ps (b), 6 ps (c), 7 ps (d), 8 ps (e), and 100 ps (f), respectively; (g): folded structure of the protein model after 5 ns, which was used as initial configuration for the dynamics simulations described in the text; the bold lines represent chemical bonds; four numbered circles mark atoms referred to in the text.

# 3 Relevance of the Simplified Protein Model

The severe simplifications inherent to the protein model described above enforce a close inspection concerning its relation to real proteins as well as a number of caveats with respect to the interpretation of the MD-studies described below. In particular we want to check in what sense our study of the low-frequency configurational dynamics of our simplified model can serve as a tool to characterize the corresponding dynamics of real proteins. To that aim we will first discuss some of the properties determining the shape of the energy and free energy landscape, respectively. Subsequently, we will check, whether the model fulfills our expectations.

The neglect of any internal structure of the 100 residues which make up the protein model, represents the most obvious simplification. Since the residues are represented by van der Waals spheres, most short-range, residue-specific interactions, such as H-bonds, which contribute to the formation of rigid secondary structure elements in proteins, are absent in our model. As a consequence, the model structure is expected to exhibit larger flexibility as compared to proteins. An additional increase of flexibility should arise from the absence of side-groups and a corresponding lack of site-specific sterical restraints. Thus, free energy barriers for conformational transitions in our model are expected to be lower than in real proteins. The neglect of angle torsion barriers should have similar consequences.

Secondly, the residue masses have been chosen smaller than those of amino acids by about one order of magnitude. However, this difference does not affect the *quality* of the dynamics: according to Newton's laws, scaling of masses merely corresponds to a shift of time scales — in the present case by a factor of three to four. As a result, also conformational motions of the model will occur at a correspondingly faster time scale as compared to proteins.

We expect that as a result of both effects, the reduction of free energy barriers as well as the time scaling of the dynamics of the simplified model, the number of conformational transitions, which occur in the course of our simulations is large enough as to permit their statistical analysis.

Finally, our simple effective solvent model does not include stochastic and frictional forces, which are exerted by solvent molecules onto protein surfaces. Comparisons of MD-simulations of protein models *in vacuo* with simulations in solvent have shown, that such solvent-induced forces reduce the inertial character of vibrational modes in proteins and decorrelate these motions.[57] Accordingly, our *in vacuo* simulations should exhibit a much slower decay of the short-time displacement autocorrelation functions of surface residues than the one determined in more realistic protein-solvent models, i.e., the simplifications should enhance memory effects at fast time scales.

The folded structure of our model (Fig. 2 (g)) seems to be built up from structural motifs comparable to secondary structure elements found in proteins, such as a 'helix' at the bottom of the model as well as 'loops' at the left and right sides and at the top. The combination of these motifs resembles typical tertiary structures of small globular proteins (compare, e.g., the structure of Crambin, as shown in Ref. 62). This structural similarity has not explicitly been put into our model, but, instead, results from the particular choice of chain-chain interactions. Analogously, we did expect that also realistic *dynamical* properties should appear as a consequence of the model design.

To confirm that expectation, we computed various properties, the combination of which is known to be characteristic to protein dynamics, and compared them with simulations of more realistic protein models or experiments. In our analysis, we will proceed from short time scales (femtoseconds) to longer ones (nanoseconds).

The short-range interactions, which determine the high-frequency dynamics of the protein model, have been chosen in close correspondence to those of hydrocarbons, as defined by the CHARMm force field.[23] Hence, the protein model should exhibit reasonably realistic dynamical properties at a time scale below some 100 femtoseconds.

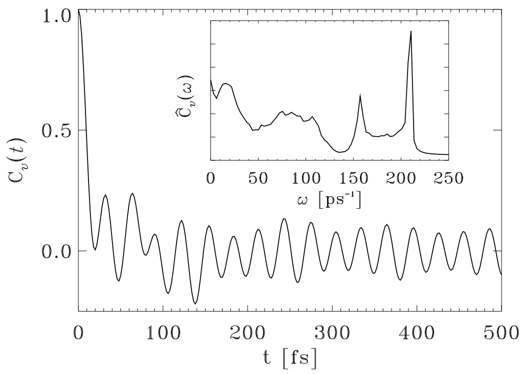


Figure 3: Normalized velocity autocorrelation function $C_v(t) = \langle \mathbf{v}(0) \cdot \mathbf{v}(t) \rangle / \langle \mathbf{v}(0)^2 \rangle$ derived from an average over a one nanosecond trajectory using the velocity vector $\mathbf{v}$ of one particular atom of the protein model. The inset shows the corresponding spectrum, computed using a Fourier transform of the velocity autocorrelation function. This spectrum is comparable to that of more realistic protein models: The two sharp peaks originate from fast bond-stretch vibrations, whereas the broad bands in the low-frequency region of the spectrum are predominantly caused by the stochastic character of inter-atomic van der Waals contacts.

As an example, Figure 3 shows the normalized velocity autocorrelation function $C(t) = \langle \mathbf{v(0)} \cdot \mathbf{v(t)} \rangle / \langle \mathbf{v(0)}^2 \rangle$, derived from an average over a one-nanosecond trajectory using the velocity vector $\mathbf{v}$ of an arbitrarily chosen residue of the protein model. The inset shows the corresponding spectrum, which has been derived in a way similar to that employed in Ref. 57 in order to enable a comparison with the results of MD-simulations of a detailed model of RNAse-A presented therein. Similar features are sharp peaks in the range 150 to $250\mathrm{ps}^{-1}$, originating from fast bond stretch vibrations, as well as broad bands in the region below $150\mathrm{ps}^{-1}$, arising from bond angle vibrations and, particularly, the noisy character of van der Waals collisions. The dynamics of more realistic models differs from that of the simplified protein model in that it typically gives rise to a larger number of peaks in the high frequency spectrum, due to a heterogeneity of bond stretch frequencies, which is absent in our model.

