

Stochastic Dynamics Simulation for Macromolecules

A. Windemuth and K. Schulten
Beckman Institute and Department of Physics
University of Illinois
Urbana, IL 61801

January 31, 1991

Abstract

A new method for the simulation of macromolecules is proposed. The method is derived from classical Newtonian Dynamics by the substitution of random variables for atomic velocities. It combines features of both Newtonian Dynamics and Monte-Carlo simulation and reproduces the time scale of motion correctly. The resulting dynamics is equivalent to solving the Langevin equation of an overdamped system, obeys the Einstein relation and corresponds to a canonical ensemble description. We argue that the probability distribution for the positions of atoms is reproduced better than by Newtonian dynamics, due to a neglect of quantum effects in the latter. The stochastic nature of the proposed algorithm allows numerical restrictions on the length of integration intervals to be relaxed and simulation times to be extended beyond those of the Verlet algorithm for Newtonian Dynamics. The method underestimates, however, cross correlations of atomic velocities that give rise to concerted inertial motions of groups of atoms.

1 Introduction

Structural and dynamical properties of biological macromolecules like proteins and DNA, consisting of several thousand atoms and more, have been subject to intense study both experimentally and theoretically[1, 2, 3, 4, 5]. While experiments such as NMR-relaxation, fluorescence depolarisation and biochemical kinetics can furnish data over a wide range of relevant time scales, theoretical descriptions based on computer simulations have been confined to time scales shorter than a few nanoseconds, i.e. time scales which are not directly relevant for biomolecular functions. In this situation any method is most welcome that can reduce computational effort spent on molecular dynamics and, thereby, increase the accessible time scale of simulation.

Traditionally, two classes of methods have been employed to explore the configurational space and phase space of a molecule or of an ensemble of molecules: Monte-Carlo simulation[6, 7] and classical Newtonian Dynamics[8, 9]. Both methods treat a molecule as a system of particles interacting according to a heuristic potential function describing chemical bonds and non-bonding interactions[10, 11, 12, 13].

The method of Monte-Carlo simulation places emphasis on finding the probability distribution of states in the space of possible conformations. A new configuration of the molecule is chosen by randomly modifying the previous configuration and is then accepted or rejected according to the value of the potential. The potential gradient is not taken into account and does not have to be calculated. This method represents a canonical ensemble, because total energy is not conserved and temperature enters as a given parameter. Dynamical processes, especially large scale motions, are at best poorly described by standard Monte-Carlo methods.

The method of Newtonian dynamics, on the other hand, solves the classical equations of motion for a system of interacting atoms. The calculated trajectory provides a detailed picture of the dynamics of a macromolecule, making use of the potential gradient and correctly describing inertial effects. In its pure form, this method represents a microcanonical ensemble, since Newton's equation conserves total energy; the temperature defined through the kinetic energy, i.e. $\frac{3}{2}NkT = E_{\text{kin}}$, fluctuates. Several extensions of the method to describe a canonical ensemble have been suggested. One is rescaling of velocities during simulation to keep the temperature constant. Another is introducing random external forces and friction to simulate a heat bath[14].

More elaborate schemes have been proposed that use various synthetic (i.e. not existing in nature) Hamiltonians that are designed to ensure the maintenance of thermodynamic constraints on systems. They can be proven to yield the correct canonical ensemble while at the same time retaining the deterministic character of the simulation[15, 16, 17].

Such deterministic Newtonian dynamics methods neglect quantum effects. This generally leads to an overestimate of thermal motion in the real system. As the following analysis will show, erroneous results have to be expected for calculations concerned with heat capacity and energy dissipation due to the occurrence of high frequency vibrational modes in the simulation that cannot be excited above the ground state in the real system.

The energy of a molecular bond is

$$E_b(r) = k_b(r - r_o)^2; \quad k_b \approx 400 \text{ kcal mol}^{-1} \text{ \AA}^{-2} \quad (1)$$

which, for a bond between two carbon atoms, results in a transition energy of the first excited state

$$\hbar\omega = \hbar\sqrt{\frac{2k}{M_C}} \approx 0.1\text{eV}. \quad (2)$$

This value is much larger than the thermal energy per degree of freedom at room temperature ($k_B T \approx 0.025\text{eV}$). Therefore, such bond vibrations are confined to the quantum mechanical ground state and contribute little to heat capacity, i. e. exchange little energy with the other degrees of freedom of a macromolecule. Also, the probability density of bond lengths is Gaussian, in contrast to the $1/\cos x$ -distribution that a classical simulation yields. Since many vibrational modes in a biological macromolecule have frequencies ω such that $\hbar\omega > k_B T$ the above line of reasoning implies that Newtonian Dynamics cannot be considered as the single benchmark method against which all other molecular dynamics methods have to be measured.

In this article we want to present a new method for molecular dynamics simulation, the method of Uncorrelated Stochastic Dynamics and its variant, Correlated Stochastic Dynamics, in an effort to avoid the unnecessary exact computation of high frequency motions while retaining more of the dynamics than is possible when using Monte-Carlo methods. Like Newtonian Dynamics, however, the method proposed in this paper will not be able to describe the correct heat capacity of macromolecules. For this purpose a genuinely quantum mechanical method would have to be devised.

The method of Stochastic Dynamics does not conserve energy, i.e. it does not describe the atoms of a macromolecule as a microcanonical ensemble, but rather introduces random forces corresponding to a fixed temperature. Total energy then fluctuates around an equilibrium value determined by this temperature. The method, hence, describes the atoms of a macromolecule as a canonical ensemble and as such should be satisfactory for large macromolecules such as proteins.

In this first study we will formulate Stochastic Dynamics in close analogy to the widely used Newtonian Dynamics algorithm in that we express the equations of motion in terms of Cartesian coordinates and velocities, rather than in terms of internal and external coordinates and velocities. A Cartesian description allows one to incorporate the Stochastic Dynamics algorithm into conventional molecular dynamics programs. However, such an approach does not permit an obvious extension of the Stochastic Dynamics method which is to focus the description of macromolecules to a subset of degrees of freedom considered as a canonical ensemble. These degrees of freedom would be treated *explicitly* while the remaining *implicit* degrees of freedom would be accounted for by the application of heuristic thermal noise. Such a separation of *explicit* and *implicit* degrees of freedom will be qualitatively different if it involves only motions of such high frequency that due to quantum effects thermal excitation cannot be expected (see the discussion above) or if it also involves degrees of freedom which behave essentially classically. In the former case one would not expect exchange of thermal energy, i.e. the presence of quantum mechanical implicit degrees of freedom should not induce noise acting on the explicit degrees of freedom. Rather, these degrees of freedom should be described according to a suggestion by Nelson [18]. The separation of classical degrees of freedom could yield an increase of the integration step size proportional to the ratio between the highest frequencies of the implicit and of the explicit degrees of freedom. In principle, one may separate out all degrees of freedom of a protein except the slowest ones like dihedral angles and group rotations, and only integrate these degrees of freedom explicitly. However, since the implicit degrees of freedom in such a case exchange thermal energy with the explicit degrees of freedom one needs to account for the implicit degrees of freedom in terms of a heuristic thermal noise. Such a description requires preliminary studies on the effect of time scale separation and we consider the present investigation a first step in this direction.

