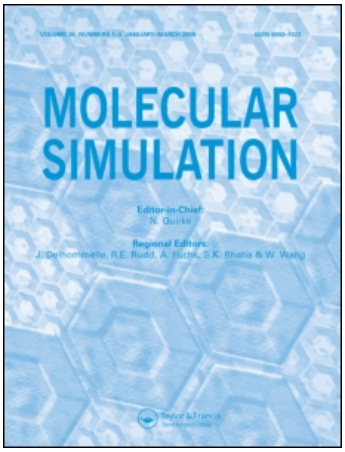

Molecular Simulation
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713644482

Increasing the Time Step and Efficiency of Molecular Dynamics Simulations: Optimal Solutions for Equilibrium Simulations or Structure Refinement of Large Biomolecules

Urs Stocker[a]; Daniel Juchli[a]; Wilfred F. van Gunsteren[a]
[a] Swiss Federal Institute of Technology Zurich, Hönggerberg, Switzerland

Online publication date: 26 October 2010

To cite this Article Stocker, Urs , Juchli, Daniel and van Gunsteren, Wilfred F.(2003) 'Increasing the Time Step and Efficiency of Molecular Dynamics Simulations: Optimal Solutions for Equilibrium Simulations or Structure Refinement of Large Biomolecules', Molecular Simulation, 29: 2, 123 — 138
To link to this Article: DOI: 10.1080/0892702031000065791
URL: http://dx.doi.org/10.1080/0892702031000065791

Molecular Simulation, 2003 Vol. 29 (2), pp. 123–138

Increasing the Time Step and Efficiency of Molecular Dynamics Simulations: Optimal Solutions for Equilibrium Simulations or Structure Refinement of Large Biomolecules

URS STOCKER, DANIEL JUCHLI and WILFRED F. VAN GUNSTEREN*

Swiss Federal Institute of Technology Zurich, CH-8093 ETH Hönggerberg, Switzerland

In biomolecular simulation, the different force field terms determine the motional frequencies found in a particular simulation. By eliminating the high frequencies, longer time steps can be achieved in equilibrium and refinement (non-equilibrium) simulations, sometimes at the expense of decreasing physical correctness of the simulated system. The different force field terms used in molecular simulations can be constrained or softened leading to a maximum time step of 5 fs for a simulation, in which static equilibrium properties are not affected by the changes in the force field parameters introduced, but dynamical quantities are. Using a very smoothed, soft interaction function and increased hydrogen masses, a time step of 14 fs can be reached. However, it should only be used in refinement simulations, in which the unphysical properties of the force field and dynamics are of minor importance.

Keywords: Molecular dynamics simulation; Constraints; Time step; Equations of motion; Biomolecules

INTRODUCTION

Since its introduction in the sixties [1] the computer simulation method of molecular dynamics (MD) has become a well-known method to investigate the equilibrium properties of a variety of molecular systems [2], in particular biomolecular systems [3–5]. In the eighties, MD simulation was introduced as a technique to efficiently search configurational space in the process of structure determination and refinement of proteins based on experimental NMR data [6,7]. This application of MD simulation has since then become a standard tool in structure determination and refinement using NMR and spectroscopic or X-ray diffraction data [8–10]. Although in both applications of MD simulation techniques: (i) study of equilibrium properties and (ii) structure refinement, the same methodology is used, their characteristic features and choice of optimal parameters are rather different:

i. By simulating the real atomic motions over a sufficiently long time period the equilibrium properties of a biomolecular system can be calculated from the simulated trajectory. Newton's or Lagrange's equations of motion are integrated using time steps of the order of 1 fs for as long as possible a period, nowadays in the order of nanoseconds (or 10^6 steps). The goal is to produce a Boltzmann-weighted ensemble of structures, which are, moreover, properly dynamically related or correlated. So, the use of proper equilibrium MD is mandatory in obtaining meaningful results.
ii. In structure refinement, MD simulation is only used as a technique to search conformational space for structures that satisfy best a large number of spatial restraints or conditions, which originate from NMR spectroscopic or X-ray diffraction measurements: nuclear overhauser enhancement (NOE) atom–atom distance bounds, *J*-coupling constants, X-ray diffraction intensities, etc. We write "best", since there will generally not exist a single molecular structure which satisfies all spatial restraints, because the measurement involves averaging over many different structures (in time and space) leading to average restraints. The goal is to produce one or a handful (10, 20, 50, depending on the flavour of the particular research group) of structures that satisfy as good as possible a given set of

*Corresponding author.

ISSN 0892-7022 print/ISSN 1029-0435 online © 2003 Taylor & Francis Ltd
DOI: 10.1080/0892702031000065791

spatial restraints. Here, the use of MD simulation techniques is validated only by the quality of the final structures generated.

Although both applications put different demands on the MD simulation techniques used, they share a need for extensive sampling of conformational space, that is, simulation periods as long as possible at a given accuracy and computational cost. In other words, increasing the time step and efficiency of MD simulations is of interest to both applications.

The four basic limitations of classical MD simulation in describing a biomolecular system are the following.

1. The use of classical–dynamical equations of motion to simulate motion along the chosen degree of freedom, thereby neglecting quantum effects.
2. The finite size of the system or number of degrees of freedom that can be simulated, currently typically 10^4–10^5 atoms or degrees of freedom.
3. The finite accuracy of the energy function or force field representing the energy of the system as function of the degrees of freedom along which the motion is simulated.
4. The finite length of or time period covered by a simulation, which limits the sampling and time scale of processes that can be simulated, currently typically 10^8 s.

In order to lengthen the time period covered by an MD simulation a variety of approaches has been tried in the past.

1. Reduction of the number of degrees of freedom to be sampled to the minimum that is essential to the particular properties of the system one is interested in. Examples of removing of, hopefully, non-essential degrees of freedom are the following.
    i. In structure refinement of proteins, solvent degrees of freedom are generally omitted [11,12], although it is experimentally known that protein structure is sensitive to solvent composition.
    ii. Aliphatic hydrogen-atom degrees of freedom can be removed by treating CH_n-groups as united (effective) atoms, an approximation which reduces the number of atom–atom interactions in proteins by a factor of 4, in membranes by a factor of up to 9, without significant loss of accuracy [13].
    iii. If hydrogen atoms are to be kept as interaction sites, but their positions need not be sampled, they can be treated as so-called virtual atoms or interaction sites, which have zero mass (no inertia) and a position that is completely determined by (generally neighbour) non-virtual atoms (which have non-zero mass) [14–17].
    iv. Bond-length or bond-angle degrees of freedom can be eliminated from the system by treating them as constrained [11,12,18,19].
2. Simplification of the energy function or force field leading to a reduction in the number and complexity of atom–atom interactions to be evaluated, and so to a reduced computational cost. Since the calculation of non-bonded interactions dominates the cost of a molecular simulation, a considerable speed-up may be obtained at the expense of loss of force field accuracy. Examples are the following.
    i. Neglect of (long-range) electrostatic interactions [11,12].
    ii. Use of a very short (smaller than 0.8 nm) cut-off radius for non-bonded interactions or neglect of attractive van der Waals interactions [11,20].
3. Use of multiple-time-step (MTS) algorithms to integrate the equations of motion. When integrating the forces arising from different terms in a force field that are characterized by different relaxation times, it can be efficient to use different integration time steps Δt for the different force field terms. If for the more costly force field terms a longer time step can be used, the efficiency of the integration of the equations of motion is enhanced. MTS algorithms have been used for integrating bond-stretching and bond-angle-bending forces separately from the remaining forces [21–23], and for integrating long-range Coulomb and van der Waals forces separately from the remaining ones in the so-called twin-range method [4].
4. Reduction of the frequencies of the fastest motions in the system will allow for a longer time step Δt to be used: Algorithms to integrate equations of motion forward in time are based on the approximation that the forces or their derivatives are constant during the integration time step Δt, which approximation is of course more exact the smaller Δt is. This approximation is exact in the special case that the force is constant or in other words, if the derivative of the force or the second derivative of the energy function V, viz., the curvature of V equals zero. So the length of the integration time step Δt is limited by the size of the local curvature of V or in physical language, by

the highest-frequency (ν_{max}) motions in the molecular system:

$$\Delta t \ll \nu_{max}^{-1} = \tau \quad (1)$$

For example, for the one-dimensional classical harmonic oscillator with potential energy function

$$V(x) = \frac{1}{2}Kx^2 \quad (2)$$

and mass m, one has

$$\nu = \frac{(K/m)^{1/2}}{2\pi} \quad (3)$$

and the curvature equals the force constant K. The larger the curvature K of $V(x)$, the higher the frequency ν of the motion and the smaller time step Δt must be used. Expressions (1)–(3) show different ways the time step in an MD simulation can be lengthened.

a. Redistribution of the total mass of a molecular system over the atoms such that the range of motional frequencies is reduced. When the inertia of the fastest motions is increased, ν_{max} is reduced thereby allowing a longer time step [24]. Of course, dynamical properties are modified when modifying atomic masses non-uniformly, but thermodynamic properties remain unchanged. For flexible molecules with internal spatial constraints such as bond lengths or angles this is not true due to metric-tensor effects [25]. By making the atomic masses more equal, that is by increasing the masses of hydrogens and reducing the masses of non-hydrogen atoms, a lengthening of the time step by 50% can be obtained for proteins [26] and for water [27].
b. Softening or smoothing of the high-frequency (most strongly curved) interaction terms in the potential energy function V will also allow the use of longer integration time steps Δt. This technique is often used in the earlier stages of structure refinement in order to enlarge the radius of convergence of the restraining simulation and in order to avoid a reduction of Δt when simulating at high temperature in simulated annealing refinement.
c. High frequency motions can also be eliminated from a molecular system by the application of atom–atom distance constraints. Two types of distance constraints can be distinguished.