These high frequency modes represent the first layer of a hierarchy of time scales in proteins, within each of which specific dynamical processes can be observed.[45,63,64] These range from bond stretching modes ($\approx$ 30 fs), bond angle- and dihedral vibrations (few 100 fs), collective motions involving groups of atoms (some 10 ps), to conformational transitions occurring within a wide range of time scales above 100 ps. Inspection of dynamical details at different resolutions in time has shown, that such hierarchy actually is reproduced by our simulations, which cover time scales differing by more than six orders of magnitude. As an example, Figure 4 shows the time development of the distance between two arbitrarily chosen residues at decreasing time resolutions. Proceeding within Fig. 4 from top to bottom, one observes characteristic fluctuations, which occur at time scales increasing at each step by a factor of ten and originate from the four different dynamical processes enumerated above.

A characteristic feature of protein dynamics is the existence of a variety of conformational substates, which are interconnected by conformational transitions. The model, too, exhibits such conformational transitions, which reveal themselves as sudden structural rearrangements. In particular, the rapid atomic distance changes apparent in the bottom three pictures of Fig. 4 are due to such conformational transitions.

From these observations we infer, that, despite its simplifications, our protein model exhibits a set of structural and dynamical properties, which are characteristic for proteins. In particular, the model should be rather well-suited for a simulation study of the long-time conformational dynamics of proteins.

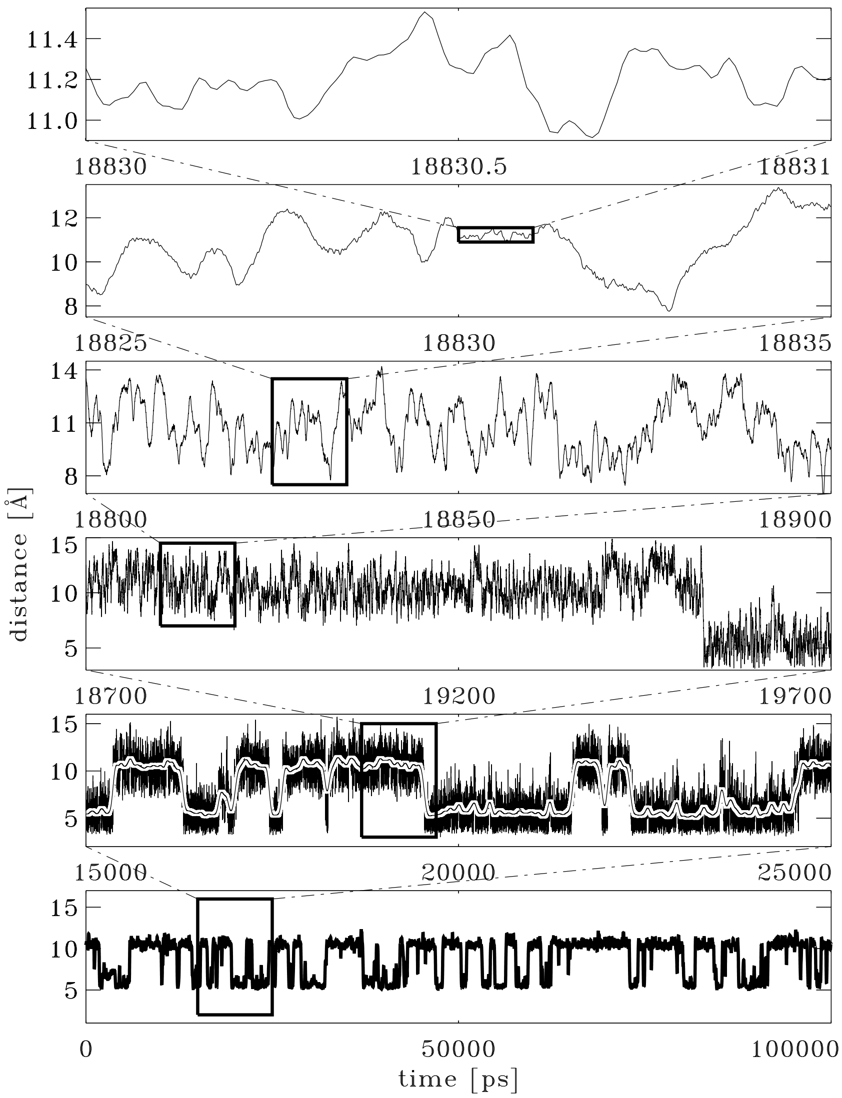


Figure 4: Distance (in Å) between residues # 12 and # 36 (cf. Fig. 2 (g)) during an MD-simulation of 100 nanoseconds length. The series of plots shows the distance fluctuations at time scales increasing by a factor of ten at each magnification step; the bold lines in the bottom two pictures represent smoothed data. Note that the depicted simulation has not been included within the analysis carried out in the following Section.

# 4 Conformational Dynamics

We studied the conformational dynamics of our protein model by means of extended MD-simulations covering a total of 232 nanoseconds. Using the well-known Verlet-algorithm with an integration step size of one femtosecond, 232 simulations, each of one nanosecond duration, were carried out[e]. All 232 simulations started with almost identical initial conditions, derived from the structure resulting from the folding procedure described in Section 2 by minute random modifications of atomic positions ($10^{-6}$ Å). Nevertheless, these 232 simulations are essentially independent from each other, since the chaoticity inherent in the dynamics guarantees a rapid decorrelation of the initially similar trajectories within a few picoseconds.

An integration time step size as short as one femtosecond may seem to contradict the purpose of our simplified model, i.e., the *reduction* of the computational effort that has to be spent to study conformational dynamics. Note, however, that here the main computational speed up is due to the decrease of the number of degrees of freedom as compared to more realistic protein models. Furthermore, as described in the preceding Chapter, conformational transitions in our model are expected to be accelerated, implying *fewer* integration steps per conformational transition. As we shall see below, these expectations will be confirmed, in that the simulations actually will exhibit a large number of conformational transitions within the time span of our simulations.