In Section 2 of this paper we will introduce the method of Uncorrelated Stochastic Dynamics. The method replaces the deterministic atomic velocities in a difference equation derived from the Newtonian equations of motion by random variables. In Section 3 we will then modify this approach by including a time correlation of these random variables which accounts for the finite relaxation time of the velocity auto-correlation function. The size of the integration time step will be considered in Section 4. We will demonstrate the potential of our method for large reductions in computing time. In Section 5 calculations on the protein *bovine pancreatic trypsin inhibitor* (BPTI)[19] are presented, which were carried out using the molecular dynamics program MD developed by us[20].

2 Uncorrelated Stochastic Dynamics

Like other simulation methods, the Stochastic Dynamics methods suggested in this paper are based on a difference equation which describes how a system alters its state during a time step of length Δt . To derive this difference equation we first define the displacement $\Delta \mathbf{x}_i$ of an atom $i, i = 1, 2, \dots, N$ during a time interval Δt as

$$\Delta \mathbf{x}_i = \mathbf{x}_i(t + \Delta t) - \mathbf{x}_i(t) = \frac{1}{m_i} \int_0^{\Delta t} d\tau \mathbf{p}_i(t + \tau). \quad (3)$$

Expanding the momentum \mathbf{p}_i and applying Newtons equation of motion

$$\dot{\mathbf{p}}_i = -\nabla_i U(\mathbf{x}) = -\frac{\partial}{\partial \mathbf{x}_i} U(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N) \quad (4)$$

we obtain the discretized equation

$$\Delta \mathbf{x}_i = \frac{1}{m_i} \mathbf{p}_i(t) \Delta t - \frac{1}{2m_i} \nabla_i U[\mathbf{x}(t)] \Delta t^2 + O(\Delta t^3) \quad (5)$$

where $U(\mathbf{x}) = U(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N)$ is the potential function of the system. Omitting terms of the order Δt^3 and substituting

$$\mathbf{p}_i = m_i \frac{\mathbf{x}_i(t + \Delta t) - \mathbf{x}_i(t - \Delta t)}{2\Delta t} \quad (6)$$

leads to the well known propagation equation of the Verlet method

$$\mathbf{x}_i(t + \Delta t) = 2\mathbf{x}_i(t) - \mathbf{x}_i(t - \Delta t) + \frac{\Delta t^2}{m_i} \nabla_i U [\mathbf{x}(t)], \quad (7)$$

which is a second order predictor-corrector algorithm[11] for numerical integration of the classical equations of motion.

In the Uncorrelated Stochastic Dynamics method proposed in this paper the momenta \mathbf{p}_i in (5) are replaced by a Gaussian random process $\vec{\xi}_i(t)$. Accordingly, the displacements are

$$\Delta \mathbf{x}_i = \Delta \vec{\xi}_i - \frac{1}{2m_i} \nabla_i U \Delta t^2. \quad (8)$$

In this expression, the mean square of the Gaussian random variable $\Delta \vec{\xi}_i$ representing atomic velocities is related to thermal kinetic energy by

$$\langle \Delta \vec{\xi}_i^2 \rangle = \frac{\Delta t^2}{m_i^2} \langle \mathbf{p}_i^2 \rangle = \frac{3k_B T}{m_i} \Delta t^2. \quad (9)$$

According to (8), atomic displacements consist of two terms, one of which is random, whereas the other is determined by the gradient of the potential function. This observation motivates a comparison of the difference equation (8) to a corresponding Langevin equation which is similarly divided into stochastic and deterministic terms.

The Langevin equation for an overdamped system is[21]

$$\dot{\mathbf{x}}_i = \vec{\xi}_i(t) - \gamma_i \nabla_i U(\mathbf{x}), \quad (10)$$

where $\vec{\xi}_i(t)$ is a random process (Gaussian white noise) defined by

$$\begin{aligned} \langle \vec{\xi}_i(t) \rangle &= 0 \\ \langle \vec{\xi}_i(t) \vec{\xi}_j(t') \rangle &= 6D_i \delta_{ij} \delta(t - t') \end{aligned} \quad (11)$$

and γ_i is a mobility constant. The factor 6 is due to the fact that the $\vec{\xi}_i$ are three-dimensional vectors. The diffusion constants D_i and γ_i are related by the well-known Einstein relation [21]

$$D_i = \gamma_i k_B T, \quad (12)$$

a special case of the fluctuation-dissipation theorem, which ascertains that the system described by (10) reaches thermal equilibrium described by a Boltzman distribution $p(\mathbf{x}) = \exp(-U(\mathbf{x})/k_B T)$. The displacement resulting from (10) after a time interval Δt is

$$\Delta \mathbf{x}_i = \Delta \vec{\xi}_i - \gamma_i \nabla_i U \Delta t \quad (13)$$

where $\Delta \vec{\xi}_i$ is a random variable defined by

$$\Delta \vec{\xi}_i = \int_t^{t+\Delta t} d\tau \vec{\xi}_i(\tau). \quad (14)$$

The mean-square random displacement can be calculated from (14) and (11) as

$$\langle \Delta \vec{\xi}_i^2 \rangle = \int_t^{t+\Delta t} d\tau \int_t^{t+\Delta t} d\tau' \langle \vec{\xi}_i(\tau) \vec{\xi}_i(\tau') \rangle = 6D_i \Delta t. \quad (15)$$

Comparison of (8) and (13) shows that the Uncorrelated Stochastic Dynamics method is equivalent to integration of the Langevin equation (10) if the mobility constant γ_i is chosen to be

$$\gamma_i = \frac{1}{2m_i} \Delta t. \quad (16)$$

Comparison of (9), (15) and (16) shows that the Einstein relation (12) is satisfied for the Uncorrelated Stochastic Dynamics method.

3 Correlated Stochastic Dynamics

By substituting a random variable for particle momenta, time correlations caused by inertia are completely lost. In order to compensate for the one-particle part of these correlations we extended the method of Stochastic Dynamics to account for velocity auto-correlation. However, for the sake of computational feasibility we do not include cross-correlation of velocities of different atoms.

