Conformational Dynamics Simulations of Proteins

Markus Eichinger, Berthold Heymann, Helmut Heller, Helmut Grubmüller, and Paul Tavan

Institut für Medizinische Optik, Theoretische Biophysik,
Ludwig-Maximilians-Universität München,
Oettingenstraße 67, D-80538 München, Germany

Abstract. Molecular dynamics (MD) simulations of proteins provide descriptions of atomic motions, which allow to relate observable properties of proteins to microscopic processes. Unfortunately, such MD simulations require an enormous amount of computer time and, therefore, are limited to time scales of nanoseconds. We describe first a fast multiple time step structure adapted multipole method (FAMUSAMM) to speed up the evaluation of the computationally most demanding Coulomb interactions in solvated protein models, secondly an application of this method aiming at a microscopic understanding of single molecule atomic force microscopy experiments, and, thirdly, a new method to predict slow conformational motions at microsecond time scales.

1 Introduction

In many cases the detailed knowledge of dynamic processes at the atomic level is essential to understand protein function, e.g., ligand binding or enzymatic reactions. Through a microscopic description of interatomic forces [1] and atomic motions, molecular dynamics (MD) simulations [2,3] can serve as a tool to interpret experimental data and to make predictions, which can guide future experiments. In such simulations, the motions are computed by numerically solving Newton's equations. Here, the forces are derived from an empirical energy function accounting for chemical binding forces as well as van der Waals and electrostatic interactions between partially charged atoms.

For the study of protein dynamics quite large simulation systems — typically comprising several $10,000$ atoms — are required. The system must be that large because the native protein environment (water or lipids) strongly affects the dynamics of the protein [4–8] and, therefore, has to be included into the simulation system. The large number of atoms provides a first reason why MD simulations of proteins pose a computational challenge. A second reason is that femtosecond integration time steps are necessary to enable sufficiently smooth descriptions of the fastest degrees of freedom. Thus, MD simulations of such systems are currently limited to nanoseconds (i.e., a few million integration steps) even if the most powerful supercomputers and efficient algorithms are used. Although there are a number of biochemically important processes which occur at such very fast time scales and have been successfully studied by MD simulations [9,10], most biochemical processes occur at much slower scales and, therefore, are currently inaccessible to conventional MD methods. This technical limitation motivates substantial efforts taken by many groups to determine suitable approximations which ideally should allow more efficient simulations without seriously affecting relevant features of the system, which may be grouped into specialized integration schemes and multiple time stepping [11–31] (see also the chapter by Schlick and Berne within this book), multipole methods [32–38], as well as grid and Ewald methods [39–42]. Most of the efforts focus on the efficient computation of the electrostatic interactions within the protein and between protein and solvent [see also the chapter by W.v. Gunsteren et al. within this book], since, typically, this is the computationally most demanding task.

As an example for an efficient yet quite accurate approximation, in the first part of our contribution we describe a combination of a structure adapted multipole method with a multiple time step scheme (FAMUSAMM — fast multistep structure adapted multipole method) and evaluate its performance. In the second part we present, as a recent application of this method, an MD study of a ligand-receptor unbinding process enforced by single molecule atomic force microscopy. Through comparison of computed unbinding forces with experimental data we evaluate the quality of the simulations. The third part sketches, as a perspective, one way to drastically extend accessible time scales if one restricts oneself to the study of conformational transitions, which are ubiquitous in proteins and are the elementary steps of many functional conformational motions.

2 Efficient MD-Simulation Methods

In order to solve the classical equations of motion numerically, and, thus, to obtain the motion of all atoms the forces acting on every atom have to be computed at each integration step. The forces are derived from an energy function which defines the molecular model [1–3]. Besides other important contributions (which we shall not discuss here) this function contains the Coulomb sum

$$U \propto \sum_{i}^{N} \sum_{j<i} \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|}$$

over all pairs of atoms (i, j) with partial charges q_i at positions $\mathbf{r}_i$. The evaluation of this sum dominates the computational effort in MD simulations as it scales quadratically with the number N of charged particles.

A very simple — and in fact quite widely used — approximation completely neglects long range electrostatic interactions beyond a certain cut-off distance [43] of typically $8 - 15$ Å. For systems which are significantly larger than this cut-off distance the computation of the remaining Coulomb interactions then scales with N instead of N^2. However, such truncation leads to serious artifacts concerning the description of the structure and dynamics of proteins [44,24,45], and more accurate methods which include the long range interactions should be preferred. Multipole methods and multiple-time-step methods are well established and widely used for this purpose. We briefly sketch both methods and subsequently show how their combination allows highly efficient simulations.

2.1 Efficient Multipole Methods

Multipole methods approximate the long-range forces originating from a group of point charges by truncated multipole expansions of their electrostatic potential. Using a hierarchy of grids for subdivision of space, nested at multiple scales, and a corresponding hierarchical organization of charge groups and multipole expansions [33] a computational complexity of $O(NlogN)$ can be achieved. By additionally using a hierarchy of local Taylor expansions for the evaluation of the electrostatic potential in the vicinity of a group of particles Greengard and Rokhlin have constructed the so-called *fast multipole method* (FMM) that even scales with $O(N)$ for large systems [34,35].

For MD simulations of biomolecules the FMM-type grouping of charges, defined by a fixed and regular subdivision of space, requires multipole expansions of rather high order (more than 6 terms of the expansion) as to achieve sufficient numerical accuracy [34]. If, instead, as shown in Figure 1, the charge grouping is adapted to specific structural and dynamical properties of the simulated biomolecules, the multipole expansions can be truncated at quite low orders, e.g., after the second order, while maintaining sufficient accuracy [36–38].

In the FAMUSAMM framework, e.g., we have grouped locally stable groups of typically three to ten covalently bound atoms into so-called structural units (level 1 in Fig. 1). By construction, these structural units either carry integer elementary charges or are uncharged, but dipolar. Test simulations show that for distances $> 10\,\text{Å}$ already the lowest non-vanishing multipole moments of these structural units provide a sufficiently accurate description of the electrostatic forces within biomolecules with an error below 2%. The objects of the next hierarchy level (level 2 in the figure) are formed by grouping structural units into clusters. For interaction distances $> 15\,\text{Å}$ the electrostatic potential of those objects, again, can be approximated by their lowest multipole moment. Extending this scheme to higher hierarchy levels, such a structure adapted multipole method (SAMM) provides a substantial speed-up for MD simulations as compared to the conventional, grid-based methods [38,46].