## 4.1 Theory of Conformational Substates

Generally, a coarse-grained effective description of a (microcanonical) system with a Hamiltonian $\mathcal{H}(\mathbf{q}^N, \mathbf{p}^N)$, where $\mathbf{q}^N$ denote the $3N$ Cartesian coordinates of the $N$ particles within the system and $\mathbf{p}^N$ their momenta, can be achieved by explicitly considering the dynamics of only a few degrees of freedom $c_i(\mathbf{q}^N, \mathbf{p}^N)$, commonly referred to as 'conformational coordinates'.[65] If the remaining degrees of freedom are regarded as a heat bath, the resulting system of reduced dimension belongs to a *canonical ensemble*.

The free energy landscape of that sub-system is a *potential of mean force*,[66] $W(\mathbf{c})$, which determines, together with the heat bath, the dynamics of the conformational coordinates $\mathbf{c}$. By means of the *canonical* (projected) phase space density $\rho_c(\mathbf{c})$:

$$\rho_c(\mathbf{c}) = \int d\mathbf{q}'^N d\mathbf{p}'^N \rho(\mathbf{q}'^N, \mathbf{p}'^N)\delta(\mathbf{c} - \mathbf{c}') \,, \tag{1}$$

[e] Part of these 232 simulations were carried out using modified Verlet integration algorithms, as described in Refs. 24, 38. However, we do not consider these algorithmic differences to seriously affect our results.

$W(\mathbf{c})$ can be derived from the phase space density $\rho(\mathbf{q}^N, \mathbf{p}^N)$ generated by the dynamics of the *entire* system:

$$W(\mathbf{c}) = -k_b T \ln \rho_c(\mathbf{c})\,. \qquad (2)$$

As an illustration, Figure 5 shows a contour-plot of the free energy landscape of our protein model, which has been determined from $\rho_c(\mathbf{c}) = \rho_c(d_{12,36}, d_{12,87})$ according to (2). Here, the two distances $d_{12,36}$ and $d_{12,87}$ between atoms #12 and #36 as well as #12 and #87, respectively, have been chosen as conformational coordinates (cf. Fig. 2). The canonical ensemble, from which we derived $\rho_c(\mathbf{c})$, has been generated from the complete set of 232 1-ns-trajectories by recording both distances every 8 fs in the course of the simulations. Discrete values for $\rho_c(\mathbf{c})$ have then been determined on a grid of 0.5 Å resolution by computing a two-dimensional histogram using square bins of 0.5 Å side length. This choice represents a compromise between the conflicting aims of high resolution and low statistical fluctuations of frequency counts per bin. The contour plot in Fig. 5 represents a smoothened version of the two-dimensional histogram.

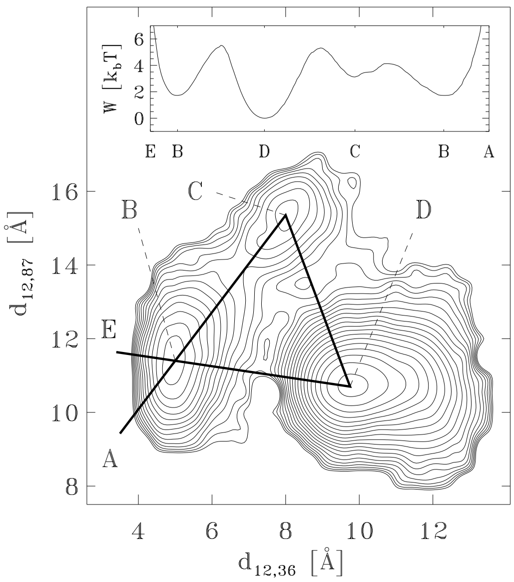


Figure 5: Contour-plot of the free energy landscape $W(d_{12,36}, d_{12,87})$ derived from a projection of the phase space density $\rho_c$ onto two conformational coordinates, namely the distances between atoms #12 and #36, and atoms #12 and #87, respectively (cf. Fig. 2). The inset shows the values of the energy landscape in units of $k_b T$ along a hypothetical reaction coordinate (bold line) connecting the points A-B-C-D-B-E.

The inset of Fig. 5 illustrates the shape of the free energy landscape $W$ by plotting its value in units of $k_b T$ (at $T = 300K$) along an arbitrarily chosen 'reaction coordinate' (bold line) passing through the three minima. $W$ has been interpolated using a cubic spline function.

The energy landscape in Fig. 5 exhibits three distinct minima, denoted as 'B', 'C', and 'D', respectively. As suggested on the basis of experimental data by Frauenfelder[4] such regions of low *free* energy in conformational space, separated by free energy barriers, generally define distinct *conformational substates* of a protein. Accordingly, we define our model to be in substate 'B', 'C', or 'D', respectively, if its conformation lies in the corresponding region of the free energy landscape. Like the Brownian motion of a particle coupled to a heat bath, the dynamics of the system within the energy landscape shown in Fig. 5 is diffusive. Occasionally, the fluctuating forces generated by the heat bath drive the system across one of the energy barriers and induce a conformational transition, which reveals itself as a rapid change in the sterical structure of the model. Such sudden structural transitions are, e.g., apparent in the lower parts of Fig. 4.

Note, that the above definition of conformational substates differs from the approach commonly employed for theoretical explorations of substate hierarchies.[67,68] In these studies, the distribution of thermally accessible local minima of *potential energy* within configurational space is studied, and it is assumed, that these local minima or clusters thereof can provide information on the distribution of conformational substates. At low temperatures, where entropic contributions are small and safely can be neglected, the *potential* energy landscape definitely can serve as a tool for the analysis of conformational substates. However, at room temperature the suggested relation between accessible minima of potential energy and conformational substates, defined as minima of *free* energy, is questionable. In contrast, our approach allows the study of conformational substates *at physiological temperatures*, as it refers to the *free* energy landscape within conformational space. Therefore, if applied to realistic protein models, our approach enables comparisons of theory and experiment. Admittedly, much larger computational effort is involved in such analysis, because a sufficiently dense sampling of phase space by extended simulations is required for the determination of $\rho_c$. At present, that computational effort restricts our method to studies of simplified protein models.