In such a description, displacements $\Delta \vec{\xi}_i$ at different times are not independent, instead, they are generated according to an Ornstein-Uhlenbeck process[22]

$$\Delta \vec{\xi}_i(t + \Delta t) = (1 - k) \Delta \vec{\xi}_i(t) + \Delta \vec{\eta}_i. \quad (17)$$

This process implies, that the value of $\Delta\vec{\xi}_i(t+\Delta t)$ contains a fraction $(1-k)$ of the last random displacement $\Delta\vec{\xi}_i(t)$. $\Delta\vec{\eta}_i$ represents uncorrelated Gaussian white noise as described by (11). The parameter k can be chosen to reproduce the observed decay of the velocity auto-correlation function $\langle\vec{v}_i(t)\cdot\vec{v}_i(0)\rangle$ of atom i . This decay indicates how long an atom is likely to move along its original direction before being diverted by interacting with other atoms. Such

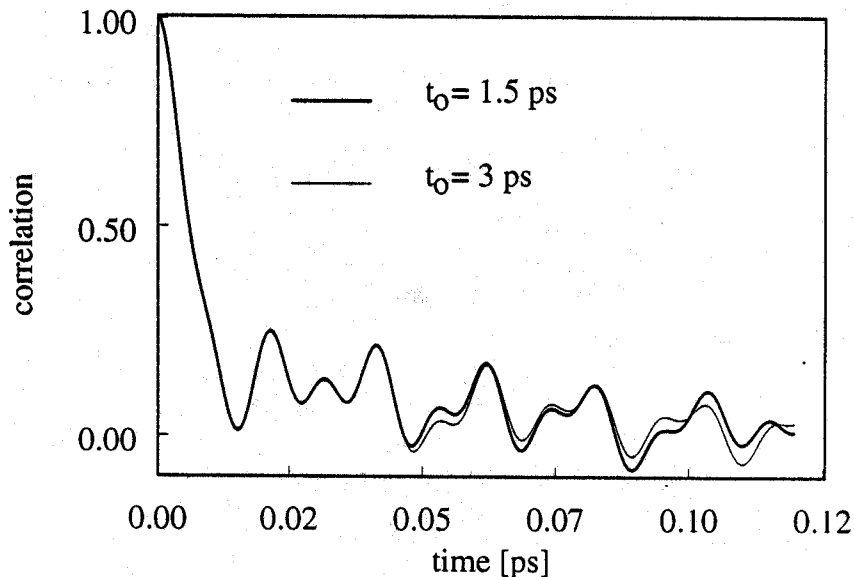


Figure 1: Velocity auto-correlation function $C_v(t)$ resulting from a Newtonian Dynamics simulation of BPTI. $C_v(t)$ defined in (18) has been calculated for two different values of t_0 taking the average $\langle \dots \rangle$ over a time $T = 0.1$ ps. As one can see a relaxation time of $\tau = 5$ fs describes the initial decay of $C_v(t)$. Oscillations, due to the finite number of degrees of freedom[23], are not accounted for by the stochastic model (17).

velocity autocorrelation functions have been determined in earlier molecular dynamics simulations by Brooks et al. [14] and by Nadler et al. [5] for different proteins. The decay times estimated from our simulation are somewhat

smaller than those in [14] (10-20 fs), which in turn are smaller than those calculated in [5].

Figure 1

represents the normalized velocity auto-correlation function

$$C_v(t) = \frac{1}{N} \sum_{i=1}^N \frac{\langle \vec{v}_i(t_0) \vec{v}_i(t_0 + t) \rangle - \langle \vec{v}_i \rangle^2}{\langle \vec{v}_i^2 \rangle - \langle \vec{v}_i \rangle^2} \quad (18)$$

averaged over all atoms of the protein bovine pancreatic trypsin inhibitor (BPTI). The function was computed from a conventional molecular dynamics simulation using a time average $\frac{1}{T} \int_0^T \dots dt$ for the mean $\langle \dots \rangle$. For this approximation to be valid, we have to assume that the system is approximately ergodic and near equilibrium. To test this assumption, we evaluated $C_v(t)$ for two different times t_0 . The results shown in Fig. 1 demonstrate that $C_v(t)$ is essentially independent of t_0 .

The velocity correlation function for the correlated stochastic process described by (17) in the limit of small Δt is

$$C_v(t) = e^{-\frac{t}{\tau}} \quad (19)$$

where the parameter k is related to τ through

$$1 - k = e^{-\frac{\Delta t}{\tau}}. \quad (20)$$

In our calculations we assumed a value of 5 fs for the decay time τ which accounts for the decay of the observed velocity correlation function shown in Fig. 1. We will refer to the simulation method based on (8) and (17) as Correlated Stochastic Dynamics.

For the sake of simplicity we have assumed identical decay times τ for all atomic velocities. However, the Correlated Stochastic Dynamics is computationally feasible also with a choice of individual decay times for atoms.

The amplitude of the random variable $\Delta \eta_i$ in (17) can be determined by calculating the mean square of both sides of (17), recognizing that $\Delta \vec{\eta}_i(t)$ is independent of $\Delta \vec{\xi}_i(t)$, which only depends on $\Delta \vec{\eta}_i$ from previous time steps. This yields

$$\langle \Delta \vec{\xi}_i^2 \rangle = (1 - k)^2 \langle \Delta \vec{\xi}_i^2 \rangle + \langle \Delta \vec{\eta}_i^2 \rangle, \quad (21)$$

from which the desired mean square value can be derived with the help of (9), i.e.

$$\langle \Delta \vec{\eta}_i^2 \rangle = k(2 - k) \langle \Delta \vec{\xi}_i^2 \rangle = k(2 - k) \frac{3k_B T}{m_i} \Delta t^2. \quad (22)$$

If one chooses the random variables $\Delta\vec{\eta}_i$ as a Gaussian process with mean square amplitude (22) the Correlated Stochastic Dynamics method yields the proper thermal equilibrium and, hence, indeed describes a canonical ensemble.

