Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Molecular Graphics and Modelling 22 (2004) 349–357
Journal of Molecular Graphics and Modelling
www.elsevier.com/locate/JMGM

New perspectives on multiple-copy, mean-field molecular dynamics methods

Christopher Adam Hixson, Jermont Chen[1], Zunnan Huang, Ralph A. Wheeler*

Department of Chemistry and Biochemistry, University of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

Abstract

Mean-field molecular dynamics (MD) techniques are designed to improve phase-space sampling in MD simulations. Reviewed here are theoretical and practical contributions from our group, and the ideas our contributions are based upon, beginning with the original time-dependent Hartree technique, locally enhanced sampling (LES), and a new, purely classical derivation of multiple-copy, mean-field equations of motion. This view is used to provide new insights into approximations inherent in LES, as well as the basis for a recently proposed method, ensembles eXtracted from atomic coordinate transformations (the EXACT approximation).

Keywords: Mean-field molecular dynamics; Time-dependent Hartree technique; Locally enhanced sampling

1. Introduction

Molecular dynamics (MD) has been a part of computational chemistry for some time [1–3] and has proven its value for solving problems in classical mechanics and quantum mechanics [1,4,5]. Despite this fact, MD has several well known limitations. One well-known limitation is the difficulty an MD simulation experiences in moving from one energy well to another. For example, if one is interested in using MD to find global minima on a complicated potential energy surface, a long simulation must be performed before there is much chance that all available areas of phase space are explored. Practically, even this may not be enough, but there are methods available to address this difficulty [6–14].

A popular method of addressing the above problem uses the same laws of physics in the solution that are the source of the problem. This method, called simulated annealing [6,12,15], raises and lowers the temperature to facilitate the movement of the system between energy minima. Raising the temperature increases the probability that a barrier crossing can occur; subsequent temperature lowering traps the system, presumably in a lower energy well.

Other possible solutions for the sampling problem have been proposed [7,8,16]. One group of methods, though, uses an approximation of the laws of physics in order to hasten energy minima transitions. These methods, called mean-field methods [10,17,18], divide the system into a bath and a smaller, more interesting section. The smaller section is copied several times, and the simulation is run using a mean-field approximation. Stated another way, a group of non-interacting copies of the smaller system are allowed to interact with the larger bath. The force each particle in the copied sections experiences is the force it would normally feel, while the force each particle in the bath experiences is the averaged force of the bath's interaction with the copied sections. Despite the fact that the trajectories generated with such methods necessarily do not correspond to physically possible trajectories [19–24], they have been used to find global minima [25–28], to study non-equilibrium behavior [29,30], to enhance free-energy calculations [27], and to search for molecules that bind to an enzyme active site [31–33]. Though many mean-field methods exist [10,17,21,33], this work will focus on the most commonly used method, called locally enhanced sampling (LES) [18], and our contribution, called ensembles eXtracted from atomic coordinate transformations (the EXACT approximation) [19,24].

The mean-field approximation, though resulting in desirable features can also cause defects in trajectories and quantities derived from them [19–24]. The consequences of such approximations include the fact that mean-field trajectories violate the equipartition of energy theorem [20–22] and that local minima of a mean-field trajectory do not necessarily correspond to minima on a physically accurate trajectory

* Corresponding author.
E-mail address: rawheeler@ou.edu (R.A. Wheeler).
[1] Present address: Department of Chemistry, Headquarters, US Air Force Academy, 235 Fairchild Dr., Suite 2N255, USAF Academy, CO 80840-6230, USA.

1093-3263/$ – see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2003.12.002

[23]. The equipartition of energy violation is especially important and has been studied by many workers [19–22,24].

This group's contributions [19,24,28] to this field include both theoretical and practical contributions important for understanding and improving mean-field methods. For example, we have provided two new approaches that result in mean-field equations of motion. These methods each provide unique insights. Further, we developed a method that can be used to control the degree of approximation included in a calculations, allowing a researcher to decide the degree of sampling or accuracy a calculation should posses.

This paper is divided into three sections. The first reviews studies that inspired our work, the second reviews our theoretical contributions, and the last section shares some of our practical experience and characterizes some results.

2. Theoretical background

2.1. Original TDH approximation

Gerber, Buch, and Ratner [17] originally devised a mean-field method by taking the classical limit of the time-dependent self-consistent field method (TDSCF). This method is described below by first describing the TDSCF method, and then explaining the path to the classical limit that defines the self-consistent trajectory bundles method [17]. This method was used to characterize the classical vibrations of small molecules, and its copies are designed to represent different normal modes of the molecule's vibrations.

The TDSCF method starts with the assumption that a system can be described by a Hartree-product wave function,

$$\Psi(\bar{x}, t) = \prod_{i=1}^{N} \psi_i(\bar{x}_i, t), \tag{1}$$

where ψ is the wavefunction, $\bar{x}$ a vector describing all N degrees of freedom in the system, ψ_i a wavefunction which describes the ith normal mode, and $\bar{x}_i$ is a vector that describes that normal mode. Time is represented by t.