As can be seen in Fig. 4, the conformational substates of our protein model are stable on the time scale of several ten picoseconds; conformational transitions occur on scales above a few hundred picoseconds. In order to study, whether memory effects are present in the dynamics on these two time scales, we will now consider two simple stochastic models.

The first model will refer to and will be derived from the protein dynamics *within* one particular conformational substate. We choose a one-dimensional Langevin model, which describes the Brownian motion of a particle in a potential of mean force, $W(c)$. We will derive this effective potential from the simulation employing the same procedure that was used to compute the free energy landscape shown in Fig. 5. The second model serves to describe the conformational dynamics of transitions *between* the three conformational substates apparent in Fig. 5 on the longer time scale above several hundred picoseconds. We will choose a Markov model which neglects the protein dynamics within each substate.

## 4.2 A Langevin Model

We now try to describe the dynamics of our protein model in the vicinity of the energy minimum 'D' (cf. Fig. 5) as a Brownian motion. As a conformational coordinate, we chose the distance $c = d_{12,36}$, which allows to separate substate 'B' from substate 'D' (cf. Fig. 5). Since the dynamics within substate 'D' determines the transition rate from 'D' to 'B', comparison of this particular rate determined from the Langevin model with the corresponding rate observed in the simulation will provide a check whether that model is applicable.

The time evolution of the stochastic model is described by the Langevin equation:

$$m\ddot{c} = -m\gamma\dot{c} - \nabla W(c) + \xi(t)\,, \tag{3}$$

where $m$ denotes the effective mass[69] of a Brownian particle, the motion of which is governed by a heat bath a potential of mean force, $W(c)$. The influence of the heat bath is described by the friction coefficient $\gamma$ and a fluctuating force $\xi(t)$. The dots represent time derivatives. We assume, that $\xi(t)$ can be modelled by Gaussian (white) noise, and then we check, whether this assumption, which implies a neglect of memory effects, is correct.

Figure 6 (left) shows the potential of mean force, $W(c)$, which has been computed according to (1) and (2), respectively, where only those conformations, which belong to state D or B, have been used for the calculation of $\rho_c$ according to (1). The considerable length of the trajectory allowed a quite accurate determination of $W(c)$. As a check, we computed $W(c)$ using only half of the trajectory. Compared to the full statistics, no significant deviations were observed (data not shown).

In Fig. 6 the transition under consideration, $B \to D$, corresponds to a transition from the deep minimum across the energy barrier to the left. To calculate its rate from the Langevin model, the parameters $m$ and $\gamma$ have to be specified; the amplitude of $\xi(t)$ then follows from the dissipation fluctuation theorem.

As indicated by harmonic fits (dashed lines) in Fig. 6 (left), at the energy minimum and at the barrier top, the shown energy landscape can be described by a harmonic double well. Diffusive motion in such a double well potential can be described analytically and, therefore, enables a simple determination of the friction coefficient $\gamma$ as well as of the effective mass $m$, which enter into (3) as adjustable parameters, from our simulations. For that purpose, we make use of the velocity autocorrelation function, $C(t) = \langle \dot{c}(0) \cdot \dot{c}(t) \rangle$, computed from an average using a selected 1ns-trajectory, which did not leave state 'D'. In our harmonic approximation we assume $W(c) \approx \frac{1}{2} m \omega_0^2 c^2$ . $C(t)$ can then be derived analytically and $\gamma$ can be determined by comparison with the simulation. We used the velocity autocorrelation instead of a displacement correlation, because it relaxes more rapidly, its relaxation depends sensitively on the friction coefficient, and it is rather insensitive to the characteristics of the mean force potential.[57]

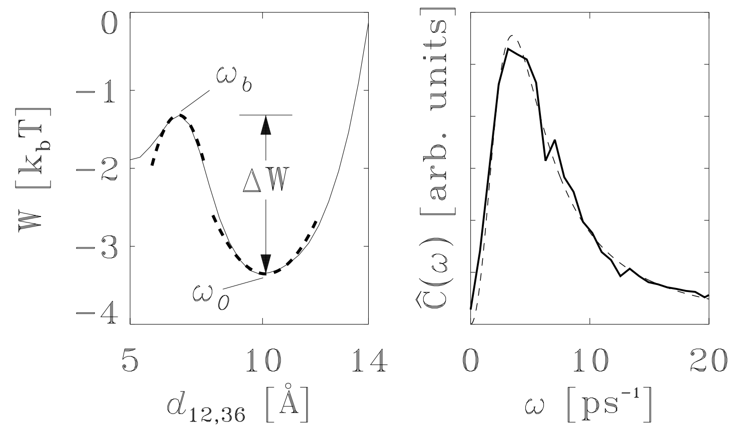


Figure 6: Left: potential of mean force $W(c)$ for the conformational coordinate $c$ in units of $k_b T$ (solid line); harmonic fits as described in the text are shown as dashed lines; right: low-frequency part of the spectrum $\hat{C}(\omega)$ of the autocorrelation function $C(t) = \langle \dot{c}(0) \cdot \dot{c}(t) \rangle$ derived from a 1-ns-trajectory (bold line); fit of an expression derived from a harmonic oscillator model (dashed line).

Figure 6 (right) shows the low-frequency part of the spectrum of $C(t)$ (solid line). In the diffusive harmonic oscillator model the spectral density of the velocity autocorrelation function is given by[69]

$$\hat{C}(\omega) \propto \frac{\gamma \omega^2}{(\omega_0^2 - \omega^2)^2 + \gamma^2 \omega^2} \, ,$$

provided that $\omega_0 > \gamma/2$, which is assumed to hold for the present application. A fit (dashed line) of this expression to the low-frequency part of the observed velocity autocorrelation spectrum yields $\gamma = 6.0 \, ps^{-1}$ and $\omega_0 = 3.5 \, ps^{-1}$. In the harmonic oscillator model, this corresponds to an effective mass of $m = 7.4$ atomic units. The excellent quality of the fit demonstrates the applicability of the harmonic oscillator model. Moreover, the obtained values for $\gamma$ and $\omega_0$, respectively, justify the assumption of moderate friction ($\omega_0 \gtrsim \gamma/2$).