4 The Discretisation Time Step

The computing time required for a molecular dynamics simulation is determined by the integration time step Δt . This time step is dictated by the fastest motion in the molecule, usually stretch vibrations with periods in the femtosecond range. Figure 2 shows the numerical fluctuations of the total

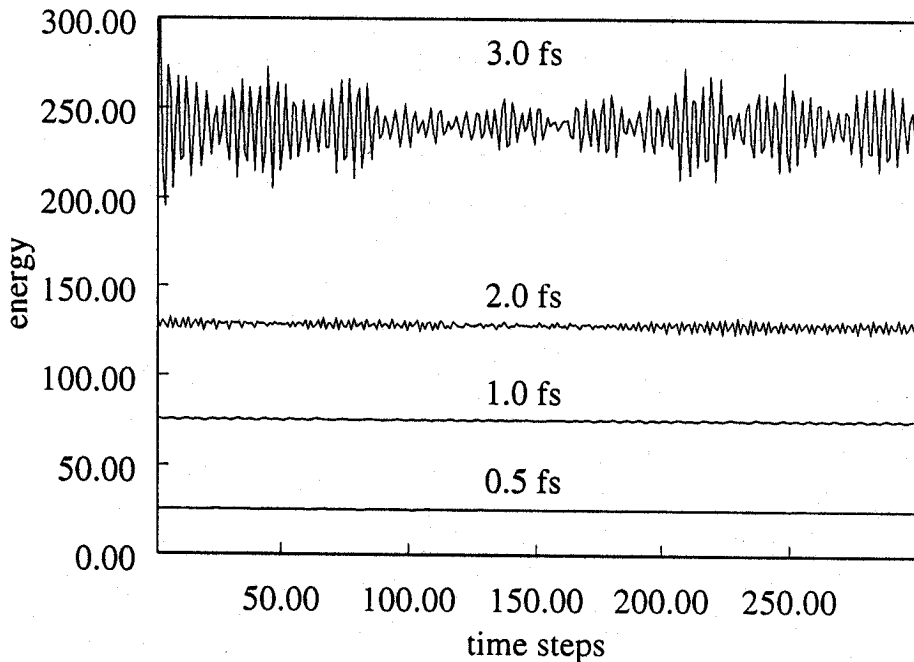


Figure 2: Total energy during Newtonian Dynamics simulation of BPTI using several different time steps Δt . Energy units are arbitrary. The curves for different Δt are offset to allow easy comparison of the fluctuations. Because the law of energy conservation holds under Newtonian Dynamics simulation, the fluctuations seen are numerical errors due to the finite size of Δt .

energy (potential and kinetic) during our molecular dynamics simulation of BPTI. The fluctuations depend sensitively on the time step Δt . When Δt becomes comparable to the 10 fs period of C—H bond stretch vibrations numerical errors lead to a violation of energy conservation. Stochastic Dynamics is inherently less sensitive to numerical inaccuracies regarding high frequency vibrations. This makes it possible to use longer time steps Δt . In order to determine an upper bound to the time step we investigated the behaviour of potential energy when the coordinates are shifted by a displacement of all atomic coordinates $\Delta \vec{x}_i$. For this purpose the equilibrated structure of BPTI was changed by a displacement

$$\Delta \vec{x}_i = -\lambda \frac{\nabla_i U}{\|\nabla_i U\|}, \quad (23)$$

where the norm $\|\dots\|$ is defined as

$$\|\vec{x}_i\| = \sqrt{\sum_{i=1}^N |\vec{x}_i|^2}. \quad (24)$$

The potential energies arising after such displacement $E_{\text{elec}}(\vec{x}_i + \Delta \vec{x}_i)$, $E_{\text{vdw}}(\vec{x}_i + \Delta \vec{x}_i)$, $E_{\theta}(\vec{x}_i + \Delta \vec{x}_i)$ describing Coulomb, van der Waals, bond angle forces, respectively, and the total potential energy $E_{\text{pot}}(\vec{x}_i + \Delta \vec{x}_i)$ were calculated for different displacement lengths λ . The results are shown in Fig. 3. For small λ , the potential surface is approximately linear, which means that all energy values decrease with increasing λ . This is the region where discrete integration is almost exact. The value of λ for one step of the Verlet integration method (5) and (7)

$$\lambda = \left\| \frac{\vec{p}_i}{m_i} \right\| \Delta t \quad (25)$$

with a time step of $\Delta t = 0.5$ fs is $\lambda = 0.0075 \text{ \AA}$. This value, indicated in Fig. 3, has been calculated by evaluating (25) explicitly during the simulation and corresponds very well to an estimate derived from (25) assuming an equilibrium mean velocity for each atom, i.e.

$$\left\| \frac{\vec{p}_i}{m_i} \right\| = \sqrt{\sum_{i=1}^N \frac{3k_B T}{m_i}}. \quad (26)$$

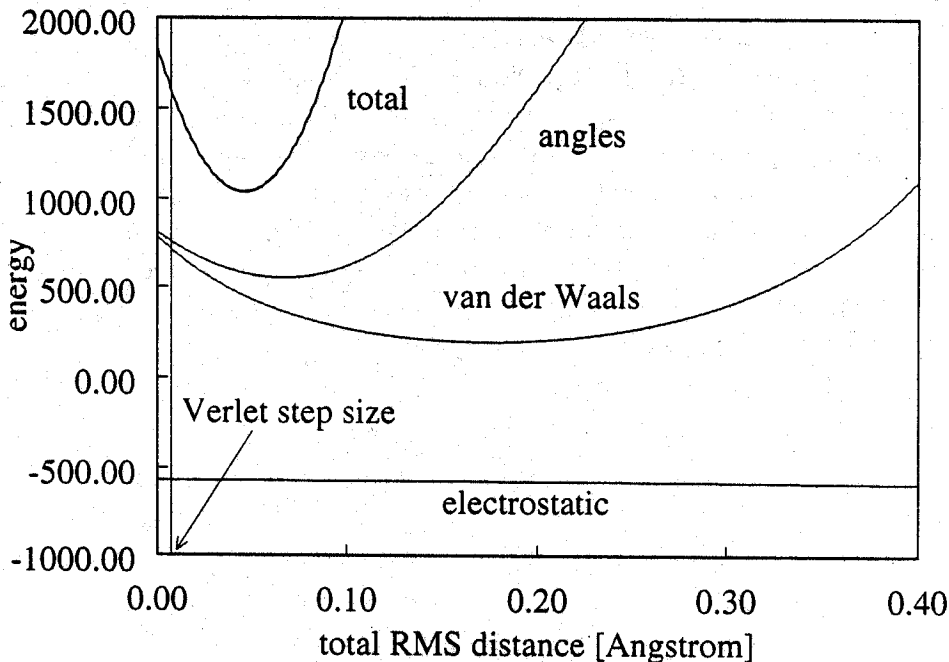


Figure 3: Potential energy plotted against the displacement of atoms in the direction of the negative potential gradient $-\nabla_i U$. In the immediate vicinity of the point where the gradient has been calculated, potential energy naturally decreases. However, due to the complex curvature of the energy landscape at a certain distance energy again increases, indicating the maximum displacement that can be tolerated during one simulation step.