The performance of this first version of SAMM [36] can be further enhanced by additionally utilizing FMM-strategies [34,38]. Here, in the vicinity of a given object (e.g., a structural unit or a cluster) the electrostatic potential originating from distant charge distributions is approximated by a local

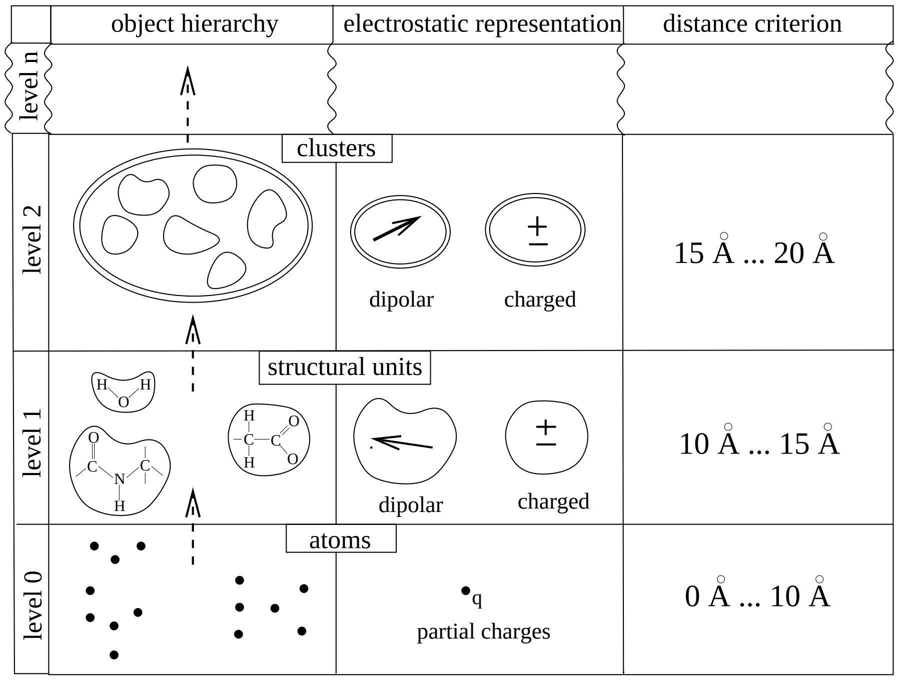

Fig. 1. Structure adapted hierarchical description of Coulomb interactions in biological macromolecules. Filled circles (level 0) represent atoms, structural units (level 1) are surrounded by a single-line border, and clusters (level 2) are surrounded by a double-line border.

Taylor expansion. Specifically, the basic tasks involved in the FMM aspect of SAMM are:

Task 1: Calculate the first non-vanishing multipole moment of the electrostatic potential of composed objects (i.e., structural units and clusters).

Task 2: Add up electrostatic potential contributions to local Taylor expansions of all objects on each hierarchy level. (Contributions to the local Taylor expansion of a selected object arise from all other objects on the same hierarchy which fulfill the distance criterion given in Fig. 1.)

Task 3: Transform ("inherit") local Taylor expansions from a upper hierarchy level to the next lower hierarchy level.

Task 4: Explicitly calculate the Coulomb interactions between atoms which are closer than about 10 Å.

In the next section we will illustrate how to further speed up the SAMM method by introducing multiple-time-stepping.

2.2 SAMM with Multiple-time-stepping

In general, multiple-time-step methods increase computational efficiency in a way complementary to multipole methods: The latter make use of regularities *in space*, whereas multiple-time-stepping exploits regularities *in time*. Figure 2 illustrates the general idea:

As sketched in the right part of the figure, forces between distant atoms generally exhibit slower fluctuations than forces between close atoms. Therefore, without significant loss of accuracy, the more slowly fluctuating forces may be computed using larger integration step sizes. As shown in the left part of the figure, the required classification of forces can be implemented, e.g., by grouping atom pairs into distance classes. The slowly fluctuating forces arising from outer distance classes may then be evaluated less frequently (filled squares) than the fast ones and, instead, are extrapolated (open squares) from previously computed forces at the time steps in between.

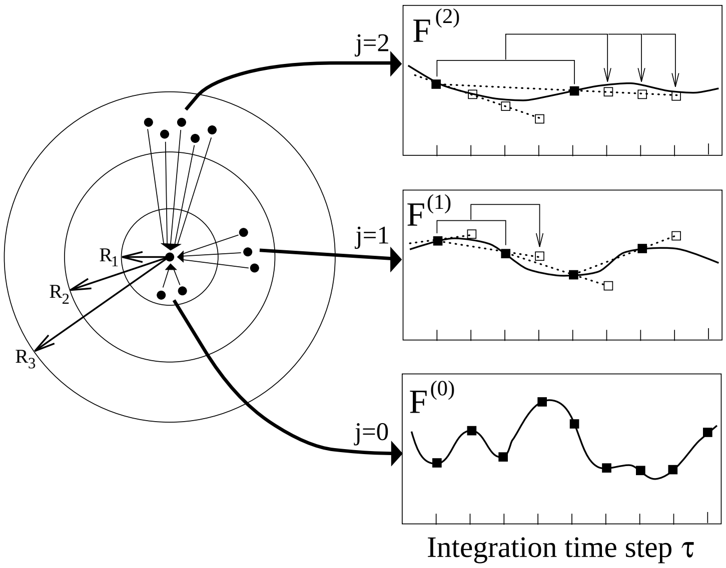

Fig. 2. Distance classes $j = 0, 1, 2, \ldots$ (left) are defined for an atom (central dot) by a set of radii R_{j+1}; the right curves sketch the temporal evolution of the total force $\mathbf{F}^{(j)}$ acting on the selected atom originating from all atoms in distance class j; shown are the exact forces (solid line), their exact values to be computed within the multiple time step scheme (filled squares), linear force extrapolations (dotted lines), and resulting force estimates (open squares).

This hierarchical extrapolation procedure can save a significant amount of computer time as it avoids a large fraction of the most time consuming step, namely the exact evaluation of long range interactions. Here, computational speed is gained at the cost of an increased demand for memory, e.g., for each atom and each distance class two previously computed forces have to be kept in memory.

In the framework of the fast structure adapted multipole method the memory demand could be drastically reduced. This was achieved by applying the multiple-time-step scheme (we used the so-called DC-1d scheme [24]) to the interactions between charge groups (structural units and clusters) rather than to the forces acting on individual atoms. In the following we give a short description of this tight and efficient combination. We termed the resulting algorithm FAMUSAMM (*multiple-time-step structure-adapted multipole method*) [46,47].