    1. Holonomic (time-independent) constraints can be implemented in two ways: (i) by formulating Lagrange equations of motion in generalized (e.g. torsional) coordinates [28,29], or (ii) by formulating these equations in Cartesian coordinates (i.e. using Newton's equations) and then using Lagrange multipliers to satisfy the constraints at each MD time step [30–32].

    2. Soft or adiabatic (variable length) distance constraints can only be implemented using the latter technique. The difference with respect to hard or holonomic constraints is that the length of a constrained distance is not a constant throughout the simulation, but varies per integration time step without involving kinetic energy. Reich [33] applies a smoothing operator to the Hamiltonian of a non-constrained system, which suppresses high-frequency oscillations and formulates its effect in terms of a soft-constrained Hamiltonian. An alternative is to adjust the length of a constrained distance adiabatically at each time point such that no strain is built up along the constrained degree of freedom.

How the various techniques to reduce the frequency of the fastest motions can be applied and combined most effectively to lengthen the MD time step depends on the hierarchy of frequencies originating in different types of interatomic interactions governing the degrees of freedom of a particular molecular system. It also depends on the goal of a simulation. In standard structure refinement [11,12], solvent degrees of freedom are omitted, and only solute torsional-angle degrees of freedom are sampled using a simplified and smoothed energy function, with the goal of obtaining a model structure of a protein. The severe approximations allow for fast refinement at the cost of loss of physical content, which is no problem as long as the quality of the obtained model structure can be independently assessed. On the other hand, in equilibrium simulations to study the dynamics of a protein only a few approximations can be accommodated without significant loss of accuracy, e.g. elimination of the bond-length degrees of freedom and aliphatic hydrogen degrees of freedom.

It has been suggested, that by integrating Lagrange equations of motions for torsional-angle degrees of freedom using an implementation in generalized coordinates, conformational space is more efficiently sampled than by using Cartesian coordinates [11,12]. However, the coordinate system used will not determine the efficiency of sampling, the choice of degree of freedom along which the sampling is done may, of course, affect the efficiency. Here, coordinate system and degree of freedom are confused. In fact, one can integrate Lagrange equations of motion for

torsional degrees of freedom using Cartesian coordinates and Lagrange multipliers as efficient as when integrating Lagrange equations of motion formulated in torsional, generalized coordinates.

In order to enable a rational choice of approximations aimed at speeding up biomolecular simulations, we review the topic and list the various aspects and choices involved in "Methods". In "Results" and "Discussion" we illustrate the findings with examples: a simple model chain of 100 CH_n atoms and a small protein, bovine pancreatic trypsin inhibitor (BPTI), both using simulations *in vacuo*. "Discussion" contains a number of practical conclusions.

METHODS

In this section, we briefly review and list a number of aspects and concepts of molecular simulation that are of importance when analysing the computational efficiency and accuracy of equilibrium or structure refinement simulation.

Simulation and Sampling Methods

Since biomolecular conformational space is too large to be exhaustively sampled, one generally has to use in biomolecular modeling or structure refinement heuristic methods for sampling and searching for low-energy conformers. An overview of types of methods to search and sample conformational space can be found in Ref. [34]. Only a subset of the great variety of methods has been tried in structure refinement based on spectroscopic or diffraction data. The most widely used sampling methods are the following.

1. Non-Boltzmann sampling, such as
    a. repeated distance geometry calculations;
    b. structural database sampling;
    c. random or gradient-driven variation of torsional angles sampling.
2. Boltzmann sampling, such as
    a. conventional or configuration-bias Monte Carlo (MC) simulation [35]
    b. molecular dynamics (MD) simulation [2]
    c. stochastic dynamics (SD) simulation [36]

The efficiency of the sampling is generally restricted by the general nature of the energy surface or the function V. The occurrence of high-energy barriers between local minima may inhibit proper sampling. Therefore, techniques have been developed to enhance the sampling power of the methods [34]. Only methods of type 1.c and of type 2 involve stepping through conformational space. In gradient-driven step-methods force-biased MC, MD, or SD simulations, the maximum possible step size is determined by the local curvature of the energy function, and in the dynamics simulation methods by the mass distribution along the degrees of freedom that are sampled.

Hierarchy of Motional Frequencies in Biomolecular Systems

In biomolecular systems a hierarchy of motional frequencies originating in different types of interatomic interactions can be distinguished [10,27,37]. In order of decreasing frequency or increasing smoothness of the corresponding atom–atom interaction term in the biomolecular force field we have (Table II):

i. Bond-stretching vibrations with an approximate oscillation or relaxation time $\tau_{\mathrm{I}} \approx 10$ fs for bonds involving a hydrogen atom and $\tau_{\mathrm{I}} \approx 20$ fs for bonds involving only carbon or heavier atoms.
ii. Bond-angle bending vibrations with $\tau_{\mathrm{II}} \approx 20$ fs for bond angles involving hydrogen atoms and $\tau_{\mathrm{II}} > \approx 40$ fs for bond angles involving only carbon or heavier atoms.
iii. Improper dihedral angle vibrations due to force field terms used to impose the proper chirality at chiral CH_1 united atoms or to impose planarity on ring structures with conjugated double bonds with $\tau_{\mathrm{III}} > \approx 30$ fs.
iv. Torsional-angle vibrations around double bonds (e.g. peptide bonds) with $\tau_{\mathrm{IV}} \approx 30$ fs.
v. Water or solvent librational vibrations with $\tau_{\mathrm{V}} \approx 30$ fs.
vi. Torsional-angle vibrations around single bonds with $\tau_{\mathrm{VI}} \approx 40$ fs for torsional angles involving a hydrogen atom and $\tau_{\mathrm{VI}} > \approx 80$ fs for torsional angles involving only carbon or heavier atoms.
vii. Motions dominated by (not softened) van der Waals contacts and short-range (e.g. hydrogen bonding) Coulomb interactions with $\tau_{\mathrm{VII}} \approx 150$ fs.
viii. Motions dominated by long-range Coulomb (ionic, dielectric) interactions with relaxation time $\tau_{\mathrm{VIII}} > \approx 2000$ fs.

It is this hierarchy that is exploited to enhance the efficiency of a simulation through the use of longer time steps Δt [38].

Choice of Molecular Model and Force Field

In its most simple form a biomolecular force field contains the following terms, where we have used

the functional form of the GROMOS96 force field [17,39] as example.

1. Bond-stretching terms

$$V^{\text{bond}}(\vec{r}(t)) = \frac{1}{4}\sum_{n=1}^{N_b} K_{b_n}\left[b_n^2(t) - b_{0_n}^2\right]^2 \qquad (4)$$

in which b_{0_n} is the ideal bond length and $b_n(t)$ is the actual (at time point (t)) bond length of the bond with sequence number n in the list of N_b bonds in the molecular system, and K_{b_n} determines the strength of the n-th bond, which determines the curvature or smoothness of term (4).

2. Bond-angle bending terms

$$V^{\text{angle}}(\vec{r}(t)) = \frac{1}{2}\sum_{n=1}^{N_\theta} K_{\theta_n}[\cos(\theta(t)) - \cos(\theta_{0_n})]^2 \qquad (5)$$

in which θ_{0_n} is the ideal bond angle and $\theta_n(t)$ the actual bond angle value of the n-th bond angle in the list of N_θ bond angles, and K_{θ_n} is the force constant determining the curvature of term (5).

3. Improper dihedral angle terms

$$V^{\text{har}}(\vec{r}(t)) = \frac{1}{2}\sum_{n=1}^{N_\xi} K_{\xi_n}[\xi_n(t) - \xi_{0_n}]^2 \qquad (6)$$

in which ξ_{0_n} is the ideal improper dihedral, $\xi(t)$ the actual improper dihedral angle value, and K_{ξ_n} is the force constant of the n-th improper dihedral.

4. Torsional (or proper) dihedral angle terms

$$V^{\text{trig}}(\vec{r}(t)) = \sum_{n=1}^{N_\varphi} K_{\varphi_n}[1 + \cos(\delta_n)\cos(m_n\varphi_n(t))] \qquad (7)$$

in which δ_n is the phase shift (0 or π), m_n the multiplicity, $\varphi_n(t)$ the actual value and K_{φ_n} is the force constant of the n-th torsional angle.