At larger displacements the non-linear nature of the energy function takes over and the total potential energy rises steeply with λ . The upper bound for the length of a time step is the point where the total potential energy rises above the value at the starting point. Any attempt to integrate with such a step size would lead to a divergence of the total potential energy. It can be seen from Fig. 3, though, that a time step of ten times the size of the Verlet time step would still be acceptable from this point of view. Tests with both Correlated and Uncorrelated Stochastic Dynamics have shown that a stable simulation can be achieved with a time step of 3 fs, which is six times longer than step sizes required for the conventional Molecular Dynamics method. This implies that the Stochastic Dynamics methods yield a sixfold increase

in accessible simulation time over that accessible by Newtonian Dynamics.

5 Comparison of methods

The three types of dynamics simulations described above, Newtonian Dynamics, Uncorrelated Stochastic Dynamics and Correlated Stochastic Dynamics, were tested on the protein BPTI. The potential function employed for all simulations is PARAM11.PRO of the CHARMM program for molecular dynamics[10]. The simulations were carried out by means of the molecular dynamics program MD developed by us previously [20] in the C programming language. This program is based on CHARMM [10] in that it is input and output compatible with CHARMM, and in that it uses the same potential function. So far, the program has been implemented on a VAX/750, a Convex C1, a Silicon Graphics workstation, a NeXT Computer, a Motorola 88000 processor, a SUN/4, a CDC Cyber vector processor, a Cray 2 and a Connection Machine CM2[24].

The initial structure of BPTI for the simulations was taken from the Brookhaven Protein Data Bank[19]. This structure was minimized and equilibrated using standard procedures to generate a relaxed starting configuration used in all three simulations. The results reported here were extracted from the trajectory during the simulation and stored for later analysis, the full trajectories were not saved due to storage space limitations.

The calculation of 1 ps of Newtonian Dynamics took about 6.6 hours of CPU-time on a Convex C1 using a time step of $\Delta t = 0.5$ fs as recommended in [10]. For the Uncorrelated and Correlated Stochastic Dynamics methods, due to the larger time step of $\Delta t = 3$ fs, the same period required only 1.1 hours of CPU-time. For the Correlated Stochastic Dynamics method, a correlation time of 5 fs was assumed, in accordance with the relaxation time of the velocity auto-correlation function in Fig. 1.

5.1 Energies

The properties we like to compare first for the three simulations are the total energy and various components of the potential energy. In Fig. 4 the potential energy for bond angles, electrostatic forces, the van der Waals-energy and the total energy (including kinetic energy) are compared for the three sim-

ulation methods. In the Newtonian Dynamics simulation, the law of energy conservation requires the total energy to be constant, which is borne out in the Verlet simulation. In the Stochastic Dynamics simulations total energy is not conserved due to the random forces; instead the system behaves as a canonical ensemble with a fixed temperature of 300 K. If bond stretching and angle bending are considered as independent harmonic degrees of freedom, each contributing $\frac{1}{2}k_B T$ to the potential energy, the expected potential energies are of the order of 157 kcal/mol for bond stretching and 250 kcal/mol for angle bending. These values were calculated from $k_B T \approx 0.6$ kcal/mol and from the fact that the molecular structure of BPTI includes 582 bonds and 834 angles. The actual potential energy values observed in the Newtonian Dynamics simulation are larger, measuring about 250 kcal/mol and 300 kcal/mol, respectively. The excess energy can be attributed to distortion of the molecule by van der Waals and electrostatic forces, which tend to increase bond stretch and bending potential energies by induction of structural stress. The comparison of Newtonian Dynamics simulation with both kinds of Stochastic Dynamics simulation shows that the potential energy is smaller when described by Uncorrelated Stochastic Dynamics and larger when described by Correlated Stochastic Dynamics. We conclude that the correlation introduced in (17), which does not depend on the potential function, tends to drive the atoms up against potential barriers, thereby increasing the potential energy. This effect is much less pronounced in Newtonian Dynamics, in which case the correlation of velocities $\vec{p}_i(t)/m_i$ and $\vec{p}_i(t + \Delta t)/m_i$ takes the gradient of the potential function into account.

5.2 Coordinate Drift

To evaluate the capability of the different simulation methods to sample phase space, we calculated the overall coordinate drift, defined by

$$d(t) = \|\mathbf{x}_i(t) - \mathbf{x}_i(t_0)\| \quad (27)$$

$d(t)$ is a measure of the diameter of the probability distribution in phase space and, thus, indicates how large a region of phase space a method can explore. The coordinate drift was calculated at every time step during the simulations. The result is shown in Fig. 5. The slope at the beginning of the simulations is the same for all three methods, showing that the time

scale is reproduced correctly. The probability distribution resulting from the Uncorrelated and Correlated Stochastic Dynamics methods expands in phase space much more slowly as compared to that resulting from Newtonian Dynamics simulation. This can be understood by considering the expansion of a free system, i.e. of a system of atoms without a potential function. The corresponding coordinate drifts are also indicated in Fig. 5. In the case of the Newtonian Dynamics simulation the free expansion is linear, i.e. the system behaves like an ideal gas without container. The Uncorrelated Stochastic Dynamics simulation of a free system is equivalent to a multi-dimensional random walk and the expansion in phase space $d(t)$ is proportional to \sqrt{t} . The presence of a potential restricts the accessible phase space and, thereby, diminishes the expansion of the probability distribution in all three cases. This is demonstrated clearly in Fig. 5. The addition of time correlation to Stochastic Dynamics has a significant effect: Correlated Stochastic Dynamics leads to a faster expansion as compared to Uncorrelated Stochastic Dynamics, the expansion being even faster than that of the free random walk.

5.3 Ring Rotation

To evaluate the effect of Stochastic Dynamics on collective motions of groups of atoms, the rotation of rings of tyrosine and phenylalanin side chains of BPTI has been monitored during the simulations. These rings are embedded in the protein and are usually prevented by the surrounding protein matter from rotating by more than a few degrees. However, they have been observed experimentally to occasionally 'flip', i.e. rotate by 180° . Such flips are activated processes which are too rare to be reproduced directly in computer simulations[25]. Nevertheless, the equilibrium motions of rings can be used to test the accuracy of the description of such collective motions by simulation techniques. Figure 6 shows the rotation angles of some of the rings in BPTI as a function of time. Differences between the methods are apparent. Large scale motions are nearly absent in the Uncorrelated Stochastic Dynamics simulation. The introduction of a finite correlation time (5 ps) leads to a significant increase of the ring angle fluctuations. As with the coordinate drift, the results of Correlated Stochastic Dynamics simulations are intermediate between those of Uncorrelated Stochastic Dynamics and those of Newtonian Dynamics.