A detailed analysis of fast SAMM has shown [47] that the most time consuming tasks are task 2 and task 4 described above. In task 2 for each hierarchy level (except for level 0) a local Taylor expansion is calculated for each object. Note that here we refer to expansions which comprise only contributions from objects of the same hierarchy level which, in addition, fulfill the distance criterion given in Fig. 1. From each of these local expansions, *approximated* electrostatic forces $\tilde{\mathbf{F}}^{(j)}$ acting on the atoms contained in the associated object could be computed and, in analogy to the exact forces $\mathbf{F}^{(j)}$ used in the multiple time step scheme described above (see Fig. 2), the multipole-derived forces $\tilde{\mathbf{F}}^{(j)}$ could be extrapolated by multiple time stepping. We further improved that obvious scheme, however, in that we applied multiple time step extrapolations to the *coefficients* of the local Taylor expansions instead. That strategy reduces memory requirements by a significant factor without loss of accuracy, since the number of local Taylor coefficients that have to be kept for the extrapolation is smaller than the number of forces acting on all atoms of the respective object.

Additionally, to optimize task 4, we applied a conventional, atom pair interaction based multiple-time-step scheme to the force computation within the innermost distance class. Here, for atom pairs closer than 5 Å, the Coulomb sum is calculated every step, and for all other atom pairs the Coulomb sum is extrapolated every second step from previously explicitly calculated forces.

This completes the outline of FAMUSAMM. The algorithm has been implemented in the MD simulation program EGO_VIII [48] in a sequential and a parallelized version; the latter has been implemented and tested on a number of distributed memory parallel computers, e.g., IBM SP2, Cray T3E, Parsytec CC and ethernet-linked workstation clusters running PVM or MPI.

2.3 Computational Performance

Here we want to document that FAMUSAMM actually provides an enhanced computational efficiency both as compared to SAMM as well as to the reference method which is characterized by exact evaluation of the Coulomb sum. To that aim we have carried out a series of test simulations for systems of varying size ranging from 500 to $40,000$ atoms. We used the sequential version of EGO_VIII. All simulations were executed on a DEC-ALPHA 3300L (175 MHz) workstation equipped with 96 MB RAM. Figure 3 shows that the average computation time required for one MD integration step scales linearly with system size for systems comprising more than about $1,000$ atoms.

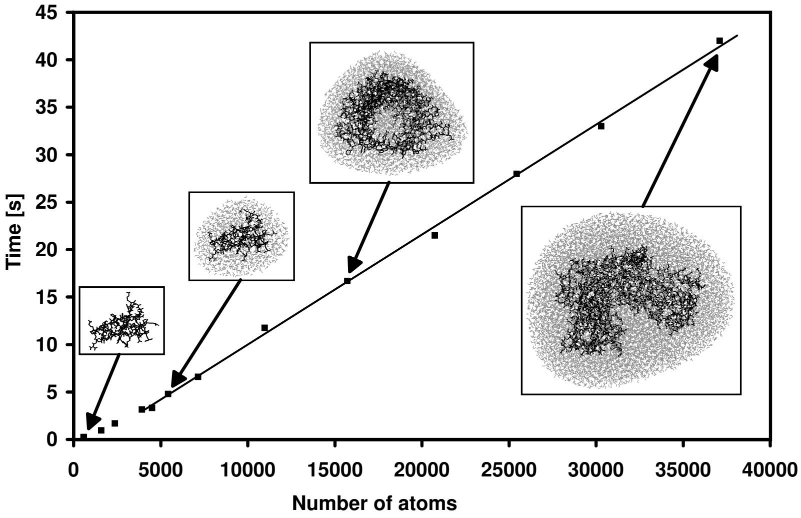

Fig. 3. Average computation time for one step using EGO_VIII on a DEC-Alpha 3300L workstation (175 MHz) for simulation systems of varying size. The insets show some of the protein-water systems used for the benchmark simulations.

For large systems comprising $36,000$ atoms FAMUSAMM performs *four* times faster than SAMM and as fast as a cut-off scheme with a 10 Å cut-off distance while completely avoiding truncation artifacts. Here, the speed-up with respect to SAMM is essentially achieved by the multiple-time-step extrapolation of local Taylor expansions in the outer distance classes. For this system FAMUSAMM executes by a factor of 60 faster than explicit evaluation of the Coulomb sum. The subsequent Section describes, as a sample application of FAMUSAMM, the study of a ligand-receptor unbinding process.

3 Microscopic Interpretation of Atomic Force Microscope Rupture Experiments

That simulation study [49] aimed at a microscopic interpretation of single molecule atomic force microscope (AFM) experiments [50], in which unbinding forces between individual protein–ligand complexes have been measured

(Fig. 4, top). In particular we asked, what interatomic interactions cause the experimentally observed unbinding forces.

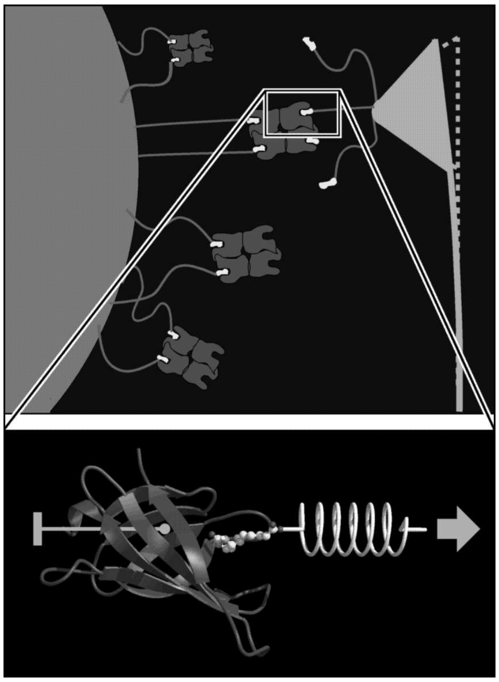

Fig. 4. Typical AFM rupture experiment (top): Receptor molecules are fixed via linker molecules to a surface (left); in the same way, ligand molecules are connected to the AFM cantilever (right). When pulling the cantilever towards the right, the pulling force applied to the ligand can be measured. At the point of rupture of the ligand–receptor complex the measured force abruptly drops to zero so that the rupture force can be measured.
Computer rupture simulation (bottom): In the course of an MD simulation of the ligand–receptor complex at atomic detail the ligand is pulled towards the right with a 'computer spring', while the receptor (drawn as a ribbon model) is kept in place. From the elongation of the 'spring' the pulling force during the unbinding process is computed, and, thereby, a 'force profile' is obtained. The rupture force is interpreted as the maximum of this force.