An upper limit $k_{D \to B}$ for the transition rate in the case of moderate friction can be obtained using Kramers' theory,[70]

$$k_{D \to B} = \frac{\sqrt{\gamma^2/4 + \omega_b^2} - \gamma/2}{\omega_b} \, \frac{\omega_0}{2\pi} \, \exp(-\Delta W / k_b T) \, , \qquad (4)$$

where $-\frac{1}{2}m\omega_b^2c^2$ is a harmonic fit to the potential at the barrier top $b$ (cf. Fig. 6 (left)) and $\Delta W$ is the barrier height. With $\omega_b = 6.5\,\mathrm{ps}^{-1}$ one obtains $k_{D\rightarrow B} = 46.5\,\mathrm{ns}^{-1}$. However, this prediction largely overestimates the rate obtained from the MD-simulation, which is lower by a factor of nearly 25, namely $1.89\,\mathrm{ns}^{-1}$!

Obviously, the Langevin model could not reproduce the observed conformational transition rates. This failure must be attributed to the only unjustified assumption in that model, namely that of memory-free, Gaussian noise as a description of the influence of all degrees of freedom within the system on the dynamics of the conformational coordinate $c$. Thus one is forced to the conclusion, that memory effects, caused by correlations between many degrees of freedom, strongly influence the short time scale dynamics within conformational states at a picosecond time scale. One consequence of these memory effects is a 25-fold reduction of the particular transition rate under consideration.

A closer inspection of the MD-trajectories showed indeed a frequent crossing of the barrier top in Fig. 6 (left), as predicted by (4), but, in most cases, *without* a consecutive transition, i.e., the system 'remembered', where it came from. Presumably, a better estimate for the transition rate could be obtained by applying the method of reactive flux[22] (a review can be found in Ref. 46). However, this method requires the preparation of an ensemble near the barrier top. The question, whether this is possible *without* relying on extended conformational sampling, must be left open within the present paper.

## 4.3 A Markov Model

The question arises, whether memory-effects also show up at longer time scales, e.g., in the range of few 100 picoseconds. Viewed at that time scale, the dynamics of the protein model is characterized by transitions between the three distinct conformational states apparent in Fig. 5. Accordingly, we shall now neglect the fast protein dynamics *within* each conformational substate and instead focus on the (discrete) conformations dynamics governed by transitions *between* the three substates of our model. To study memory effects at that time scale we will check, whether the distribution of transition times between the conformational states can be described by a memory-free stochastic dynamics, i.e., we compare the observed conformational dynamics with the three-state continuous Markov process depicted in Figure 7.

This Markov process has three states $S \in \{B, C, D\}$, which shall represent the three conformational states defined in Fig. 5 and which are shown together with all possible transitions (arrows) and corresponding transition rates, $k_{ij}$ $(i, j \in \{B, C, D\})$. The probabilities $P_i(t)$ to find the model in state $S_t = i$ at time $t$

obey the master equation

$$\frac{d}{dt}P_i(t) = \sum_{j=B,C,D} k_{ij}P_j(t) - \sum_{j=B,C,D} k_{ji}P_i(t) \quad , \quad i = B, C, D\,.$$

The time-independent transition rates $k_{ij}$ are defined by the conditional probability of finding the system in state $i$ at any time $t = 0$, if it has been in state $j$ before at time $t = -\Delta t$:

$$k_{ij} = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(S_0 = i | S_{-\Delta t} = j) \quad , \quad i \neq j\,, k_{ii} = 1 - \sum_{j \neq i} k_{ji}\,.$$

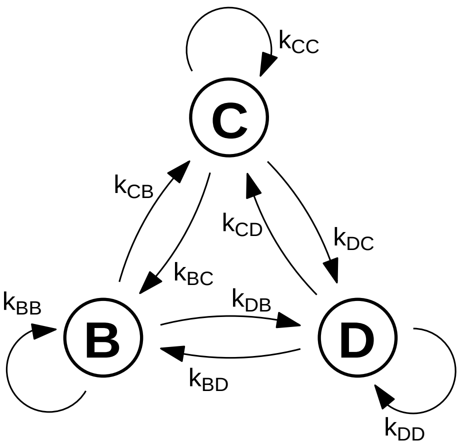


Figure 7: Transition diagram of the three state Markov process used to describe the transition dynamics of the protein model (see text). Three states, B,C, and D, corresponding to the three maxima of the configuration space density projection in Fig. 5, are connected via conformational transitions (arrows). The stochastic dynamics of that model is specified by the values of nine transition rates $k_{ij}$.

In order to decide, whether the observed conformational dynamics is Markovian, we need to check (cf. Ref. 66), whether the probability for a particular transition depends on the history of the process, i.e., whether the conditional probabilities listed below are independent of the time intervals $\Delta_k$, $k = 0, 1, 2, \ldots$:

$$\begin{aligned} P_{ij}(\Delta_1) &:= P(S_0 = i | S_{-\Delta_1} = j) \\ P_{ij_1j_2}(\Delta_1, \Delta_2) &:= P(S_0 = i | S_{-\Delta_1} = j_1 \wedge S_{-\Delta_1-\Delta_2} = j_2) \\ P_{ij_1j_2j_3}(\Delta_1, \Delta_2, \Delta_3) &:= P(S_0 = i | S_{-\Delta_1} = j_1 \wedge S_{-\Delta_1-\Delta_2} = j_2 \wedge S_{-\Delta_1-\Delta_2-\Delta_3} = j_3) \\ &\vdots \qquad \vdots \end{aligned} \tag{5}$$

Of course, the limited set of available data (a total of 238 conformational transitions) does not allow to verify all the above conditions. Instead, we restrict our analysis to a comparison of only the first and second order conditional probabilities, $P_{ij}(\Delta_1)$ and $P_{ij_1j_2}(\Delta_1, \Delta_2)$, respectively. In addition, we consider only probabilities $P_{ij_1j_2}(\Delta) := P_{ij_1j_2}(\Delta,\Delta)$ derived from equally spaced instances of time. In applying this slightly less rigorous test, we assume that 'partial amnesia' is not likely to appear within the dynamical systems considered here, that is, our analysis does not apply to processes, which exhibit long-term memory, but at the same time do *not* show short-term memory. This assumption is supported by the results of the preceding section, where it was shown, that the dynamics of our model does exhibit short-term memory.