5. Non-bonded (Lennard–Jones, Coulomb) interaction terms

$$V^{\text{LJ}}(\vec{r}(t)) = \sum_{\text{non-bondedpairs}(i,j)} \frac{1}{\alpha_{\text{LJ}}(i,j)C_{126}(i,j) + (r_{ij}(t))^6} \times \left[\frac{C_{12}(i,j)}{\alpha_{\text{LJ}}(i,j)C_{126}(i,j) + (r_{ij}(t))^6} - C_6(i,j)\right] \qquad (8)$$

and

$$\mathrm{t}V^{\text{CRF}}(\vec{r}(t)) = \sum_{\text{non-bondedpairs}(i,j)} \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_1} \times \left[\frac{1}{[\alpha_{\text{C}}(i,j) + (r_{ij}(t))^2]^{1/2}} - \frac{\frac{1}{2}C_{rf}(r_{ij}(t))^2}{\left[\alpha_{\text{C}}(i,j) + R_{rf}^2\right]^{3/2}} - \frac{1 - \frac{1}{2}C_{rf}}{R_{rf}}\right] \qquad (9)$$

in which r_{ij} is the actual distance between atoms i and j. $\alpha_{\text{LJ}}(i,j)$ and $\alpha_v(i,j)$ are so-called soft-core parameters for the Lennard–Jones and charge interactions [40], respectively, which can be used to convert the hard-core singularity at $r_{ij} = 0$ for $\alpha_{\text{LJ}}(i,j) = \alpha_{\text{C}}(i,j) = 0$ into a smooth soft core for α_{LJ}, $\alpha_{\text{C}} > 0$. The r^{-6} and r^{-12} Lennard–Jones parameters for atom pair (i,j) are $C_6(i,j)$ and $C_{12}(i,j)$, respectively and

$$C_{126}(i,j) = \begin{cases} \frac{C_{12}(i,j)}{C_6(i,j)} & \text{if } C_6(i,j) \neq 0 \\ 0 & \text{if } C_6(i,j) = 0 \end{cases} \qquad (10)$$

The partial charge of atom i is denoted by q_i and the constant

$$C_{rf} = \frac{(2\varepsilon_1 - 2\varepsilon_2)(1 + \kappa R_{rf}) - \varepsilon_2(\kappa R_{rf})^2}{(\varepsilon_1 + 2\varepsilon_2)(1 + \kappa R_{rf}) + \varepsilon_2(\kappa R_{rf})^2} \qquad (11)$$

determines the Poisson–Boltzmann reaction field [41] contribution due to a medium with relative dielectric permittivity ε_2 and inverse Debye screening length κ outside the cut-off sphere with radius R_{rf}, and ε_1 is the relative permittivity inside this sphere. The dielectric permittivity of vacuum is ϵ_0.

The smoothness of the energy function consisting of terms (4)–(9) can be increased by decreasing the force constants K_{b_n}, K_{θ_n}, K_{ξ_n}, and K_{φ_n} in terms (4)–(7) and by increasing the values of the soft core parameters α_{LJ} and α_{C} in terms (8) and (9).

Choice of Degrees of Freedom to be Simulated or Sampled

The hierarchy of motional frequencies in a molecular system (as determined physically or artificially by the force field terms used) will determine which

degrees of freedom should be omitted in order to allow for a longer time step Δt.

1. Omission of aliphatic hydrogen degrees of freedom has two advantages. The number of non-bonded atom–atom interactions in the summations of Eqs. (8) and (9) is more than halved for proteins and lipids. Secondly, the high-frequency ($\tau_l \approx 10$ fs) hydrogen motions need not be integrated.
2. If the aliphatic hydrogen atoms are kept as virtual interaction sites (with zero mass) only the second advantage is retained.
3. This is also true, when the bonds or in addition the bond angles involving hydrogen atoms are constrained in a simulation.
4. Omitting water or solvent degrees of freedom essentially yields only the first advantage since their typical motional frequencies differ not much from those in proteins. However, since solute–solvent and solvent–solvent interactions easily account for 80–99% of the computational effort in explicit solvent simulations, omitting solvent degrees of freedom yields by far the largest gain in computational efficiency.

When omitting degrees of freedom, their (mean) effect on the remaining explicitly simulated degrees of freedom should be as much as possible incorporated into the energy function for these degrees of freedom. Mean solvation force terms have been reviewed in Ref. [42].

Equations of Motion

The classical equations of motion have been given by Lagrange in a most general form

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L(q,\dot{q})}{\partial \dot{q}_i}\right) = \frac{\partial L(q,\dot{q})}{\partial q_i} \quad i = 1, 2, \ldots, N_{\mathrm{df}} \tag{12}$$

where q_i denote the generalized coordinates, $\dot{q}_i$ their time derivatives and the Lagrangean $L(q,\dot{q})$ is the kinetic energy $K(\dot{q})$ minus the potential energy $V(q)$ of the system which contains N_{df} degrees of freedom. When using Cartesian coordinates $q \equiv x$, one has $K(\dot{x}) = (1/2)m\dot{x}^2$ and Eq. (12) reduce to Newton's equations of motion (for N_{df} degrees of freedom)

$$m_i \frac{\mathrm{d}^2 x_i}{\mathrm{d}t^2} = \frac{-\partial V(x_1, x_2, \ldots, x_{N_{\mathrm{df}}})}{\partial x_i},$$

$$i = 1, 2, \ldots, N_{\mathrm{df}} \tag{13}$$

When considering branched polymers, the choice of internal coordinates, bond lengths, bond angles, and torsional angles seems to be natural. However, the equations of classical dynamics (12) expressed in the N_{df} internal, generalized coordinates $q_i \equiv \theta_i$

$$\sum_{j=1}^{N_{\mathrm{df}}} a_{ij} \frac{\mathrm{d}^2\theta_j}{\mathrm{d}t^2} = \frac{-\partial V(\theta_1, \theta_2, \ldots, \theta_{N_{\mathrm{df}}})}{\partial \theta_i} - \sum_{j=1}^{N_{\mathrm{df}}} b_{ij}\left(\frac{\mathrm{d}\theta_j}{\mathrm{d}t}\right)^2 - \sum_{j=1}^{N_{\mathrm{df}}}\sum_{k=1}^{N_{\mathrm{df}}} c_{ijk}\left(\frac{\mathrm{d}\theta_j}{\mathrm{d}t}\right)\left(\frac{\mathrm{d}\theta_k}{\mathrm{d}t}\right)$$

$$1, 2, \ldots, N_{\mathrm{df}} \tag{14}$$

are considerably more complex than when expressed in Cartesian coordinates. They contain two additional summations over the number of degrees of freedom and two additional quadratic (i.e. non-linear) terms in the generalized velocities. Equations (14) have been presented in different forms [19,28,29,43–48], and the coefficients a_{ij}, b_{ij}, and c_{ij} depend on the atomic masses and the molecular topology of the polymer considered.

Choice of Coordinate System

Since the Cartesian (Newtonian) equations of motion (13) do not involve the explicit coupling of the equations through the summation over the index j in Eq. (14) and lack the two non-linear terms depending on the generalized velocities in Eq. (14), Cartesian equations of motion have been the method of choice in MD and SD simulations of biomolecular systems for efficiency reasons. Even in the presence of spatial constraints, such as bond-length and bond-angle constraints, it is possible to use a Cartesian coordinate system, while imposing the constraint conditions at every time step through the Lagrange multiplier technique [49]. When only bond lengths are constrained, it takes for a protein only a few percent of the total simulation effort to determine the Lagrange multipliers [50]. This is done by iteratively solving the constraint equations [30,32]. When bond-angle degrees of freedom are to be constrained in addition, the simple SHAKE algorithm [30] becomes inefficient [18]. However, using techniques proposed in Ref. [31] the convergence of solving the constraint equations can be improved considerably. In this case it takes for a protein at most 10–20% of the total simulation effort to determine the Lagrange multipliers.

Simulation of a protein through the equations of motion in generalized, torsional coordinates (14) requires a larger computational effort, since at each time step a set of N_{df} non-linear equations is to be solved. One iteration to this end may take as much computational effort as the calculation of all forces and energies, thereby roughly doubling the overall computational effort [11]. Secondly, it is practically impossible in the framework of generalized,

torsional coordinates to treat chain closure of a chain of atom–atom distance constraints while maintaining flexibility within the closed chain. As a consequence, disulfide bridges or other polypeptide chain closures are treated as unconstrained degrees of freedom in simulations based on Eq. (14). This implies that their high-frequency motions are still present in the system, unless they are removed using other techniques (e.g. potential-energy smoothing).

In summary, the use of equations of motion in non-Cartesian coordinates should be avoided if computational efficiency is to be optimized.