Both the AFM rupture experiments as well as our simulation studies focussed on the streptavidin-biotin complex as a model system for specific ligand binding. Streptavidin is a particularly well-studied protein and binds its ligand biotin with high affinity and specificity [51]. Whereas previous experiments (see references in Ref. [49]) and simulation studies [52] referred only to bound/unbound states and the associated kinetics, the recent AFM rupture experiments have provided a new and complementary perspective on ligand binding by focussing at atomic details of binding/unbinding *pathways*: The former were described in terms of binding free energies as thermodynamic quantities, which are independent of the particular reaction pathway; the latter relate to *forces*, which actually depend on details of the unbinding reaction path and, therefore, can provide new insights into these details.

To enable an atomic interpretation of the AFM experiments, we have developed a molecular dynamics technique to simulate these experiments [49]. From such 'force simulations' rupture models at atomic resolution were derived and checked by comparisons of the computed rupture forces with the experimental ones. In order to facilitate such checks, the simulations have been set up to resemble the AFM experiment in as many details as possible (Fig. 4, bottom): the protein–ligand complex was simulated in atomic detail starting from the crystal structure, water solvent was included within the simulation system to account for solvation effects, the protein was held in place by keeping its center of mass fixed (so that internal motions were not hindered), the cantilever was simulated by use of a harmonic 'spring potential' and, finally, the simulated cantilever was connected to the particular atom of the ligand, to which in the AFM experiment the linker molecule was connected.

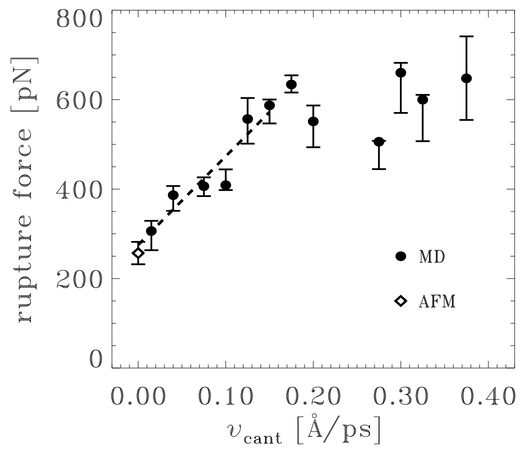

Fig. 5. Theory vs. experiment: rupture forces computed from rupture simulations at various time scales (various pulling velocities v_{cant}) ranging from one nanosecond ($v_{\mathrm{cant}} = 0.015$ Å/ps) to 40 picoseconds ($v_{\mathrm{cant}} = 0.375$ Å/ps) (black circles) compare well with the experimental value (open diamond) when extrapolated linearly (dashed line) to the experimental time scale of milliseconds.

However, one significant difference between the AFM experiment and its simulations cannot be avoided at present: Whereas the AFM experiment takes place at a millisecond time scale, our simulations had to be completed within the nanosecond time scale. So, in fact, in the simulation the pulling velocity had to be chosen about six orders of magnitude larger than in the AFM experiment!

In a first attempt to bridge these six orders of magnitude, we performed a series of rupture force simulations using pulling velocities ranging from 0.375 to 0.015 Å/ps. As can be seen in Fig. 5, we observed a linear dependency of the computed rupture forces in the velocity range between 0.15 and 0.015 Å/ps. This suggests that simple friction dominates the non-equilibrium effects in this regime described by a friction coefficient of 20 pN s/m. A simple linear extrapolation of the computed rupture forces to the experimental time scale shows agreement between theory and experiment. Clearly, this first step has not yet solved the question how to bridge the six orders of magnitude gap between theory and experiment (cf. also [53]). To answer that question, a better understanding of the physics of rupture experiments using simplified models on the one hand (cf., e.g., Ref. [54]) and, on the other hand, a careful analysis of the atomic processes which cause the velocity dependent rupture forces is necessary.

One of the results of an MD rupture simulation is the pulling force as a function of time or cantilever position $z_{\mathrm{cant}}(t)$, called the *force profile*. Figure 6 shows an example, derived from an extended 1 ns-simulation, where a pulling velocity of 0.015 Å/ps was used. The apparent multitude of force maxima mirrors the complexity of the energy landscape traversed by the biotin on its way out of the binding pocket. The peaks of this force profile can be attributed to the rupture and formation of individual hydrogen bonds and water bridges shown in the snapshots of Fig. 7, which characterize the main steps of the rupture process. The rupture forces in Fig. 5 are the maxima of the corresponding force profiles.

We will not discuss here in detail our atomic model of the unbinding process derived from our simulations and sketched in Fig. 7, but restrict ourselves to two unexpected features. One is that the rupture of the initially very strong hydrogen bonds between the ligand and the residues of the binding pocket (Fig. 7 A) does not entail immediate unbinding. Rather, the complex is stabilized by a transient network of water bridges and other transient hydrogen bonds, which form during the unbinding process (Fig. 7 B and C). Only after subsequent rupture of these hydrogen bonds the maximum force — the rupture force — is reached and the biotin rapidly moves out of the entry of the binding pocket (Fig. 7 D). As another feature we observed, towards the end of the unbinding process, a second force maximum, which we attribute to a strong transient hydrogen bond and several water bridges between biotin and the entry of the binding pocket (Fig. 7 E). Crossing of that second barrier, which cannot yet be resolved in the AFM experiment, completes the unbinding process.