Estimates for the first and second order conditional probabilities $P_{ij}(\Delta)$ and $P_{ij_1j_2}(\Delta)$ were computed from frequency counts derived from the 232 nanoseconds of simulation data by

$$
\begin{aligned}
P_{ij}(\Delta) &\approx \frac{N_{ij}(\Delta)}{N_j(\Delta)} \\
P_{ij_1j_2}(\Delta) &\approx \frac{N_{ij_1j_2}(\Delta)}{N_{j_1j_2}(\Delta)} .
\end{aligned}
$$

Here, the frequency counts $N_j$, $N_{ij}$, and $N_{ij_1j_2}$, where $i,j,j_1,j_2 \in \{B,C,D\}$, are defined in analogy to (5): $N_j(\Delta)$ denotes, how often in the course of the simulations the model has been found in substate $j$; $N_{ij}(\Delta)$ denotes the number of times it has been found in substate $i$, *if* it has been in substate $j$ a particular time span $\Delta$ before, and $N_{ij_1j_2}$, how often it has been found in substate $i$, *if* it has been in substate $j_1$ before *and* in substate $j_2$, $2\Delta$ before:

$$
\begin{aligned}
N_j &= \sum_k \delta_{j,S(k\Delta')} , \\
N_{ij} &= \sum_k \delta_{i,S(k\Delta'+\Delta)}\delta_{j,S(k\Delta')} , \\
N_{ij_1j_2} &= \sum_k \delta_{i,S(k\Delta'+2\Delta)}\delta_{j_1,S(k\Delta'+\Delta)}\delta_{j_2,S(k\Delta')} .
\end{aligned}
$$

The frequency counts were determined using a sample rate of $\Delta' = 0.5\,ps$. Here, $\delta_{i,S(t)}$ denotes the Kronecker symbol, being unity, if the protein model is in conformational state $i$ at time $t$, and zero otherwise.

In order to test the hypothesis that corresponding first order and second order conditional probabilities are equal, in which case the observed conformational dynamics would be Markovian, the statistical significance of observed deviations of the probability estimates $P_{ij_1j_2}$ from $P_{ij}$ had to be determined. Accordingly, we estimated the corresponding standard deviations $\sigma_{ij_1j_2}$ from our data by

$$
\sigma_{ij_1j_2}(\Delta) \approx \sqrt{P_{ij_1j_2}(\Delta)(1 - P_{ij_1j_2}(\Delta))/N_{j_1j2}(\Delta)} , \qquad (6)
$$

using the Gaussian approximation for binomial probability distributions.

From the set of 27 possible probability distributions $P_{ij_1j_2}(\Delta)$, $i,j_1,j_2 \in \{B,C,D\}$ which we examined, Figure 8 shows a selection of six typical cases.

These are the second order conditional probability estimates $P_{DDD}$, $P_{DDB}$, $P_{DDC}$, $P_{BCD}$, $P_{BDD}$, and $P_{CCD}$ computed in the range $\Delta = 50 \ldots 500\,\mathrm{ps}$ (bold, solid lines) which are compared with the corresponding first order conditional probability estimates, $P_{ij}(\Delta)$ (bold, dashed). Also shown is a $1\sigma$-interval (thin, solid) as well as a $2\sigma$-interval (thin, dashed dotted), computed according to (6).

The main observation is that the deviations of the second order conditional probabilities from the first order probabilities do not exceed $2\sigma$, i.e., with a significance of 2% the data are consistent with the hypothesis that the conformational dynamics of our model is indeed Markovian. This is also true for the remaining 21 probability distributions not shown in Fig. 8. For a closer analysis we note, that although the three second order probabilities, $P_{DDD}$, $P_{DDB}$, and $P_{DDC}$ (upper half of Fig. 8) correspond to the same first order probability, $P_{DD}$, the particular error ranges vary considerably as a result of the different population densities of the three states. Whereas the broad error ranges of $P_{DDC}$ or $P_{BCD}$, resulting from a low population density of state $C$ do not provide much significance, the narrow ranges apparent for $P_{DDD}$ or $P_{BDD}$, resulting from a large number of transitions between these highly populated states, provide a good check for non-Markovian memory-effects. The largest deviation ($\approx 2\sigma$) was observed for $P_{DDC}$ in the time range below 60 ps. One may speculate, that this deviation might be due to a minor memory-effect present in the rapid decay dynamics from conformational state $C$ into $D$.

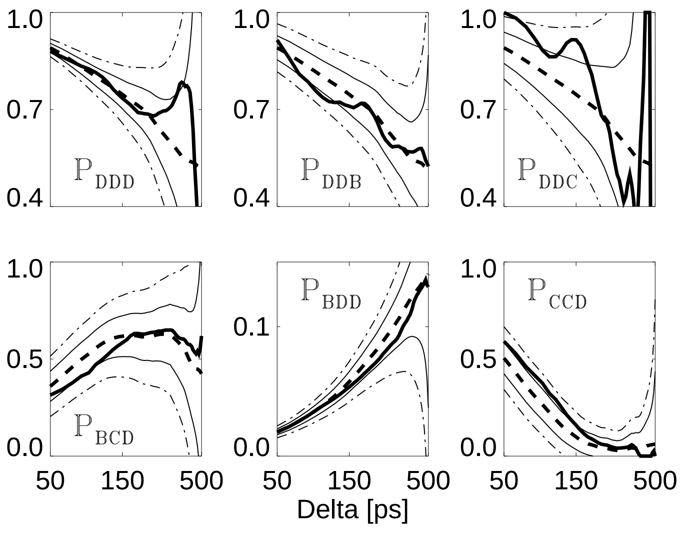


Figure 8: Comparison of six selected second order conditional probabilities $P_{ij_1j_2}$ (bold, solid) with the corresponding first order conditional probabilities $P_{ij}$ (bold, dashed), as described in the text. The statistical error due to the limited data set is depicted as a $1\sigma$-range (thin, solid lines) centered at $P_{ij}$, and a $2\sigma$-range (thin, dashed-dotted), respectively.