6 Conclusion

We have presented above a molecular dynamics method which allows one to describe motions of macromolecules as part of a canonical ensemble. The method takes advantage of the fact that in macromolecules velocities are correlated only on a very short time scale, typically a few femtoseconds, and accordingly describes velocities by a stochastic model with the characteristics of a Maxwell distribution and finite or vanishing correlation times. The present study demonstrates the feasibility of such approach. In the following we like to discuss modelling techniques, some already mentioned above, which become available in the framework of Stochastic Molecular Dynamics.

The computational procedures involved in the Stochastic Dynamics algorithm are very similar to those of the conventional Newtonian Dynamics method. This opens up a class of further developments of the algorithm identical to those suggested for Newtonian dynamics simulations. The most notable development appears to be that the algorithm lends itself to implementation on various massive parallel computers, e.g. of the SIMD type [24] or of the MIMD type [26]. A further improvement would involve new algorithms for the computation of non-bonded forces, i.e. the Fast Multipole Algorithm [27] or the distance class algorithm [28]; the latter algorithm can be integrated in a straightforward way into the method of Stochastic Dynamics. Also established computational techniques such as the stochastic boundary method [14], the optimization routines involving artificial forces to match NMR and X-ray crystallographic data [29, 30], and the thermodynamic perturbation method [31] can be readily connected with Stochastic Dynamics.

Two obvious extensions specific for Stochastic Dynamics have been mentioned above. First, one can assign individual relaxation times for velocity auto-correlation of atoms. More cumbersome, but possible, is an inclusion of velocity cross-correlation for groups of n_j atoms, e.g. atoms belonging to amino acid side chains with rings. In this case the N uncoupled Eqs. (17) are replaced by sets of n_j ($\sum_j n_j = N$) coupled equations, the tensorial coupling to be determined either from Newtonian dynamics or from a model of rigid ring motion.

The most promising avenue for further development should stem from the possibility to separate the degrees of freedom of a macromolecule into explicit and implicit degrees, i.e. to focus the dynamics on the relevant (explicit) de-

degrees of freedom, viewing this subsystem as a canonical subensemble in the whole macromolecule. For many important problems in macromolecular dynamics, e.g. the description of activated processes or of protein folding, the relevant degrees of freedom are very slow and a separate description would allow one to increase integration step sizes by orders of magnitude. Such a possibility would necessitate, however, that the remaining implicit degrees of freedom, which establish a heat bath for the explicit degrees of freedom, are accounted for by a proper noise term. We suggest that systematic explorations be carried out to establish to what extent this avenue furnishes proper descriptions. We have already commented above that the separation of degrees of freedom which represent high frequency vibrations and, therefore, should be treated quantum mechanically would be particularly simple since (i) these degrees of freedom exchange little energy with the classical degrees of freedom, thus contributing little thermal noise, and since (ii) according to the theory of Nelson[18] they can be described probabilistically. A Nelson-type description would bypass velocities altogether and assign position values (bond distances and bond angles) according to the respective ground state wave functions.

The development outlined hinges on the development of computational techniques which appear to be long overdue in the field of molecular dynamics, namely, a formulation of equations of motion in terms of generalized coordinates rather than Cartesian coordinates. Certainly this task is not simple. The Stochastic Dynamics method which only involves displacement and force terms, but no inertial terms stemming from second order derivatives, should provide a simpler avenue to such a description than the Newtonian Dynamics method. Generalized coordinates may be chosen to be identical to the so-called internal coordinates, i.e. bond lengths, bond angles, torsions. However, even better results could be achieved if one would choose local rotational and vibrational modes to represent degrees of freedom corresponding to group vibrational modes and group rotations. Such modes can be obtained from vibrational analysis routines, e.g. those in [32, 33]. The generalized coordinate approach would require development of a new generation of dynamics programs, the birth of which hopefully will not be prevented by inertial effects exerted by the vast codes of existing molecular dynamics routines.

7 Acknowledgements

This work has been supported by a fellowship of the Böhlinger Ingelheim Fonds to A.W., by the University of Illinois at Urbana-Champaign and by the National Science Foundation, USA, through the National Center for Supercomputer Applications.

References

- [1] M. Karplus and J. A. McCammon. Dynamics of proteins: Elements and function. *Ann. Rev. Biochem.*, 53:263, 1983.
- [2] H. Treutlein, K. Schulten, J. Deisenhofer, H. Michel, A. Brünger, and M. Karplus. Molecular dynamics simulation of the primary processes of the photosynthetic reaction center of *rps. viridis*. In J. Breton and A. Vermeglio, editors, *The photosynthetic bacterial reaction center: Structure and dynamics*, pages 139–150, London, 1988. Plenum Press.
- [3] G. Lipari, A. Szabo, and R. M. Levy. Protein dynamics and NMR relaxation: comparison of simulations with experiment. *Nature*, 300:197, 1982.
- [4] R. M. Levy, R. P. Sheridan, J. W. Keepers, G. S. Dubey, S. Swaminathan, and M. Karplus. Molecular dynamics of myoglobin at 298_o K: results from a 300-ps computer simulation. *Biophys. J.*, 48:509, 1985.
- [5] W. Nadler, A. T. Brünger, K. Schulten, and M. Karplus. Molecular and stochastic dynamics of proteins. *Proc. Natl. Acad. Sci. USA*, 84:7933, 1987.
- [6] Kim Sharp, Richard Fine, Klaus Schulten, and Barry Honig. Brownian dynamics simulation of diffusion to irregular bodies. *Phys. Chem.*, 91:3624, 1987.
- [7] Zhenqin Li and H. A. Scheraga. Monte carlo-minimization approach to the multiple-minima problem in protein folding. *Proc. Natl. Acad. Sci. USA*, 84:6611, 1987.
- [8] J. A. McCammon. Protein dynamics. *Rep. Prog. Phys.*, 47:1, 1984.