Choice of Numerical Algorithm to Integrate Newton's Equations of Motion

A great variety of time integration algorithms has been tried over the years in MD simulation [2,51]. Since Newton's equations of motion (13) are time-invariant and do not contain terms dependent on the atomic velocities, simple algorithms that are time-invariant are most efficient to integrate them [51,52]. The leap-frog algorithm [53] and the equivalent Verlet algorithm [54] are most efficient.

Use of Multiple-time-step Algorithms

One way to exploit the hierarchy of motional frequencies in biomolecular systems is to use multiple-time-step (MTS) algorithms, in which different time steps Δt_{I}, $\Delta t_{\mathrm{II-VII}}$ or Δt_{VIII} are used, each satisfying condition (1) with respect to the oscillation or relaxation times τ_{I}, $\tau_{\mathrm{II-VII}}$, or τ_{VIII} of the various atom–atom interaction terms discussed in "Hierarchy of motional frequencies in biomolecular systems" section, when integrating the contributions of the different forces. Since evaluating the non-bonded force terms (8) and (9) is by far the most costly, it is efficient to integrate high-frequency bond stretching and bond-angle bending forces ($\tau_{\mathrm{I-II}}$) [22,23] and the very low-frequency non-bonded forces (τ_{VIII}) [4] with time steps shorter and longer, respectively, than the standard time step by which the remaining forces are integrated ($\tau_{\mathrm{II-VII}}$).

Redistribution of Atomic Mass Along Degrees of Freedom

Another way to exploit the hierarchy of motional frequencies is to change atomic masses such that the fastest motions are slowed down and the slowest motions are sped up. A simple approach is to increase the masses of hydrogens to those of carbons or nitrogens, which allows the MD time step to be lengthened, however, at the cost of distortion of the dynamics of the system. Thermodynamic properties, however, are not influenced by mass changes. This technique is also used in structure refinement using MD along torsional angle degrees of freedom [11].

We note that a scaling of the total mass of a molecular system makes not much sense: At constant temperature, the scaling of all atomic masses with a factor λ (or scaling the unit of mass by a factor of λ^{-1}) is equivalent to scaling the time with a factor $\lambda^{-1/2}$. The increase in size of the time step due to the mass increase will be precisely offset by an equally big reduction of the time unit.

Softening or Smoothing the Highest-frequency Interaction Terms

The oscillation or relaxation time of the highest-frequency motions can be reduced by decreasing the force constant of the corresponding force field terms. For example, a reduction of the strength K_{b_n} of the bond stretching term (4) by a factor 10 will allow a three times longer time step, as indicated by relation (3). The price to be paid is an increase of the bond-length fluctuations by a factor of 3. Similar conditions apply to the bond-angle term (5) and the improper dihedral angle term (6).

A peculiar situation may occur when using very soft bond-angle forces with the torsional energy term (7) switched on. If the bond-angle i–j–k defined by atoms i, j, and k increases towards 180° and beyond, the torsional dihedral angle i–j–k–l will make a step of 180°, which may induce a large sudden change in torsional-angle energy. This effect could be avoided by introducing a coupling between bond-angle and torsional angle forces.

Use of Hard, Soft or Adiabatic Atom–atom Distance Constraints

High-frequency motions can be eliminated from a molecular system by constraining the corresponding degrees of freedom. This can be implemented either by formulating the equations of motion in generalized coordinates or by sticking to Cartesian coordinates and using the Lagrange multiplier technique to impose the constraints. When using soft or adiabatic (variable length) constraints the latter method is most practical.

Consequences of the use of the Various Techniques and Approximations

When applying constraints, so-called metric tensor correction terms may have to be added to the interaction function V, depending on the type of constraints used [25,55]. Moreover, a physical force field calibrated for use in unconstrained simulations may have to be recalibrated for use in conjunction with hard constraints, which will rigidify

the molecules [18], or with soft or adiabatic constraints, which will mollify the molecules.

Another relevant aspect is the quantum-mechanical nature of particular intramolecular vibrations. Bond-stretching vibrations have frequencies ω or wave numbers in the range $2000-4000\,\mathrm{cm}^{-1}$. At room temperature, $k_\mathrm{B}T$ corresponds to a wave number of $200\,\mathrm{cm}^{-1}$. Thus we have (k_B is Boltzmann's constant and h is Planck's constant)

$$\frac{h}{2\pi}\omega \gg k_\mathrm{B}T \tag{15}$$

which implies that the bond-stretching vibrations are essentially of quantum-mechanical nature and that only the ground state will be populated. Treating the bonds as hard constraints is most likely to be a more correct approximation of the quantum dynamics than adiabatic dynamics or classical (harmonic oscillator) dynamics, with its different energy distribution, would be. On the other hand, zero-point energy is neglected.

Another unpleasant aspect of fully flexible molecular models is the presence of weakly coupled modes of different frequencies, which makes the energy distribution over these modes in a molecular simulation a slow process. This is a technical problem, which can be avoided by coupling the different degrees of freedom (high-frequency ones versus the rest) separately to a heat bath [17]. This problem is likely to occur when MTS algorithms are used for higher-frequency motions.

An obvious, non-desirable effect of the softening or smoothing of the highest-frequency interaction terms in V is an enlargement of the structural fluctuations of the molecules along the degrees of freedom for which the interaction has been smoothed. This may influence thermodynamic properties that depend on fluctuations.

The unpleasant consequences of using equations of motion in non-Cartesian coordinates have been discussed in "Choice of coordinate system" section.

Relative Merits of the Various Time-saving Techniques and Approximations

We now reconsider the hierarchy of motional frequencies and the corresponding force field terms given before, and discuss the relative merits of applying the different time-saving techniques to lengthen the integration time step Δt to a particular motion or force, field term.

i. Bond-stretching motion ($\tau_\mathrm{I} = 10-20\,\mathrm{fs}$). The most appropriate treatment of the bond-stretching degree of freedom is the use of hard constraints. It is a good approximation of their quantum-mechanical nature, it avoids the energy redistribution problem, and it can be very simply carried out using Cartesian coordinates (Newton's equations of motion) with the SHAKE method to determine the Lagrange multipliers [30]. Metric tensor effects are negligible [55] and force-field corrections are not necessary: the dynamics of the molecular system is not affected by the bond-length constraints [18]. Use of any of the time-saving techniques (constraining of the degree of freedom, mass redistribution along the degree of freedom, smoothing of the force field term, MTS integration) may reduce the computational effort by up to a factor of 4 through a lengthening of the time step.

ii. Bond-angle bending motion involving hydrogen atoms ($\tau_\mathrm{II} \approx 20\,\mathrm{fs}$). Constraining the bond-angle degrees of freedom involving hydrogen atoms would allow for a small gain in time step. Metric tensor effects are probably small, and the heavy atom dynamics probably not much influenced. Alternatives are to use larger hydrogen masses or to reduce the force constants for the bond-angle terms involving hydrogens slightly.

iii–vi. Bond-angle bending motion of non-hydrogen atoms, torsional motion around double bonds, improper dihedral vibrations, and water librational motion ($\tau_\mathrm{II-VI} \approx 30\,\mathrm{fs}$). In simulations of biomolecules including explicit water molecules, little gain in computational efficiency can be obtained by use of one of the time-saving techniques due to the presence of the high-frequency water molecule librations governed by the non-bonded (van der Waals, Coulomb) interactions. These limit the time step to about $\Delta t = 2\,\mathrm{fs}$. Using heavier hydrogens would allow for a somewhat larger $\Delta t = 3\,\mathrm{fs}$, at the cost of distorting dynamical properties. When simulating a macromolecule *in vacuo*, the bond-angle and stiff (double bond character) torsional-angle degrees of freedom could be treated using one of the time-saving techniques: constraining, mass redistribution, or smoothing of the force field terms. This may reduce the computational effort by up to a factor of 4. Typically, the time step is limited to about $\Delta t = 8\,\mathrm{fs}$ when using physically realistic force fields. By combining constraints, hydrogen mass scaling, and softening of the various interaction terms a time step $\Delta t = 14\,\mathrm{fs}$ can be reached. The use of bond angle constraints or so-called torsion-angle dynamics [11,12,19] introduces two

problems: (i) metric-tensor effects are not small, so should not be ignored [55], and (ii) the dynamics of the macromolecule is severely altered due to the rigidification of the molecular model, structural fluctuations and torsional-angle transitions are quenched [18]. To restore the proper physical behaviour of the molecular system a metric-tensor term should be added to the energy function V, and the force field should be recalibrated for use with bond-angle constraints.

vii. Motions dominated by van der Waals contacts, single-bond torsional interactions, and short-range Coulomb interactions, ($\tau_{\mathrm{VII}} \approx 150$ fs). These interactions dominate the essential degrees of freedom of molecular systems and should therefore not be approximated using time saving techniques. They limit the time step using physically realistic force fields to $\Delta t = 8$ fs for macromolecules and to $\Delta t = 2$ fs when water is used as solvent. As before, softening of torsional and non-bonded interaction terms allows for larger Δt, but at a loss of physical accuracy of the molecular model.

viii. Motions dominated by long-range Coulomb interactions ($\tau_{\mathrm{VIII}} > \approx 2000$ fs). Long-range electrostatic interactions can be computationally very demanding due to their dependence on a large number of atoms. Since these interactions are only changing slowly during a simulation, the use of multiple-time-step algorithms when integrating these forces can reduce the computational effort by up to a factor of 10 [4,17]. On the other hand, application of so-called particle–particle–particle–mesh (P3M) methods to evaluate electrostatic interactions [53] may reduce the computational effort by a factor of 100 over conventional Ewald summation techniques [56].