The above analysis of the distribution of conformational transition times shows, that the conformational dynamics of our model is actually well described by a Markov model suggesting that no memory effects are present at the corresponding time scale of a few hundred picoseconds.

# 5 Summary and Discussion

We studied the conformational dynamics of a simplified protein model by molecular dynamics simulations at long time scales up to $0.232\,\mu$s. The model has been designed as a minimal model, intended to include only those structural elements, which can be assumed to be essential for the low-frequency dynamical properties of proteins. As shown by our simulations the model actually exhibits properties similar to those of more realistic protein models, such as tertiary structure, vibrational spectra, a hierarchy of time scales, and the occurrence of rare transitions between conformational substates.

Two effective descriptions for the conformational dynamics of the model, which both neglect memory effects, were compared with an explicit MD-simulation. To enable an analysis, a rigorous theoretical concept of conformational substates based on the notion of free energy landscapes has been formulated and applied.

The first model, a Langevin model, describing the dynamics *within* a certain conformational state at a picosecond time scale, could not reproduce conformational transition rates derived from MD-simulations. This failure was found to be due to memory effects, caused by correlations between many degrees of freedom, which strongly influence the short time scale dynamics within conformational states at the picosecond time scale.

The second description, a Markov model, aimed at an analysis of the protein dynamics at the much longer time scale of few hundred picoseconds. Here, the analysis of the distribution of conformational transition times suggested, that the conformational dynamics of our model does not exhibit memory effects.

These findings demonstrate a qualitative change in the dynamical behavior of our model protein, when proceeding from the short time scales, which are at present accessible by MD-simulations of realistic protein models, to longer time scales: whereas memory effects play a significant role at short time scales, they appear to vanish at longer time scales. As a result the slower conformational dynamics can be described by a master equation.

Care has to be taken in the attempt to generalize these results to the dynamical behavior of real proteins. As argued in the first part of Section 3, we expect the polymer chain of real proteins to be much more flexible than that of our simple model. Correspondingly, their conformational dynamics — as far as collective motions of the polypeptide chain are involved — should occur at slower time scales. For this reason, one can not conclude, that memory effects in protein dynamics are actually absent at that hundred picosecond time scale characteristic for conformational transitions of our model. However, the results do suggest that memory effects in the dynamics of proteins generally tend to vanish at long time scales. Furthermore, because enhanced rigidity of real proteins, the time scale of few hundred picosecond can be considered as a *lower bound* for the absence of memory effects.

## 6 Acknowledgements

This work has been supported by the Deutsche Forschungsgemeinschaft (SFB 143/C1). The authors are grateful to the referee for his careful reading of the manuscript and valuable hints for its improvement.

# References


[1] J. Deisenhofer and H. Michel, The Crystal Structure of the Photosynthetic Reaction Center from Rps. Viridis, in *The Photosynthetic Bacterial Reaction Center: Structure and Dynamics*, edited by J. Breton and A. Vermeglio, pp. 1–3, London, 1988, Plenum Press.

[2] R. Henderson, J. M. Baldwin, T. A. Ceska, F. Zemlin, E. Beckmann, and K. H. Downing, *J. Molec. Biol.* **213**, 899 (1990).

[3] M. Levitt, *J. Molec. Biol.* **168**, 621 (1983).

[4] H. Frauenfelder, *Biophys. J.* **47**, 35 (1985).

[5] J. A. McCammon and M. Karplus, *Proc. Natl. Acad. Sci. USA* **76**, 3585 (1987).

[6] M. Levitt and S. Lifson, *J. Molec. Biol.* **46**, 269 (1969).

[7] J. A. McCammon, B. R. Gelin, and M. Karplus, *Nature* **267**, 585 (1977).

[8] H. S. Chan and K. A. Dill, *Ann. Rev. Biophys. Biophys. Chem.* **20**, 447 (1991).

[9] J. E. Mertz, D. J. Tobias, C. L. Brooks, and U. C. Singh, *J. Comp. Chem.* **12**, 1270 (1991).

[10] N. S. Ostlund and R. A. Whiteside, A machine architecture for molecular dynamics: the systolic loop, in *Macromolecular Structure and Specificity: Computer-Assisted Modeling and Applications*, edited by B. Venkataraghavan and R. J. Feldman, pp. 195–208, Annals of the N. Y. Acad. of Sciences 439, New York, 1985.

[11] H. Heller, H. Grubmüller, and K. Schulten, *Molec. Sim.* **5**, 133 (1990).

[12] B. R. Brooks and M. Hodošček, *Chem. Des. Autom. News* **7**, 16 (1992).

[13] W. F. van Gunsteren and H. J. C. Berendsen, *Angew. Chem. Int. Ed. Engl.* **29**, 992 (1990).

[14] D. I. Okunbor, *Canonical Integration Methods for Hamiltonian Dynamical Systems*, PhD thesis, University of Illinois at Urbana-Champaign, Department of Computer Science, 1304 W. Springfield Avenue, Urbana, Il 61801-2987, 1992.

[15] M. Buchner and B. M. Ladanyi, *Molec. Phys.* **73**, 1127 (1991).

[16] J. Turner, H. Chun, V. Lupi, P. Weiner, S. Gallion, and C. Singh, *Chem. Des. Autom. News* **7**, 33 (1992).

[17] R. D. Ruth, *IEEE Trans. Nuc. Sci.* **30**, 2669 (1983).

[18] J. Barnes and P. Hut, *Nature* **324**, 446 (1986).

[19] L. Greengard and V. Rokhlin, *Chem. Scr.* **29A**, 139 (1989).

[20] S. J. Aarseth, *Direct Methods for N-Body Simulations*, pp. 378–418, Multiple Time Scales, Academic Press, 1st edition, 1985.