In summary, our simulations provided detailed insight into the complex mechanisms of streptavidin–biotin rupture. They attribute the binding force

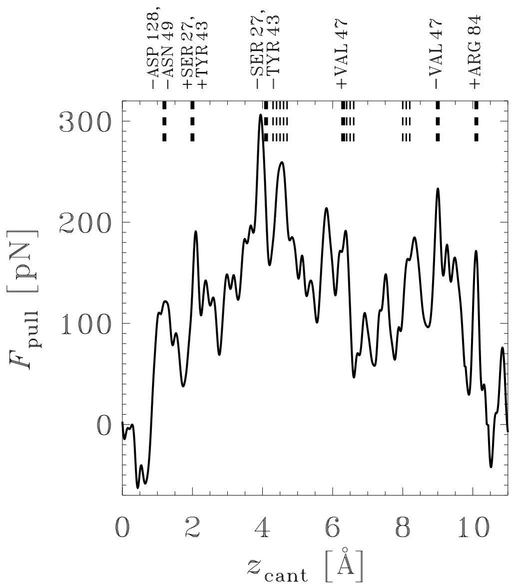

Fig. 6. Force profile obtained from a one nanosecond simulation of streptavidin–biotin rupture showing a series of subsequent force peaks; most of these can be related to the rupture of individual microscopic interactions such as hydrogen bonds (bold dashed lines indicate their time of rupture) or water bridges (thin dashed lines).

to a network of hydrogen bonds between the ligand and the binding pocket and show that water bridges substantially enhance the stability of the complex. Good agreement with experimental results was obtained. Further 'force simulations' of various systems, e.g., an antigen-antibody complex, are in progress.

4 Conformational Flooding

The previous application — in accord with most MD studies — illustrates the urgent need to further push the limits of MD simulations set by todays computer technology in order to bridge time scale gaps between theory and either experiments or biochemical processes. The latter often involve conformational motions of proteins, which typically occur at the microsecond to millisecond range. Prominent examples for functionally relevant conformational motions

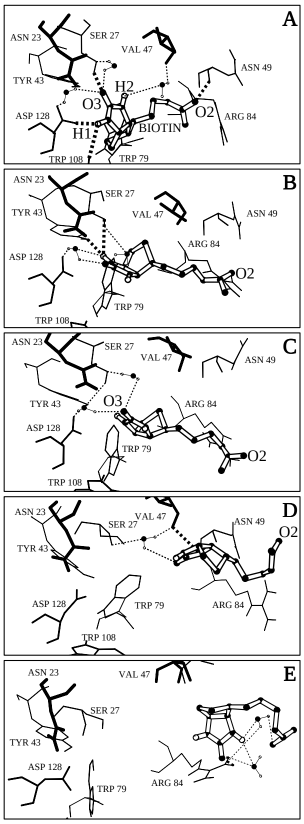

Fig. 7. 'Snapshots' of rupture taken (A) at the start of the simulation ($z_{\rm cant} = 0$), (B) at $z_{\rm cant} = 2.8$ Å, (C) at $z_{\rm cant} = 4.1$ Å, (D) at $z_{\rm cant} = 7.1$ Å, and (E) at $z_{\rm cant} = 10.5$ Å. The biotin molecule is drawn as a ball-and-stick model within the binding pocket (lines). The bold dashed lines show hydrogen bonds, the dotted lines show selected water bridges.

are the opening and closing of ion channels or, as proposed by Griffith [55] and Prusiner [56], pathogenic conformational transitions in prion proteins, the putative agents of mad cow and Creutzfeldt-Jacob diseases. Conformational motions often involve a complex and concerted rearrangement of many atoms in a protein from its initial state into a new conformation. These rearrangements, called conformational transitions, exhibit a multi-rate behaviour, which is is captured by the concept of "hierarchical conformational substates" introduced by Hans Frauenfelder [57]. According to that concept the free energy landscape of a protein exhibits a large number of nearly isoenergetic minima corresponding to the *conformational substates*, which are separated by barriers of different height [58].

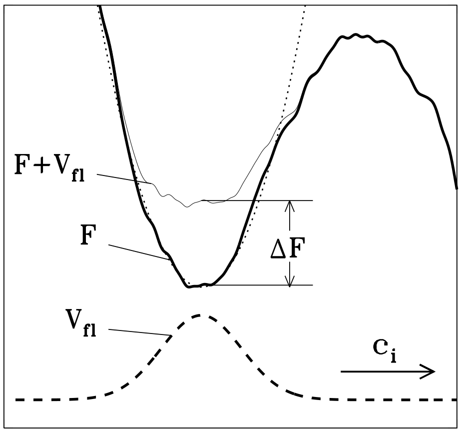

Fig. 8. 'Conformational flooding' lowers free energy barriers of conformational transitions and thus accelerates such transitions. The figure shows a one dimensional cut through the high dimensional free energy landscape F (bold line) along a particular conformational coordinate c_i. During an MD simulation the protein remains in the initial configuration (local minimum in the free energy), since the high barrier to the right cannot be overcome on an MD time scale. However, the MD simulation can serve to approximate the free energy harmonically in the vicinity of the initial configuration (dotted line) in order to derive an artificial 'flooding potential' V_{fl} (dashed line). Inclusion of this potential (thin line) in subsequent MD simulations reduces the barrier height by an amount of ΔF and thereby destabilizes the initial configuration.

Figure 8 shows a one-dimensional sketch of a small fraction of that energy landscape (bold line) including one conformational substate (minimum) as well as, to the right, one out of the typically huge number of barriers separating this local minimum from other ones. Keeping this picture in mind the conformational dynamics of a protein can be characterized as "jumps" between these local minima. At the MD time scale below nanoseconds only very low barriers can be overcome, so that the studied protein remains in or close to its initial conformational substate and no predictions of slower conformational transitions can be made.

In order to make such predictions possible, we have developed the *conformational flooding* (CF) method, which *accelerates* conformational transitions [59] and thereby brings them into the scope of MD simulations ("flooding simulations"). The method is a generalization of the "local elevation method" [60] in that it rests on a quasi harmonic model for the free energy landscape in the vicinity of the minimum representing the initial (known) conformational state. This model is derived from an ensemble of structures generated by a conventional MD simulation as will be described below and is shown in Fig. 9. From that model a "flooding potential" V_{fl} is constructed (dashed line in Fig. 8), which, when subsequently included into the poten-

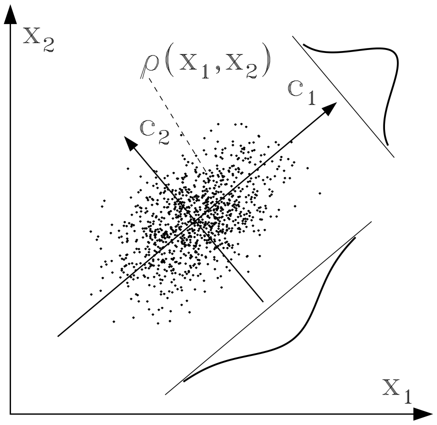

Fig. 9. Two-dimensional sketch of the 3N-dimensional configuration space of a protein. Shown are two Cartesian coordinates, x_1 and x_2, as well as two conformational coordinates (c_1 and c_2), which have been derived by principle component analysis of an ensemble ("cloud" of dots) generated by a conventional MD simulation, which approximates the configurational space density ρ in this region of configurational space. The width of the two Gaussians describe the size of the fluctuations along the configurational coordinates and are given by the eigenvalues λ_i.