Testing the Precision and Efficiency of the Various Simulation Techniques

To compare the various techniques aimed at lengthening the time step Δt in MD simulations, one should compare MD simulations (i.e. trajectories) that differ only in one aspect, all other parameters, force field terms, set-up procedures being the same. Considering the variety of techniques discussed, a comprehensive comparison would require too much effort. Therefore, we have limited ourselves to a comparison of the following properties.

In equilibrium MD simulations without coupling to temperature or pressure baths, the total energy

$$E_{\mathrm{tot}}(t) = E_{\mathrm{kin}}(t) + E_{\mathrm{pot}}(t) \tag{16}$$

should be conserved over the trajectory. This implies that the root-mean-square fluctuation of the total energy

$$\Delta E_{\mathrm{tot}} \equiv \langle [E_{\mathrm{tot}} - \langle E_{\mathrm{tot}} \rangle]^2 \rangle^{1/2} \tag{17}$$

and the drift in E_{tot} should be ideally zero. Here a trajectory average is denoted by $\langle\ \rangle$. Due to numerical and integration errors. ΔE_{tot} will be non-zero and could be used as an indication of the precision with which the equations of motion are integrated forward in time. One would expect that for longer integration time steps Δt larger values for ΔE_{tot} will be obtained. For very small time steps Δt the value of ΔE_{tot} should become independent of Δt, since in that limit, numerical errors are expected to dominate integration errors [50]. It makes little sense to calculate the relative fluctuation of the total energy $\Delta E_{\mathrm{tot}}/\langle E_{\mathrm{tot}} \rangle$ as was done [11], since the zero-point of the potential energy E_{pot} and therefore of the total energy E_{tot} can be chosen at will. This means that the relative fluctuation $\Delta E_{\mathrm{tot}}/\langle E_{\mathrm{tot}} \rangle$ can be made arbitrarily large or small depending on the choice of zero on the potential energy scale. A more useful quantity to measure the precision of the time integration is the ratio of the fluctuations in the total energy to those in the kinetic energy

$$\Delta E_{\mathrm{tot}}/\Delta E_{\mathrm{kin}}, \tag{18}$$

since the latter should be independent of the time step and of the zero-point of the potential energy. This ratio measures how well the total energy is conserved given the kinetic and potential energy fluctuations characteristic for the molecular system at the chosen temperature [50]. For a more detailed analysis of energy conservation in MD simulations we refer to Ref. [52].

In non-equilibrium MD simulations as used in structure refinement, conservation of energy or correct dynamics is of no importance. In this case the efficiency of an algorithm can be measured by determining the computational effort required to obtain a correctly folded structure starting from an arbitrary or extended structure and using spatial restraints derived from NMR or X-ray data to guide the refinement.

Molecular Dynamics Simulations

Equilibrium MD simulations were carried out for two test systems in order to determine the degree of energy conservation as a function of MD time step size Δt and the maximum possible Δt value in dependence of the setting of the following force field

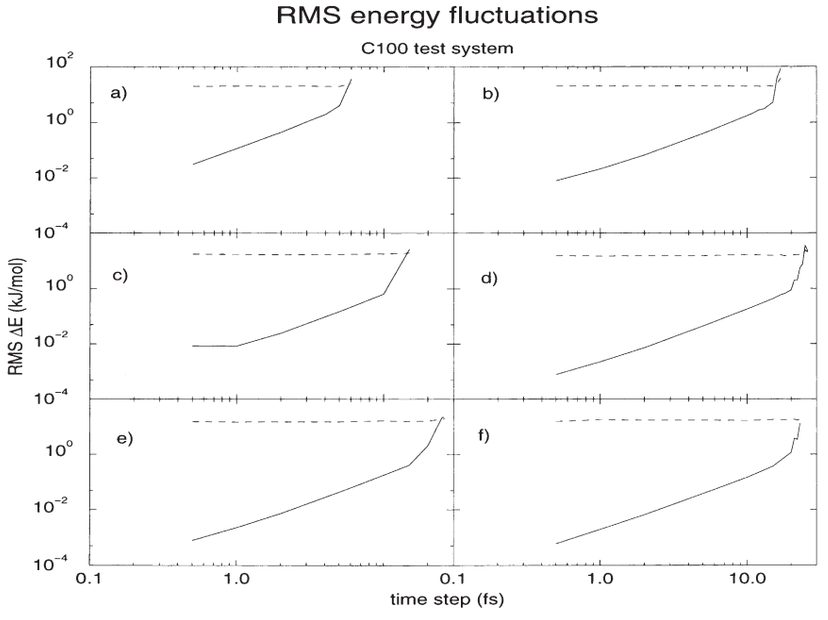

FIGURE 1 Root-mean-square fluctuations of total (E_{tot}, solid line) and kinetic (E_{kin}, dashed line) energy for the C_{100} test system for different parameter setting as function of the time step at 300 K: (a) fully flexible model, (b) bonds constrained, (c) bonds constrained, bond-angle force constants K_{0_n} scaled by a factor 0.3, (d) bonds constrained, bond-angle force constants K_{0_n} scaled by a factor 0.2, no torsional-angle forces, (e) bonds constrained, bond-angle forces constants K_{0_n} scaled by a factor 0.1, no torsional-angle forces, (f) bonds constrained, no bond-angle forces, no torsional-angle forces.

and simulation parameters:

a. application of bond-length constraints (bond-constraints),
b. hydrogen mass m_H,
c. force constant of K_{b_n} of bond-stretching terms,
d. force constant K_{θ_n} of bond-angle bending terms,
e. force constant K_{ξ_n} of improper-dihedral angle terms,
f. force constant K_{φ_n} of proper torsional-angle terms,
g. van der Waals and Coulomb soft-core parameters α_{LJ} and α_C,
h. atom partial charges.

If not explicitly stated otherwise, the force field parameter values are those of the GROMOS96 [17] force field, and the hydrogen masses are $m_{II} = 1.008$ u. The first test system consisted of a linear C_{100} chain built of 98 united CH_2 atoms and 2 united CH_3 atoms as end groups. In the united atom approach, aliphatic hydrogen atoms are incorporated into the carbon atoms and not treated explicitly. So, some of the highest frequencies were not present: (i) there are no hydrogen atoms and thus the fastest bond-stretching vibrations and the fastest bond-angle bending vibrations are not present; (ii) the system has no improper dihedral angles which would be the fastest motions in the system after bond-length vibrations and bond-angle bending vibrations. The second test system was the small (58 amino acids) protein BPTI including a variety of internal degrees of freedom. All simulations were carried out using the following conditions.

i. Vacuum boundary condition.
ii. An infinite cut-off radius, because the noise arising from the use of a cut-off scheme would heat the system and spoil the energy conservation. The fluctuation of the total energy (18) would be determined by the cut-off noise rather than by the time step used in the integration.
iii. When SHAKE [30] was applied, a relative geometric precision of 10^{-6} was used. In molecular dynamics of proteins using a cut-off radius a lower relative SHAKE tolerance of 10^{-4} can be used, since the cut-off noise will dominate the fluctuation in the total energy.