[21] R. Chicon and E. Martin, *American Journal of Physics* **59**, 759 (1991).

[22] B. J. Berne, *Molecular Dynamics and Monte Carlo Simualtions of Rare Events*, pp. 419–436, Multiple Time Scales, Academic Press, 1st edition, 1985.

[23] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comp. Chem.* **4**, 187 (1983).

[24] H. Grubmüller, H. Heller, A. Windemuth, and K. Schulten, *Molec. Sim.* **6**, 121 (1991).

[25] M. E. Tuckerman and B. J. Berne, *J. Chem. Phys.* **95**, 8362 (1991).

[26] H. Grubmüller, *Molekulardynamik von Proteinen auf langen Zeitskalen*, PhD thesis, Technische Universität München, München, 1994.

[27] R. J. Loncharich and B. R. Brooks, *Proteins* **6**, 32 (1989).

[28] W. F. van Gunsteren and H. J. C. Berendsen, *Molec. Phys.* **34**, 1311 (1977).

[29] Bennet, *J. Comp. Phys.* **19**, 297 (1975).

[30] Metropolis, *J. Chem. Phys.* **21**, 1087 (1953).

[31] Pangali, *Chem. Phys. Lett.* **55**, 413 (1978).

[32] R. M. Levy, M. Karplus, and J. A. McCammon, *Chem. Phys. Lett.* **65**, 4 (1979).

[33] M. E. Tuckerman and B. J. Berne, *J. Chem. Phys.* **95**, 4389 (1991).

[34] S. Miyazawa, *Macromolecules* **18**, 534 (1985).

[35] H. Frauenfelder, K. Chu, and R. Philipp, Physics from Proteins, in *Biologically Inspired Physics*, edited by L. Peliti, Plenum Press, New York, 1991.

[36] L. Verlet, *Phys. Rev.* **159**, 98 (1967).

[37] K. D. Gibson and H. A. Scheraga, *J. Comp. Chem.* **11**, 468 (1990).

[38] H. Grubmüller and P. Tavan, *J. Comp. Chem.* (1994), in preparation.

[39] A. Ansari et al., *Biophysical Chemistry* **26**, 337 (1987).

[40] G. U. Nienhaus, J. R. Mourant, and H. Frauenfelder, *PNAS* **89**, 2902 (1992).

[41] A. Brünger, C. Brooks, and M. Karplus, *Proc. Natl. Acad. Sci. USA* **82**, 8458 (1985).

[42] H. Treutlein, K. Schulten, A. Brünger, M. Karplus, J. Deisenhofer, and H. Michel, *Proc. Natl. Acad. Sci. USA* (1989).

[43] S. Swaminathan, T. Ichiye, W. van Gunsteren, and M. Karplus, *Biochemistry* **21**, 5230 (1982).

[44] R. H. Austin et al., *Biochemistry* **14**, 5355 (1975).

[45] A. Ansari, J. Berendzen, S. F. Browne, H. Frauenfelder, I. E. T. Iben, T. B. Sauke, E. Shyamsunder, and R. D. Young, *Proc. Natl. Acad. Sci. USA* **82**, 5000 (1985).

[46] P. Hänggi, P. Talkner, and M. Borkovec, *Rev. Mod. Phys.* **62**, 251 (1990).

[47] B. R. Brooks, R. W. Pastor, and F. W. Carson, *Proc. Natl. Acad. Sci. USA* **84**, 4470 (1987).

[48] R. Czerminski and R. Elber, *Proc. Natl. Acad. Sci. USA* **86**, 6963 (1989).

[49] R. Elber, *J. Chem. Phys.* **93**, 4312 (1990).

[50] J. Pleiss and F. Jähnig, *Eur. Biophys. J.* **21**, 63 (1992).

[51] J. P. Valleau and G. M. Torrie, in *Statistical Mechanics Part A: Equilibrium Techniques*, edited by B. J. Berne, p. 137, Plenum Press, New York, 1977.

[52] J. Cao and B. J. Berne, *J. Chem. Phys.* **92**, 1980 (1990).

[53] H. S. Chan and K. A. Dill, *J. Chem. Phys.* **95**, 3775 (1991).

[54] J. D. Honeycutt and D. Thirumalai, *Biopolymers* **32**, 695 (1992).

[55] H. Frauenfelder, S. G. Sligar, and P. G. Wolynes, *Science* **254**, 1598 (1991).

[56] W. Doster, S. Cusack, and W. Petry, *Nature* **337**, 754 (1989).

[57] W. Nadler, A. T. Brünger, K. Schulten, and M. Karplus, *Proc. Natl. Acad. Sci. USA* **84**, 7933 (1987).

[58] D. E. Smith and C. B. Harris, *J. Chem. Phys.* **92**, 1304 (1990).

[59] T. E. Creighton, *Proteins*, W. H. Freeman and Company, San Francisco, 1984.

[60] K. A. Dill, *Biochemistry* **29**, 7133 (1990).

[61] M. Prévost, S. J. Wodak, B. Tidor, and M. Karplus, *Proc. Natl. Acad. Sci. USA* **88**, 10880 (1991).

[62] R. L. Ornstein, *J. Biol. Struct. Dyn.* **7**, 1019 (1990).

[63] J. C. Smith, *Q. Rev. Biophys.* **24**, 227 (1991).

[64] M. Kurzynski, *Accounts of Chemical Research* , submitted.

[65] H. Frauenfelder, P. J. Steinbach, and R. D. Young, *Chem. Scr.* **29A**, 145 (1989).

[66] S. A. Rice and P. Gray, *The Statistical Mechanics of Simple Liquids*, Interscience Publishers, John Wiley & Sons, Inc., New York, 1965.

[67] R. Elber and M. Karplus, *Science* **235**, 318 (1987).

[68] N. Go and T. Noguti, *Chem. Scr.* **29A**, 151 (1989).

[69] S. Chandrasekhar, *Rev. Mod. Phys.* **15**, 1 (1943).

[70] H. A. Kramers, *Physica* **7**, 284 (1940).