tial energy function of the system, raises the minimum under consideration (thin line in Fig. 8) and thereby lowers the surrounding free energy barriers by an amount ΔF without severely modifying the barriers themselves. As a result, transitions over these barriers are accelerated by approximately the Boltzmann factor $\exp(\frac{\Delta F}{k_B T})$. In detail, the following steps are necessary to perform a CF simulation:

Step 1: A short conventional MD simulation (typically extending over a few 100 ps) is performed to generate an ensemble of protein structures $\{\mathbf{x} \in \mathcal{R}^{3N}\}$ (each described by N atomic positions), which characterizes the initial conformational substate. The 2-dimensional sketch in Fig. 9 shows such an ensemble as a cloud of dots, each dot $\mathbf{x}$ representing one "snapshot" of the protein.

Step 2: This ensemble is subjected to a "principal component analysis" (PCA) [61] by diagonalizing the covariance matrix $\mathbf{C} \in \{\mathcal{R}^{3N} \times \mathcal{R}^{3N}\}$,

$$\mathbf{C} := \langle (\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T \rangle \quad \text{with} \quad \bar{\mathbf{x}} = \langle \mathbf{x} \rangle$$

i.e.,

$$\mathbf{C} = \mathbf{Q}^T \mathbf{\Lambda}^{-1} \mathbf{Q}$$

with orthonormal $\mathbf{Q}$ and $\mathbf{\Lambda} = (\delta_{ij}\lambda_i) \in \{\mathcal{R}^{3N} \times \mathcal{R}^{3N}\}$, where $\langle \ldots \rangle$ denotes an average over the ensemble $\{\mathbf{x}\}$.

Step 3: The eigenvectors of $\mathbf{C}$ define $3N - 6$ collective coordinates (quasi particles) $\mathbf{q} := \mathbf{Q}(\mathbf{x} - \bar{\mathbf{x}})$, where we have eliminated the six rotational and translational degrees of freedom. From these $3N - 6$ degrees of freedom we select a number $m < 3N - 6$ *conformational coordinates* $\mathbf{c} = (c_1, \ldots, c_m)^T$ associated to the largest eigenvalues. Thus, the conformational coordinates cover most of the atomic fluctuations occurring at the 100 ps time scale. These m degrees of freedom are expected to dominate (not necessarily exclusively) conformational motion also at slower time scales [62–64].

Step 4: This PCA defines a *multivariate Gaussian model* $\tilde{\rho}^{\mathbf{c}}$,

$$\tilde{\rho}^{\mathbf{c}}(\mathbf{c}) \propto \exp\left[-\mathbf{c}^T \mathbf{\Lambda}_c \mathbf{c}/2\right]$$

of the conformational space density $\rho(\mathbf{c})$, from which the quasi harmonic approximation of the energy landscape,

$$\tilde{F}(\mathbf{c}) = -k_B T \ln\left[\tilde{\rho}(\mathbf{c})\right] = \frac{1}{2} k_B T \mathbf{c}^T \mathbf{\Lambda}_c \mathbf{c}$$

is derived (see Ref. [59]).

Step 5: From that model of the current substate we construct the *flooding potential* V_{fl} of strength E_{fl},

$$V_{\mathrm{fl}} = E_{\mathrm{fl}} \;\; \exp\left[-\frac{1}{2}\frac{k_B T \mathbf{c}^T \mathbf{\Lambda}_c \mathbf{c}}{E_{\mathrm{fl}}}\right],$$

which is included in a subsequent MD simulation within the energy function used in the conventional MD simulation before (see Fig. 8), thereby causing the desired acceleration of transitions.

As a sample application we describe simulations suggesting possible conformational transitions of the protein BPTI (Bovine Pancreatic Trypsin Inhibitor) at a time scale of several 100 nanoseconds (see Fig. 10). First we carried out a conventional MD simulation of 500 ps duration (no explicit solvent included), during which the protein remained in its initial conformational substate CS 1. The upper left part of the figure shows several snapshots of the backbone taken from that simulation; the lower left shows a projection of the 500 ps trajectory onto the two conformational coordinates with largest eigenvalues (corresponding to Fig. 9). From that ensemble we constructed a flooding potential as described above (dashed contour lines, superimposed to the CS 1-trajectory, bottom right). The flooding potential was subsequently switched on and rapidly induced a conformational transition (to the right in the figure) into another energy minimum, CS 2. After switching off the flooding potential the new conformational state of the protein remained stable, indicating that, indeed, the new minimum is separated from CS 1 by a large energy barrier. Using multi-dimensional transition state theory [59] we could estimate that in an conventional (i.e., unperturbed) MD simulation that conformational transition would have been observed only after several hundred

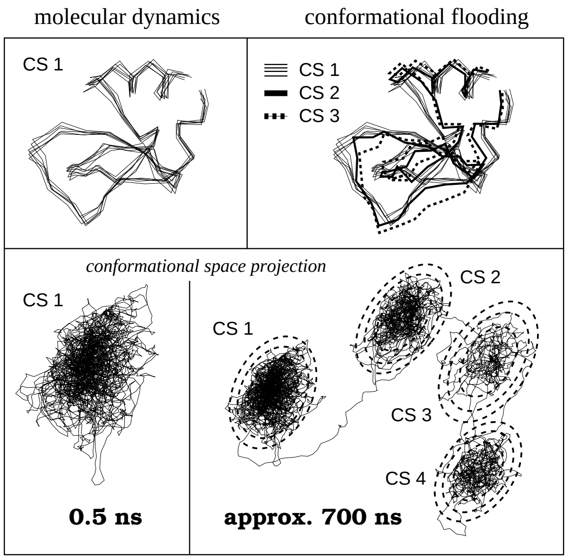

Fig. 10. 'Conformational flooding' accelerates conformational transitions and makes them accessible for MD simulations. Top left: snapshots of the protein backbone of BPTI during a 500 ps-MD simulation. Bottom left: a projection of the conformational coordinates contributing most to the atomic motions shows that, on that MD time scale, the system remains in its initial configuration (CS 1). Top right: 'Conformational flooding' forces the system into new conformations after crossing high energy barriers (CS 2, CS 3, ...). Bottom right: The projection visualizes the new conformations; they remain stable, even when the applied flooding potentials (dashed contour lines) is switched off.

nanoseconds. As shown in Fig. 10, the CF method can be applied iteratively to systematically search for further conformational substates, CS 3, CS 4 etc. The upper right part of the figure shows the backbone configuration of BPTI corresponding to the new substates.