- [9] P. G. Debrunner and H. Frauenfelder. Dynamics of proteins. *Ann. Rev Phys. Chem.*, 3:283, 1982.
- [10] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. CHARMM: A program for macromolecular energy minimization, and dynamics calculations. *J. Comp. Chem.*, 4:187, 1983.
- [11] W. F. van Gunsteren and H. J. C. Berendsen. Algorithms for macromolecular dynamics and constraint dynamics. *Molec. Phys.*, 34:1311, 1977.
- [12] M. Levitt and S. Lifson. Refinement of protein conformations using a macromolecular energy minimization procedure. *J. Molec. Biol.*, 46:269, 1969.
- [13] E. Clementi, G. Corongiu, M. Aida, U. Nieşar, and G. Kneller. Monte carlo and molecular dynamics simulations. In Enrico Clementi, editor, *MOTECC, Modern Techniques in Computational Chemistry*, pages 805–882. ESCOM, Leiden, 1990.
- [14] C. L. Brooks, A. Brünger, and M. Karplus. Active site dynamics in protein molecules: A stochastic boundary molecular-dynamics approach. *Biopolymers*, 24:843, 1985.
- [15] S. Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molec. Phys.*, 52:255–268, 1984.
- [16] D. J. Evans and G. P. Moriss. Isothermal–isobaric molecular dynamics. *Chem. Phys.*, 77:63–66, 1983.
- [17] D. J. Evans and G. P. Moriss. non-newtonian molecular dynamics. *Comp. Phys. Rep.*, 1:297–343, 1984.
- [18] E. Nelson. Connection between brownian motion and quantum mechanics. In H. Nelkowski, A. Hermann, H. Poser, R. Schrader, and R. Seiler, editors, *Lecture Notes in Physics 100, Einstein Symposium Berlin*, Berlin, 1979. Springer–Verlag.

- [19] J. Deisenhofer and W. Steigemann. Crystallographic refinement of the structure of bovine pancreatic trypsin inhibitor at 1.5 Å resolution. *Acta Cryst. B*, 31:238, 1975.
- [20] A. Windemuth. Dynamiksimulation von Makromolekülen. Master's thesis, Technische Universität München, München, 1988.
- [21] C. Kittel. *Elementary Statistical Physics*. John Wiley & Sons, Berlin, Heidelberg, New York, 1985.
- [22] C. W. Gardiner. *Handbook of Stochastic Methods*. Springer-Verlag, Berlin, Heidelberg, New York, 1985, 1985.
- [23] P. Mazur and E. Montroll. Poincaré cycles, ergodicity, and irreversibility in assemblies of coupled harmonic oscillators. *J. Math. Phys.*, 1:70, 1960.
- [24] A. Windemuth and K. Schulten. Molecular dynamics on the connection machine. *Molecular Simulation*, in press.
- [25] J. A. McCammon and M. Karplus. Dynamics of activated processes in globular proteins. *Proc. Natl. Acad. Sci. USA*, 76:3585, 1987.
- [26] H. Grubmüller, H. Heller, and K. Schulten. Molecular dynamics simulation on a parallel computer. *Molecular Simulation*, 5:133–165, 1990.
- [27] L. Greengard and V. Rohklin. A fast algorithm for particle simulations. *J. Comp. Phys.*, 73:325–348, 1987.
- [28] H. Grubmüller, H. Heller, A. Windemuth, and K. Schulten. Generalized verlet algorithm for efficient molecular dynamics simulations with long-range interactions. Submitted to *Molecular Simulation*.
- [29] A. T. Brünger, R. L. Campbell, G. M. Clore, A. M. Gronenborn, M. Karplus, G. A. Petsko, and M. M. Teeter. Solution of a protein structure with a model obtained from NMR interproton distance restraints. *Science*, 235:1049, 1987.
- [30] A. Brünger, J. Kuriyan, and M. Karplus. Crystallographic r factor refinement by molecular dynamics. *Science*, 235:458, 1987.

- [31] Shashidar N. Rao et al. Free energy perturbation calculations on binding and catalysis after mutating asn 155 in subtilisin. *Nature*, 328:551, 1987.
- [32] M. J. S. Dewar and W. Thiel. Ground states of molecules. 38. the mndo method. *J. Am. Chem. Soc.*, 99:4899, 1977.
- [33] M. J. S. Dewar and J. J. P. Stewart. AMPAC. *Quantum Chemistry Program Exchange Bulletin*, 6:24, 1986.

List of Figures

1	1	8
2	1	10
3	1	12
4	1	22
5	1	23
6	1	24

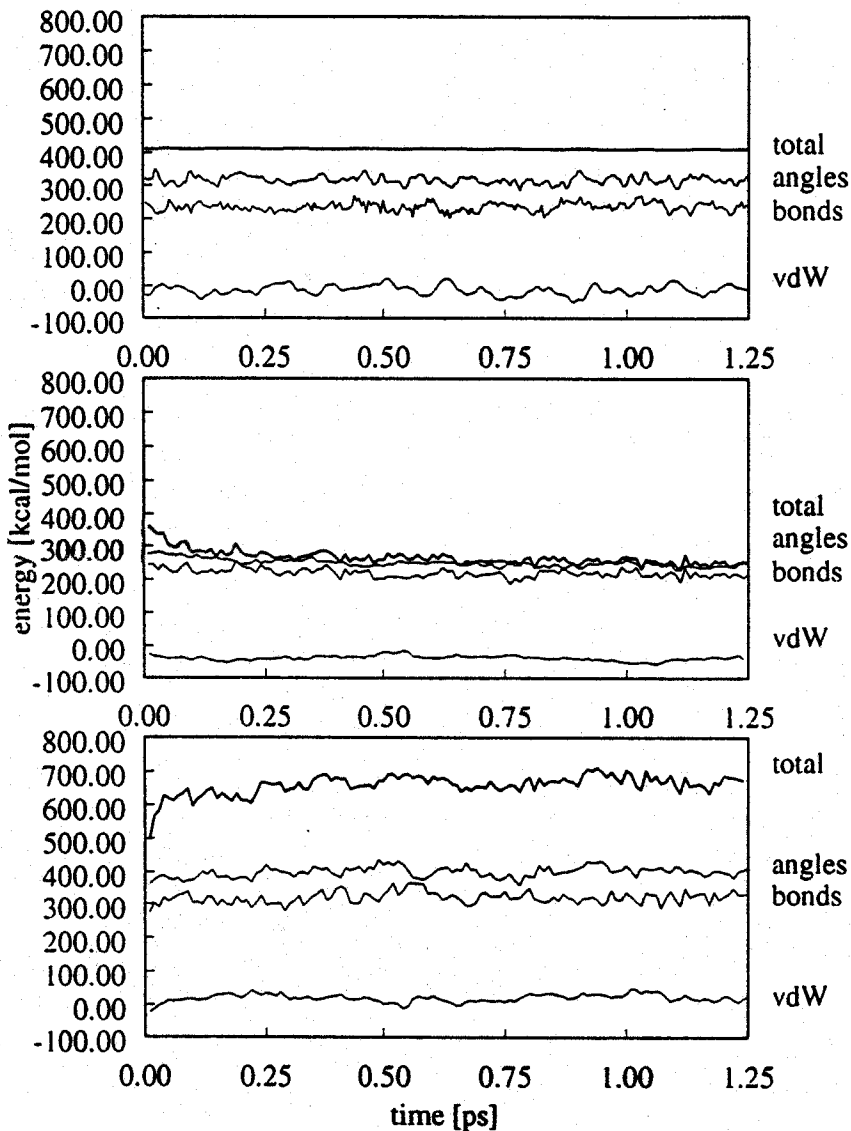


Figure 4: Energy contributions during molecular dynamics simulation using the three different simulation methods discussed in the text. While Uncorrelated Stochastic Dynamics simulation (middle) lowers the potential energy contributions somewhat in comparison with a Newtonian Dynamics simulation (top), Correlated Stochastic Dynamics (bottom) elevates the energy contributions significantly. The latter effect is due to the fact that in Correlated Stochastic Dynamics simulations the system tends to retain its movement even when encountering an energy barrier.