In order to obtain an equilibrated starting structure in the C_{100} system, an extended chain was simulated for 1 ns using standard conditions (2 fs time step, weak coupling to a temperature bath of 300 K using a coupling time τ_T of 0.1 ps [57]; bonds kept constrained using the SHAKE algorithm with a relative geometric tolerance of 10^{-4}, and twin-range cut-off radii of 0.8 and 1.4 nm). This leads to a compact, almost spherically shaped molecule. This new structure was then used as a starting structure for a simulation of 20 ps using the various force field and simulation parameter sets for which the total energy fluctuation was to be determined, still using temperature coupling to eliminate the energetic

TABLE I Oscillation or relaxation times (in fs) of various degrees of freedom or force field terms in the C_{100} simulations at 300 K. Deviations of force field parameters from GROMOS96 values are indicated by scaling factors. The maximum step size Δt is indicated and the corresponding average change of torsional-angle value per time step is given

	Force field term					
		Bonds constrained				
		Scaling factor s_{ba} for bond angle forces				
		With dihedral forces, $s_{dih} = 1.0$		*No dihedral angle forces, $s_{dih} = 0.0$*		
	No constraints, fully flexible	$s_{ba} = 1.0$	$s_{ba} = 0.3$	$s_{ba} = 0.2$	$s_{ba} = 0.1$	$s_{ba} = 0.0$
Single bond	30	–	–	–	–	–
Bond angle	30	85	128	171	171	–
Torsional angle	128	256	171	–	–	–
Max step size (fs)	6	17	17	25	25	22
Average $\Delta\varphi$ (°) per step	2.0	6.5	5.3	10.5	9.4	16.3

effect of a change in parameter settings used. Each equilibrated system was then simulated for 80 ps without temperature coupling. These 80 ps were used to determine energy conservation for the selected parameter combination as a function of Δt. For the second test system, the protein BPTI, the structure 1BPI of the Brookhaven protein databank [58] was used as a starting structure. The cysteine bridges were reduced, cysteine residues protonated. The following residues were used in their protonated form: lysine (Lys) and arginine (Arg). The solvent water molecules were removed except for the four internal water molecules in BPTI. The GROMOS96 vacuum force field [17], code 43B1, was used to simulate the protein for 1 ns at standard conditions. As for the C_{100} test system, parameters were then changed to various combinations and another 20 ps were simulated using temperature coupling to absorb the energetic effects of changing the parameters. Afterwards, 80 ps without temperature coupling were used for analysis. For determining the different frequencies and relaxation times, another 1 ps was appended to these simulations. A time step of 0.5 fs was used and every trajectory configuration was saved for analysis. All simulations were carried out using the GROMOS96 package and force field [17,39].

RESULTS

C_{100} Test System

When simulating the C_{100} test system using a fully flexible model (without SHAKE to constrain bonds), a time step of 6 fs can be achieved (Fig. 1). Using larger time steps leads to a loss of energy conservation due to integration errors. ΔE_{tot} is larger than ΔE_{kin}. The oscillation time of the vibrations of a C–C bond or a C–C–C bond-angle is around 30 fs (Table I). These are the fastest motions in the system and thus limit the time step. For a physically reasonable simulation ($\Delta E_{tot}/\Delta E_{kin} < 5 \times 10^{-2}$, Fig. 1), a time step of 1 fs should be used. The average torsional angle change $\Delta\varphi$ at the largest possible time step of 6 fs is 2°. Removing the bond stretching vibrations from the system using the SHAKE algorithm leads to an increase of the maximal time step by almost a factor of three. Use of even larger time steps causes SHAKE to fail because the distortions from ideal bond geometry become too large. The average $\Delta\varphi$ is also increased by a factor 3 indicating that the motion along torsional degrees of freedom is not inhibited by constraining bonds. Bond-angle vibrations, however, become slower compared to the flexible model, with an oscillation time of about 85 ps on average. A physically meaningful time step would lie around 4 fs for such a system. Reducing the highest frequencies of the system by scaling all bond-angle force constants K_θ by a factor of 0.3 does not affect the energy conservation very much. The maximum time step that can be used stays the same. The vibrational frequencies are slightly differently distributed between the bond-angle and torsional-angle degrees of freedom. Bond-angle vibrations become slower and torsional-angle motions faster. The average $\Delta\varphi$ per step stays about the same. Further reduction of the bond-angle force constant was not possible in the presence of torsional-angle interactions. The very soft bond-angle forces allow the bond-angles to become larger than 180°, causing a jump of 180° for the torsional angles defined by the same three atoms as define the bond-angle. These discontinuities in the torsional angle trajectories imply discontinuities in the torsional angle forces, which cause the system energy to diverge. So, when using even softer bond-angles, bond-angle force constants K_θ scaled by $s_{ba} = 0.2$ and 0.1, the torsional-angle interaction term of the force field was turned off. Both these simulations led to nearly identical results. The largest

TABLE II Oscillation or relaxation times (in fs) of various degrees of freedom or force field terms in the BPTI simulations. Deviations of force field parameters from GROMOS96 values are indicated by scaling factors. The maximum step size Δt is indicated and the corresponding average change $\Delta\varphi$ of torsion angle value per time step is given

	Force field term						
	$T = 300\,\mathrm{K}$					$T = 1000\,\mathrm{K}$	
	No constraints	*Bonds constrained*					
	$m_\mathrm{H} = 1.008\,\mathrm{u}$	$m_\mathrm{H} = 14.027\,\mathrm{u}$				$m_\mathrm{H} = 1.008\,\mathrm{u}$	$m_\mathrm{H} = 14.027\,\mathrm{u}$ $s_\mathrm{ba} = 0.4$ $s_\mathrm{imp} = s_\mathrm{dih} = 0.25$
				$s_\mathrm{ba} = 0.5$	$s_\mathrm{ba} = s_\mathrm{imp} = s_\mathrm{dih} = 0.4$		
Single bond H	10	–	–	–	–	–	–
Double bond	20	–	–	–	–	–	–
Single bond	30	–	–	–	–	–	–
Bond-angle 2H	20	20	57	102	102	20	102
Bond-angle 1H	32	22	64	73	85	32	85
bond-angle	21	57	73	85	102	73	102
improper H	51	51	51	73	85	51	128
Improper planar	51	102	64	73	128	102	85
Improper tetrahedral	73	57	47	73	85	43	85
Torsion H	34	32	73	64	171	34	171
Torsion	85	47	85	85	73	64	85
Double bond torsion	73	64	85	64	128	73	128
Max step size (fs)	2	4	6	6	14	3	12
Average $\Delta\varphi$ (°) per step	0.7	1.3	2.0	2.0	5.1	1.0	6.1

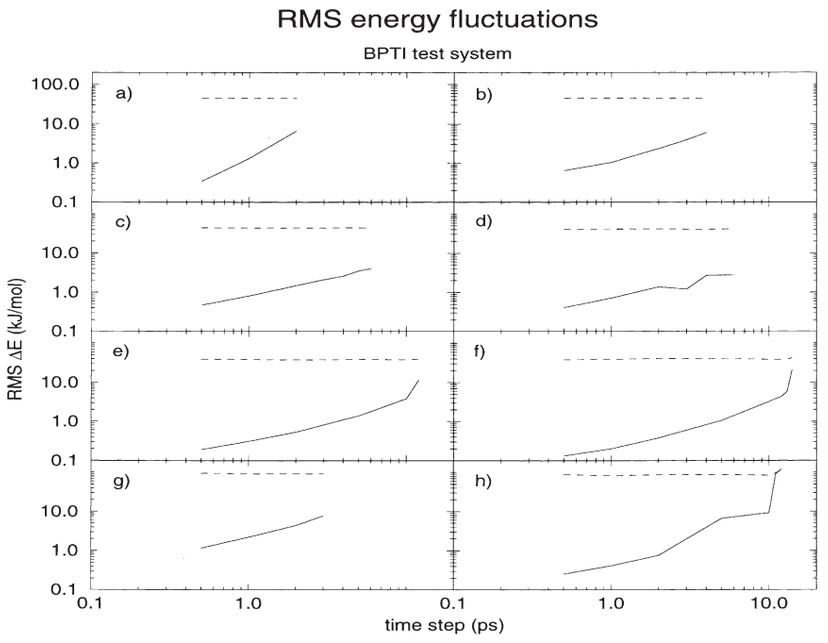

FIGURE 2 Root-mean-square fluctuations of total (solid line) and kinetic (dashed line) energy for the protein BPTI for different parameter settings as function of the time step at 300 K if not mentioned otherwise; (a) flexible bonds, (b) bonds constrained, (c) bonds constrained, hydrogen masses increased to 14.027 u, (d) bonds constrained, bond-angle force constants scaled by a factor of 0.5, hydrogen masses increased to 14.027 u, (e) bonds constrained, bond-angle force constants scaled by a factor of 0.4, improper-dihedral force constants scaled by a factor of 0.4, torsional force constants scaled by a factor of 0.4, hydrogen masses increased to 14.027 u, charges turned off, (f) bonds constrained, bond-angle force constants scaled by a factor of 0.4, improper-dihedral force constants scaled by a factor of 0.25, torsional force constants scaled by a factor of 0.25, hydrogen masses increased to 14.027 u, charges turned off, (g) bonds constrained, $T = 1000$ K, (h) bonds constrained, bond-angle force constants scaled by a factor of 0.4, improper-dihedral force constants scaled by a factor of 0.25, torsional force constants scaled by a factor of 0.25, hydrogen masses increased to 14.027 u, charges, turned off, $T = 1000$ K.

achievable time step was 25 fs, since bond-angle vibrations have an oscillation time of about 170 fs. The average torsional-angle change at $\Delta t = 25$ fs was around 10°. Completely removing bond-angle forces ($s_{ba} = 0.0$) reduced the maximum time step to 22 fs. However, the large freedom of the system (now mainly a covalently bound chain of van der Waals atoms) led to a large average $\Delta\varphi$ of 16.3°. Energy conservation was very good for all the simulations using softened or smoothed forces. At a time step of 10 fs, the criterion for a physically correct simulation, $\Delta E_{tot}/\Delta E_{kin} < 5 \times 10^{-2}$, was fulfilled. But, since the force field used is very unphysical, this time step cannot be used to simulate equilibrium conditions of a real system.