MD simulations are valuable tools if one wants to gain detailed insight into *fast* dynamical processes of proteins and other biological macromolecules at atomic resolution. But since conventional MD simulations are confined to the study of very fast processes, conformational flooding represents a complementary and powerful tool to predict and understand also slow conformational motions. Another obvious application is an enhanced refinement of Xray- or NMR-structures.

References

1. M. Levitt and Shneior Lifson. Refinement of protein conformation using a macromolecular energy minimization procedure. *J. Mol. Biol.*, 46:269–279, 1969.

2. J. A. McCammon, B. R. Gelin, and M. Karplus. Dynamics of folded proteins. *Nature (London)*, 267:585–590, 1977.
3. W. F. van Gunsteren and H. J. C. Berendsen. Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.*, 34(5):1311–1327, 1977.
4. Olle Edholm, Oliver Berger, and Fritz Jähnig. Structure and fluctuations of bacteriorhodopsin in the purple membrane. *J. Mol. Biol.*, 250:94–111, 1995.
5. M. Levitt and R. Sharon. Accurate simulation of protein dynamics in solution. *Proc. Natl. Acad. Sci. USA*, 85:7557–7561, 1988.
6. Walter Nadler, Axel T. Brünger, Klaus Schulten, and Martin Karplus. Molecular and stochastic dynamics of proteins. *Proc. Natl. Acad. Sci. USA*, 84:7933–7937, Nov. 1987.
7. H. Kovacs, A.E. Mark, J. Johansson, and W.F. van Gunsteren. The effect of environment on the stability of an integral membrane helix: Molecular dynamics simulations of surfactant protein C in chloroform, methanol and water. *J. Mol. Biol.*, 247:808–822, 1995.
8. G.H. Peters, D.M.F van Aalten, O. Edholm, S. Toxvaerd, and R. Bywater. Dynamics of proteins in different solvent systems: Analysis of essential motion in lipases. *Biophys. J.*, 71:2245–2255, 1996.
9. M. C. Nuss, W. Zinth, W. Kaiser, E. Kölling, and D. Oesterhelt. Femtosecond spectroscopy of the first events of the photochemical cycle in bacteriorhodopsin. *Chem. Phys. Lett.*, 117(1):1–7, 1985.
10. Feng Zhou, Andreas Windemuth, and Klaus Schulten. Molecular-dynamis study of the proton pump cycle of bacteriorhodopsin. *Biochem.*, 32(9):2291–2306, 1993.
11. B. Leimkuhler and R. D. Skeel. *J. Comp. Phys.*, 112:117, 1994.
12. B.J. Leimkuhler, S. Reich, and R. D. Skeel. Integration methods for molecular dynamics. In *Mathematical approaches to biomolecular structure and dynamics*, Seiten 161–185, New York, 1996. Springer.
13. R. D. Skeel, G. H. Zhang, and T. Schlick. A family of symplectic integrators: Stability, accuracy, and molecular dynamics applications. *SIAM J. Scient. COMP.*, 18:203–222, 1997.
14. A. Ahmad and L. Cohen. A numerical integration scheme for the N-body gravitational problem. *J. Comp. Phys.*, 12:389–402, 1973.
15. W. B. Streett, D. J. Tildesley, and G. Saville. Multiple time step methods in molecular dynamics. *Mol. Phys.*, 35:639–648, 1978.
16. R. C. Y. Chin, G. W. Hedstrom, and F. A. Howes. *Considerations on Solving Problems with Multiple Scales*. Academic Press, Orlando, Florida, 1985.
17. Andreas Windemuth. Dynamiksimulation von Makromolekülen. Diplomarbeit, Technical University of Munich, Physics Department, T 30, James-Franck-Street, 8046 Garching, August 1988.
18. Mark E. Tuckerman, Glenn J. Martyna, and Bruce J. Berne. Molecular dynamics algorithm for condensed systems with multiple time scales. *J. Chem. Phys.*, 93(2):1287–1291, Jul. 1990.
19. Helmut Grubmüller, Helmut Heller, Andreas Windemuth, and Klaus Schulten. Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol. Sim.*, 6:121–142, 1991.
20. Mark E. Tuckerman and Bruce J. Berne. Molecular dynamics algorithm for multiple time scales: Systems with disparate masses. *J. Chem. Phys.*, 94(2):1465–1469, January 1991.

21. Mark E. Tuckerman, Bruce J. Berne, and Glenn J. Martyna. Molecular dynamics algorithm for multiple time scales: Systems with long range forces. *J. Chem. Phys.*, 94(10):6811–6815, May 1991.
22. Robert D. Skeel, Jeffrey J. Biesiadecki, and Daniel Okunbor. Symplectic integration for macromolecular dynamics. In *Proceedings of the International Conference Computation of Differential Equations and Dynamical Systems*. World Scientific Publishing Co., 1992. in press.
23. Robert D. Skeel and Jeffrey J. Biesiadecki. Symplectic integration with variable stepsize. *Ann. Num. Math.*, 1:191–198, 1994.
24. Helmut Grubmüller. *Molekulardynamik von Proteinen auf langen Zeitskalen*. Doktorarbeit, Technische Universität München, Germany, Jan. 1994.
25. D. Okunbor and R. D. Skeel. Explicit canonical methods for Hamiltonian systems. Working document, Numerical Computing Group, University of Illinois at Urbana-Champaign, 1991.
26. Andreas Windemuth. *Advanced Algorithms for Molecular Dynamics Simulation: The Program PMD*. ACS Books, 1995.
27. D. D. Humphreys, R. A. Friesner, and B. J. Berne. Simulated annealing of a protein in a continuum solvent by multiple-time-step molecular dynamics. *J. Phys. Chem.*, 99:10674–10685, 1995.
28. P. Procacci, T. Darden, and M. Marchi. A very fast molecular dynamics method to simulate biomolecular systems with realistic electrostatic interactions. *J. Phys. Chem.*, 100:10464–10468, 1996.
29. S. J. Stuart, R. Zhou, and B. J. Berne. Molecular dynamics with multiple time scales: The selection of efficient reference system propagators. *J. Chem. Phys.*, 105:1426–1436, 1996.
30. R. Zhou and B. J. Berne. A new molecular dynamics method combining the reference system propagator algorithm with a fast multipole method for simulating proteins and other complex systems. *J. Phys. Chem.*, 103:9444–9459, 1995.
31. T. Schlick, E. Bartha, and M. Mandziuk. Biomolecular dynamics at long timesteps: Bridging the timescale gap between simulation and experimentation. *Ann. Rev. Biophys. Biom. Structure*, 26:181–222, 1997.
32. Andrew. W. Appel. An efficient program for many-body simulation. *SIAM J. Sci. Stat. Comput.*, 6(1):85–103, January 1985.
33. Josh Barnes and Piet Hut. A hierarchical $o(n \log n)$ force-calculation algorithm. *Nature (London)*, 324:446–449, December 1986.
34. L. Greengard and V. Rokhlin. On the evaluation of electrostatic interactions in molecular modeling. *Chem. Scr.*, 29A:139–144, 1989.
35. James F. Leathrum and John A. Board. The parallel fast multipole algorithm in three dimensions. Technical report, Dept. of Electrical Engineering, Duke University, Durham, 1992.
36. C. Niedermeier and P. Tavan. A structure adapted multipole method for electrostatic interactions in protein dynamics. *J. Chem. Phys.*, 101:734–748, 1994.
37. Christoph Niedermeier. *Modellierung elektrostatischer Wechselwirkungen in Proteinen: Eine strukturadaptierte Multipolmethode*. Doktorarbeit, Ludwig-Maximilians-Universität, München, Germany, 1995.
38. C. Niedermeier and P. Tavan. Fast version of the structure adapted multipole method — efficient calculation of electrostatic forces in protein dynamics. *Mol. Sim.*, 17:57–66, 1996.