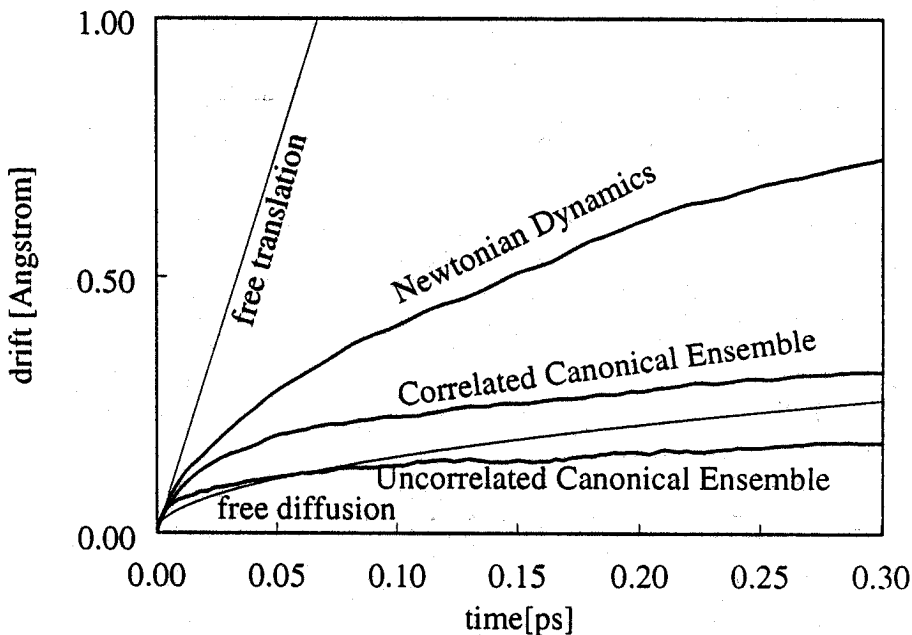


Figure 5: Overall coordinate drifts for the different types of dynamics simulations described in the text. The Newtonian Dynamics simulation clearly shows the most pronounced deviation from the starting configuration. The behavior of the Correlated Stochastic Dynamics (bottom) is halfway between the Newtonian Dynamics (top) and Uncorrelated Stochastic Dynamics (middle) cases. The light curves show the coordinate drift for a free system, i.e. a system of completely unbound atoms.

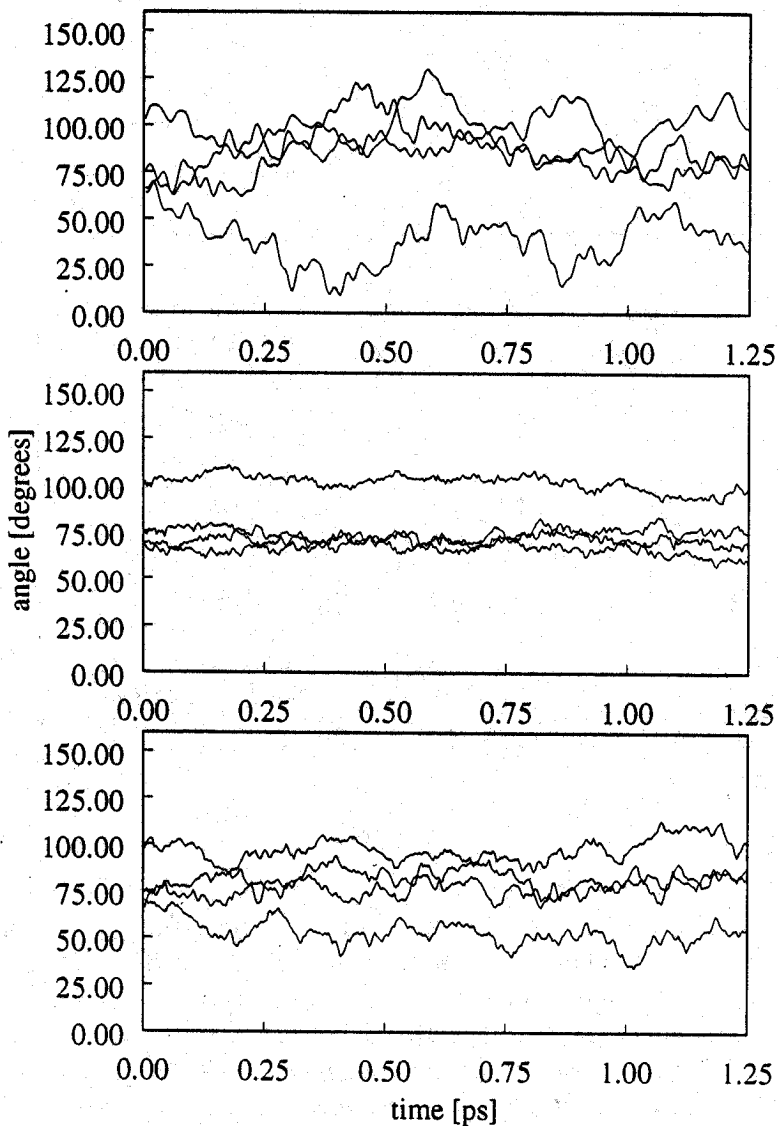


Figure 6: Ring rotation angles for the amino acid side chains TYR 10, TYR 21, PHE 4 and PHE 33 calculated during the three different simulations discussed in the text. The Newtonian Dynamics simulation (top) gives very high mobilities, whereas Uncorrelated Stochastic Dynamics (middle) effectively suppresses ring movement. For Correlated Stochastic Dynamics ring fluctuations are restored to a large degree, in spite of the small correlation time of 5ps assumed in our simulation.