BPTI Test System

Simulating the protein BPTI involves a wider spread of and also higher frequencies than the simple C_{100} test system. The fastest motions in a biomolecular simulation are the vibrations of bonds involving hydrogen atoms. The period of one oscillation of such a bond is around l0 fs (Table II). Double bonds with their higher force constant have a period of about 20 fs compared to around 30 fs for a single bond involving only heavy atoms. Using the flexible bond model, one can reach time steps of 2 fs until the total energy starts to diverge. At this time step, the average torsional-angle change $\Delta\varphi$ is 0.7° per time step. For a physically meaningful simulation with vacuum boundary conditions ($\Delta E_{tot}/\Delta E_{kin} < 5 \times 10^{-2}$, Fig. 2), a time step of 1 fs can be used. Removal of the highest-frequency motions by applying constraints using the SHAKE algorithm increases the largest possible (until geometric distortions get too large for SHAKE to correct them) time step to 4 fs. The energy conservation criterion limits the time step to 2–3 fs. The fastest motions under these conditions, which are usually applied in standard biomolecular simulations, are the vibrations of bond angles with hydrogen atoms. The oscillation periods of these bond-angle vibrations are around 20 fs. By increasing the masses of hydrogen atoms to a value similar to those of the other simulated atoms (in this simulation 14.027 u, the mass of a nitrogen atom) one can increase the time step further to 6 fs. Since static equilibrium properties do not depend on masses, physical simulations can still be carried out under these conditions as long as dynamical properties are not of interest. The high-frequency bond-angle vibrations are no longer present in the system. The fastest motions are now improper-dihedral-angle motions, although other degrees of freedom show similar frequencies. A time step of 5 fs can be applied

without the energy conservation being violated to an untolerable extent. Decreasing the bond-angle force constants in addition to the scaling of hydrogen masses brings no further improvement in time step. Under both conditions, using a time step of 6 fs, the average torsional-angle change is 2° per step. Since the frequencies corresponding to the various force field terms are very similar, further improvement could only be achieved by reducing bond-angle-, improper dihedral-, and torsional-angle force constants simultaneously. To do so, charges were set to zero, because otherwise, hydrogen atoms (which do not have a van der Waals interaction) turned out to be attracted too strongly by negatively charged neighboring atoms (e.g. in the side chain of the amino acid Arg), causing an electrostatic collapse.

The largest time step for the simulation of BPTI in vacuum was possible using the following parameter set:

hydrogen masses (m_H) = 14.027 u,
bond-angle force constants K_{0_n} scaled by a factor of 0.4,
improper-dihedral force constants K_{ξ_n} scaled by a factor of 0.4,
torsional-angle force constants K_{φ_n} scaled by a factor of 0.4,
all charges set to zero.

The largest possible time step was 14 fs with an average $\Delta\varphi$ per step of 5.1°. A further reduction of force constants (e.g. scaling K_{ξ_n} and K_{φ_n} by a factor of 0.25) did not lead to larger time steps. Like in the C_{100} test system, physically meaningful equilibrium simulations can no longer be performed using this parameter set, but it is useful for sampling conformational space. Increasing the temperature to 1000 K did not change the results very much.

DISCUSSION

Molecular dynamics using Cartesian coordinates and Newton's equations of motion is an efficient method to integrate the classical equations of motion. When water molecules solvating a biomolecule are explicitly treated, their fast librational motions limit the integration time step to 2 fs, if physically realistic dynamics is to be obtained. In the absence of explicit water, the time step can be lengthened to 6 fs when using bond-length constraints and by applying hydrogen mass scaling. The latter will not change significantly the static equilibrium properties, but will modify the dynamics. When using in addition a smoothed, very soft atomic interaction function or force field and neglecting electrostatic interactions at the expense of loosing physical correctness, a time step of 14 fs can be reached. The use of another algorithm than SHAKE to constrain bonds would probably enlarge this maximum time step even further. The time step that can be used in Cartesian dynamics is of a size similar to the ones employed in torsional-angle molecular dynamics. The use of Newton's equations of motion instead of Lagrange's equations of motion, however, has a number of advantages:

i. Newton's equations of motion in Cartesian space are much simpler than Lagrange's equations of motion using generalized, torsional, coordinates, so one can use simpler algorithms to integrate the equations of motion. This leads to higher computational efficiency, less computational noise and a reduced chance of software errors.
ii. Metric tensor effects play a non-negligible role when using generalized, non-Cartesian, coordinates and have to be corrected if a Boltzmann-weighted ensemble is to be generated.
iii. The complete freezing of bond-angle degrees of freedom without adjustment of the physical force field to these conditions reduces atomic motions and thus sampling of conformational space.
iv. The transition from a physically correct force field to a smoothed, very soft force field and back is continuous and more straight-forward than when using dynamics in torsional coordinates
v. Using generalized, torsional coordinates, it is virtually impossible to treat chain closure of a chain of particles using bond constraints while maintaining flexibility within the ring. So, disulfide bridges have to be treated as unconstrained degrees of freedom and the corresponding high-frequency motions are still present in the simulated system.

Considering all these points, for equilibrium simulations, Newton's equations of motion in Cartesian coordinates are to be preferred to Lagrange's equations of motion using generalized, non-Cartesian, coordinates.

Acknowledgements

Financial support was obtained from the Schweizerischer Nationalfonds, project number 21-57069.99, which is gratefully acknowledged.

References

[1] Rahman, A. (1964) "Correlations in the motion of atoms in liquid argon", *Phys. Rev.* **136A**, 405.
[2] Allen, M.P. and Tildesley, D.J. (1987) *Computer Simulation of Liquids* (Clarendon Press, Oxford).

[3] McCammon, J.A., Gelin, B.R. and Karplus, M. (1977) "Dynamics of folded proteins", *Nature* **267**, 585.

[4] van Gunsteren, W.F. and Berendsen, H.J.C. (1990) "Computer simulations of molecular dynamics: methodology, applications and perspectives in chemistry", *Angew. Chem., Int. Ed. Engl.* **29**, 992.

[5] Tieleman, D.P., Marrink, S.J. and Berendsen, H.J.C. (1997) "A computer perspective of membranes: Molecular dynamics studies of lipid bilayer systems", *Biochim. Biophys. Acta* **1331**, 235.

[6] van Gunsteren, W.F., Kaptein, R. and Zuiderweg, E.R.P. (1984) "Use of molecular dynamics computer simulations when determining protein structure by 2D-NMR", *Proceedings NATO/CECAM Workshop on Nucleic Acid Conformation and Dynamics* Orsay, (CECAM, France), pp 79–97.

[7] Kaptein, R., Zuiderweg, E.R.P., Scheek, R.M., Boelens, R. and van Gunsteren, W.F. (1985) "A protein structure from nuclear magnetic resonance data, lac repressor headpiece", *J. Mol. Biol.* **182**, 179.

[8] Brünger, A.T. and Nilges, M. (1993) "Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR-spectroscopy", *Quart. Rev. Biophys.* **26**, 49.

[9] van Gunsteren, W.F., Brunne, R.M., Gros, P., van Schaik, R.C., Schiffer, C.A. and Torda, A.E. (1994) "Accounting for molecular mobility in structure determination based on nuclear magnetic resonance spectroscopic and X-ray diffraction data", Methods in Enzymology: Nuclear Magnetic Resonance, (Academic Press, New York) **Vol. 239**, pp 619–654.

[10] van Gunsteren, W.F., Bonvin, A.M.J.J., Daura, X. and Smith, L.J. (1999) "Aspects of modeling biomolecular structure on the basis of spectroscopic or diffraction data", Structure Computation and Dynamics in Protein NMR, Biological Magnetic Resonance, (Plenum Press, New York) **Vol. 17**, pp. 3–35.

[11] Güntert, P., Mumenthaler, C. and Wüthrich, K. (1997) "Torsion angle dynamics for NMR structure calculation with the new program DYANA", *J. Mol. Biol.* **273**, 283.

[12] Stein, E.G., Rice, L.M. and Brünger, A.T. (1997) "Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation", *J. Magn. Res.* **124**, 154.

[13] Schuler, L.D. and van Gunsteren, W.F. (2000) "On the choice of dihedral angle potential energy functions for *n*-alkanes", *Mol. Simulat.* **25**, 301.

[14] Rahman, A. and Stillinger, F.H. (1971) "Molecular dynamics study of liquid water", *J. Chem. Phys.* **55**, 3336.

[15] Berendsen, H.J.C. and van Gunsteren, W.F. (1984) "Molecular dynamics simulations: techniques and approaches", *Molecular Liquids—Dynamics and Interactions* NATO ASI Series, (Reidel, Dordrecht) **Vol. C135**, pp 561–564.