39. B. A. Luty, I. G. Tironi, and W. F. van Gunsteren. Lattice-sum methods for calculating electrostatic interactions in molecular simulations. *J. Chem. Phys.*, 103:3014–3021, 1995.
40. U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. The smooth particle mesh ewald method. *J. Chem. Phys.*, 103:8577, 1995.
41. Brock A. Luty, Ilario G. Tironi, and Wilfried F. van Gunsteren. Lattice-sum methods for calculating electrostatic interactions in molecular simulations. *J. Chem. Phys.*, 103:3014–3021, 1995.
42. Brock A. Luty and Wilfried F. van Gunsteren. Calculating electrostatic interactions using the particle–particle particle–mesh method with nonperiodic long-range interactions. *J. Phys. Chem.*, 100:2581–2587, 1996.
43. Bernhard R. Brooks, Robert E. Bruccoleri, Barry D. Olafson, David J. States, S. Swaminathan, and Martin Karplus. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.*, 4(2):187–217, 1983.
44. Charles L. Brooks III, B. Montgomery Pettitt, and Martin Karplus. Structural and energetic effects of truncating long ranged interactions in ionic and polar fluids. *J. Chem. Phys.*, 83(11):5897–5908, December 1985.
45. Richard J. Loncharich and Bernard R. Brooks. The effects of truncating long-range forces on protein dynamics. *Proteins*, 6:32–45, 1989.
46. M. Eichinger, H. Grubmüller, H. Heller, and P. Tavan. FAMUSAMM: An algorithm for rapid evaluation of electrostatic interaction in molecular dynamics simulations. *J. Comp. Chem.*, 18:1729–1749, 1997.
47. M. Eichinger. Paralleler schneller Multipolalgorithmus mit Mehrschrittverfahren für Molekulardynamiksimulationen. Diplomarbeit, Ludwig-Maximilians-Universität, München, Germany, 1995.
48. M. Eichinger, H. Grubmüller, and H. Heller. *User Manual for EGO_VIII, Release 2.0.* Theoretische Biophysik, Institut für Medizinische Optik, Ludwig-Maximilians-Universität, Theresienstr. 37, D-80333 München, Germany, 1995. Electronic access: http://www.imo.physik.uni-muenchen.de/ego.html.
49. Helmut Grubmüller, Berthold Heymann, and Paul Tavan. Ligand binding: Molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, 271(5251):997–999, 1996.
50. E.-L. Florin, V. T. Moy, and H. E. Gaub. Adhesion forces between individual ligand-receptor pairs. *Science*, 264:415–417, Apr. 15 1994.
51. N. M. Green. Avidin. *Adv. Protein Chem.*, 29:85, 1975.
52. S. Miyamoto and P. A. Kollman. Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics/free energy perturbation approaches. *Proteins*, 16:226–245, 1993.
53. S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.*, 72:1568–1581, 1997.
54. Evan Evans and Ken Ritchie. Dynamic strength of molecular adhesion bonds. *Biophys. J.*, 72:1541, 1997.
55. J. S. Griffith. *Nature (London)*, 215:1043–1044, 1967.
56. S. B. Prusiner. *Science*, 252:1515–1522, 1991.
57. Hans Frauenfelder, Sthephen G. Sligar, and Peter G. Wolynes. The energy landscape and motions of proteins. *Science*, 254:1598–1603, 1991.

58. R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, and I.C. Gunsalus. Dynamics of ligand binding to myoglobin. *Biochem.*, 14(24):5355–5373, 1975.
59. Helmut Grubmüller. Predicting slow structural transitions in macromolecular systems: conformational flooding. *Phys. Rev. E*, 52:2893, 1995.
60. T. Huber, A. E. Torda, and W. F. van Gunsteren. Local elevation: A method for improving the searching properties of molecular dynamics simulation. *J. of Computer-Aided Molecular Design*, 8:695–708, 1994.
61. J. C. Gower. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53:325, 1966.
62. A. Amadei, A. B. M. Linssen, and H. J. C. Berendsen. Essential dynamics of proteins. *Proteins*, 17:412–425, 1993.
63. Steven Hayward, Akio Kitao, and Nobuhiro Gō. Harmonic and anharmonic aspects in the dynamics of BPTI: A normal mode analysis and principal component analysis. *Physica Scripta*, 3:936–943, 1994.
64. H. Grubmüller, N. Ehrenhofer, and P. Tavan. Conformational dynamics of proteins: Beyond the nanosecond time scale. In M. Peyard, editor, *Proceedings of the Workshop 'Nonlinear Excitations in Biomolecules', May 30–June 4, 1994, Les Houches (France)*, Seiten 231–240. Centre de Physique des Houches (France), Springer-Verlag, 1995.