[16] van Gunsteren, W.F., Boelens, R., Kaptein, R., Scheek, R.M. and Zuiderweg, E.R.P. (1985) "An improved restrained molecular dynamics technique to obtain protein tertiary structure from nuclear magnetic resonance data", *Molecular Dynamics and Protein Structure* (Polycrystal Book Service, P.O. Box 27, Western Springs, III. 60558, USA), pp. 92–99.

[17] vanGunsteren, W.F., Billeter, S.R., Eising, A.A., Hünenberger, P.H., Krüger, P., Mark, A.E., Scott, W.R.P. and Tironi, I.G. (1996) *Biomolecular Simulation: The GROMOS96 Manual and User Guide* (Vdf Hochschulverlag, Zürich, Switzerland).

[18] van Gunsteren, W.F. and Karplus, M. (1982) "Effect of constraints on the dynamics of macromolecules", *Macromolecules* **15**, 1528.

[19] Rice, L.M. and Brünger, A.T. (1994) "Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement", *Proteins* **19**, 277.

[20] Levitt, M., Hirshberg, M., Sharon, R., Laidig, K.E. and Daggett, V. (1997) "Calibration and testing of a water model for simulation of the molecular dynamics of proteins and nucleic acids in solution", *J. Phys. Chem. B* **101**, 5051.

[21] Tuckerman, M., Berne, B.J. and Martyna, G.J. (1992) "Reversible multiple time scale molecular dynamics", *J. Chem. Phys.* **97**, 1990.

[22] Watanabe, M. and Karplus, M. (1993) "Dynamics of molecules with internal degrees of freedom by multiple time-step methods", *J. Chem. Phys.* **99**, 8063.

[23] Izaguirre, J.A., Reich, S. and Skeel, R.D. (1999) "Longer time steps for molecular dynamics", *J. Chem. Phys.* **110**, 9853.

[24] Bennett, C.H. (1975) "Mass tensor molecular dynamics", *J. Comput. Phys.* **19**, 267.

[25] Ryckaert, J.-P. (1991) *Computer-Simulations in Material Science* E205 of NATO ASI Series, (Kluwer Academic Publishers, Dordrecht) **Vol. E205**, pp 43–66.

[26] Pomès, R. and McCammon, J.A. (1990) "Mass and step length optimization for the calculation of equilibrium properties by molecular dynamics simulation", *Chem. Phys. Lett.* **166**, 425.

[27] Feenstra, K., Hess, B. and Berendsen, H.J.C. (1999) "Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems", *J. Comput. Chem.* **20**, 786.

[28] Katz, H., Walter, R. and Somorjay, R.L. (1979) "Rotational dynamics of large molecules", *Comput. Chem.* **3**, 25.

[29] Mazur, A.K., Dorofeev, V.E. and Abagyan, R.A. (1991) "Derivation and testing of explicit equations of motion for polymers described by internal coordinates", *J. Comput. Phys.* **92**, 261.

[30] Ryckaert, J.-P., Ciccotti, G. and Berendsen, H.J.C. (1977) "Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes", *J. Comput. Phys.* **23**, 327.

[31] Ciccotti, G., Ferrario, M. and Ryckaert, J.-P. (1982) "Molecular dynamics of rigid systems in cartesian coordinates: a general formulation", *Mol. Phys.* **47**, 1253.

[32] Hess, B., Bekker, H., Berendsen, H.J.C. and Fraaije, J.G.E.M. (1997) "LINCS: a linear constraint solver for molecular simulations", *J Comput. Chem.* **18**, 1463.

[33] Reich, S. (1995) "Smoothed dynamics of highly oscillatory hamiltonian systems", *Physica D* **89**, 28.

[34] van Gunsteren, W.F., Huber, T., Torda, A.E. (1995) "Biomolecular modelling: overview of types of methods to search and sample conformational space", in *Conference Proceedings*, American Institute of Physics, European Conference on Computational Chemistry (E.C.C.C 1), Vol. 330, pp. 253–268.

[35] Frenkel, D. (1993) Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications, (Escom, Leiden, The Netherlands) **Vol. 2**, pp. 37–66.

[36] van Gunsteren, W.F. (1993) "Molecular dynamics and stochastic dynamics simulation: a primer", Computer Simulation of Biomolecular systems, Theoretical and Experimental Applications, (Escom, Leiden, The Netherlands) **Vol. 2**, pp. 3–36.

[37] Mazur, A.K. (1998) "Hierarchy of motions in protein dynamics", *J. Phys. Chem. B* **102**, 473.

[38] van Gunsteren, W.F. (1991) "Computer simulation of biomolecular systems: overview of time saving techniques", Advances in Biomolecular Simulations, (American Institute of Physics **Vol. 239**, pp. 131–146.

[39] Scott, W.R.P., Hünenberger, P.H., Tironi, I.G., Mark, A.E., Billeter, S.R., Fennen, J., Torda, A.E., Huber, T., Krüger, P. and van Gunsteren, W.F. (1999) "The GROMOS biomolecular simulation program package", *J. Phys. Chem. A* **103**, 3596.

[40] Beutler, T.C., Mark, A.E., van Schaik, R.C., Gerber, P.R. and van Gunsteren, W.F. (1994) "Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations", *Chem. Phys. Letters* **222**, 529.

[41] Tironi, I.G., Sperb, R., Smith, P.E. and van Gunsteren, W.F. (1995) "A generalized reaction field method for molecular dynamics simulations", *J. Chem. Phys.* **102**, 5451.

[42] van Gunsteren, W.F., Luque, F.J., Timms, D. and Torda, A.E. (1994) "Molecular mechanics in biology: from structure to function, taking account of solvation", *Ann. Rev. Biophys. Biomol. Struct.* **23**, 847.

[43] Wittenburg, J. (1977) *Dynamics of Systems of Rigid Bodies* (Teubner, Stuttgart).

[44] Bae, D.S. and Haug, E.J. (1987) "A recursive formulation for constrained mechanical system dynamics: Part I. Open loop systems", *Mech. Struct. Mach.* **15**, 359.

[45] Bae, D.-S. and Haug, E.J. (1988) "A recursive formulation for constrained mechanical system dynamics: Part II. Closed loop systems", *Mech. Struct. Mach.* **15**, 481.

[46] Jain, A., Vaidehi, N. and Rodriguez, G. (1993) "A fast recursive algorithm for molecular dynamics simulation", *J. Comput. Phys.* **106**, 258.
[47] Turner, J.D., Weiner, P.K., Chun, H.M., Lupi, V., Gallion, S. and Singh, U.C. (1993) Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications, (Escom, Leiden, The Netherlands) **Vol. 2**, pp. 535–555.
[48] Mathiowetz, A.M., Jain, A., Karasawa, N. and Goddard, III, W.A. (1994) "Protein simulations using techniques suitable for very large systems: the cell multipole method for nonbonded interactions and the Newton–Euler inverse mass operator method for internal coordinate dynamics", *Proteins* **20**, 227.
[49] Goldstein, H. (1980) *Classical Mechanics*, 2nd ed. (Addison-Wesley, Reading, MA, USA).
[50] van Gunsteren, W.F. and Berendsen, H.J.C. (1977) "Algorithms for macromolecular dynamics and constraint dynamics", *Mol. Phys.* **34**, 1311.
[51] Berendsen, H.J.C. and van Gunsteren, W.F. (1986) "Practical, algorithms for dynamic simulations", *Molecular-Dynamics Simulation of Statistical–Mechanical Systems* Proceedings of the International School of Physics Enrico Fermi, Course 97, (North-Holland, Amsterdam), pp 43–65.
[52] Mazur, A.K. (1997) "Common molecular dynamics algorithms revisited: Accuracy and optimal time steps of Störmer-leapfrog integrators", *J. Comput. Phys.* **136**, 354.
[53] Hockney, R.W. and Eastwood, J.W. (1981) *Computer Simulation Using Particles* (McGraw-Hill, New York).
[54] Verlet, L. (1967) "Computer experiments on classical fluids. I thermodynamical properties of Lennard–Jones molecules", *Phys. Rev.* **159**, 98.
[55] van Gunsteren, W.F. (1980) "Constrained dynamics of flexible molecules", *Mol. Phys.* **40**, 1015.
[56] Luty, B.A., Davis, M.E., Tironi, I.G. and van Gunsteren, W.F. (1994) "A comparison of particle–particle–particle–mesh and Ewald methods for calculating electrostatic interactions in periodic molecular systems", *Mol. Simulat.* **14**, 11.
[57] Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R. (1984) "Molecular dynamics with coupling to an external bath", *J. Chem. Phys.* **81**, 3684.
[58] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, Jr, E.F., Brice, Jr, M.D., Rodgers, Jr, J.R., Kennard, Jr, O., Shimanouchi, Jr, T. and Tasumi, Jr, M. (1977) "The protein data bank: a computer-based archival file for macromolecular structures", *J. Mol. Biol.* **112**, 535.