Inserting Eq. (1) into the time-dependent Schrödinger equation and then operating with $<|\psi_i$ gives

$$i\bar{h}\frac{\partial \phi_i}{\partial t} = h_i^{\mathrm{SCF}}(\bar{x}_i, t)\phi_i(\bar{x}_i, t), \tag{2}$$

where $\phi_i(\bar{x}_i, t)$ is $\mathrm{e}^{iwt}\psi_i(\bar{x}_i, t)$. Most importantly, h_i^{SCF} is the kinetic and potential energies of the ith normal-mode plus an extra average term. So,

$$h_i^{\mathrm{SCF}}(\bar{x}_i, t) = T_i + V_i(\bar{x}_i) + \bar{V}_i(\bar{x}_i, t). \tag{3}$$

The final term of this expression is an averaged potential of the form

$$\bar{V}_i(x_i, t) = \sum_{j \neq i} < \phi_j | V_{ij}(\bar{x}_i, \bar{x}_j) | \phi_j >, \tag{4}$$

where i and j are normal-mode labels. Thus, V_{ij} is the interaction energy between modes i and j. This term was the original basis for the development of mean-field methods, which include LES, and the self-consistent trajectory bundle method in particular. It is a "mean-field" term because it averages the interaction energy of the ith normal mode with all the other normal modes of the system.

The self-consistent trajectory bundle method is the classical correspondence of the above method, and redefines the average classical potential as

$$\bar{V}_i(\bar{x}_i, t) = \frac{1}{N}\sum_{j \neq i}\sum_{\lambda=1}^{C} V_{ij}(\bar{x}_i, \bar{x}_j^{(\lambda)}), \tag{5}$$

where i and j are normal mode labels, and $\bar{x}_j^{(\lambda)}$ describes the λth copy of the jth normal mode. So in this method, all the normal modes are simultaneously simulated, but one is specially prepared in different states. These states interact with the other normal modes according to the potential described in Eq. (5), but not with other states.

2.2. Phase space TDH

Elber and Karplus [18] begin their derivation of LES with the assumption that the classical phase-space density function can be expressed as a product of the "copied" sub-system's density and the "bath" sub-system's density:

$$\rho(\bar{X}, t) = \rho_s(\bar{X}_s, t)\rho_b(\bar{X}_b, t). \tag{6}$$

Here $\bar{X}$ is a vector representing all degrees of freedom in the system and t is the time. They further assumed that the bath's density can be written as a single delta function:

$$\rho_b(\bar{X}_b, t) = \delta(\bar{X}_b(t)), \tag{7}$$

while the copied sub-system's density can be written as a "swarm" of delta functions. Thus,

$$\rho_s(\bar{X}_s, t) = \sum_{k=1}^{C} w_{sk}\delta(\bar{X}_{sk}(t)), \tag{8}$$

where each delta function represents the positions in phase space of the various copies, and w_{sk} is a weighting function. In LES the weighting functions are generally taken to be $1/C$, where C is the number of copies. Elber and Karplus refer to this as the time-dependent Hartree (TDH) approximation, in analogy with Gerber, Buch, and Ratner's [17] previous work, described above. Using this assumed density, they derive the approximate equations of motion by proving that

$$\frac{\partial < Q_j >}{\partial t} = \left\langle \frac{\partial H}{\partial P_j} \right\rangle; \qquad \frac{\partial < P_j >}{\partial t} = -\left\langle \frac{\partial H}{\partial Q_j} \right\rangle, \tag{9}$$

where H is the Hamiltonian for one copy interacting with the bath, Q_j is a generalized coordinate of the jth particle, and P_j is the momentum of the jth particle. They then use their

assumed density to determine the LES equations of motion

$$\dot{q}_{i,k} = \frac{\partial H_k}{\partial p_{i,k}}, \quad \dot{p}_{i,k} = -\frac{\partial H_k}{\partial q_{i,k}}; \qquad \dot{Q}_i = \sum_{k=1}^{C} w_k \frac{\partial H_k}{\partial P_i},$$

$$\dot{P}_i = -\sum_{k=1}^{C} w_k \frac{\partial H_k}{\partial Q_i}. \tag{10}$$

The index i refers to the particle and the index k refers to the copy. The lower case variables refer to the copied sub-system while the uppercase variables refer to the "bath."

2.3. Limitations of LES

Because LES and other mean-field techniques employ an approximate dynamics, there are limits to the ability of these trajectories to model a Newtonian trajectory. One of the limitations include the fact that LES violates the equipartition of energy theorem [20,21]. Some have speculated [20–22] that this violation results in the "temperature disparity problem," which is a failure of the sub-system and bath temperatures to reach the same equilibrium value. Another problem is that geometry-optimization problems solved in LES are definitive only if the global energy minimum is desired. Stultz and Karplus have proven that local-minima found using LES cannot be assumed to be minima on the original energy surface [23].

3. Theoretical contributions

3.1. Classical mechanical approach

We derived [19] mean-field equations of motion in an attempt to better understand the approximations involved in methods such as LES and to improve on them. The derivation begins by writing the Hamiltonian of C non-interacting copies of the original system,

$$H = \sum_{k=1}^{C} \left(\sum_{i=1}^{S} \frac{p_{i,k}^2}{2m_i} + \sum_{i>j=1}^{S} V(q_{i,k}, q_{j,k}) + \sum_{i=1}^{S}\sum_{j=1}^{N} V(q_{i,k}, Q_{j,k}) + \sum_{i=1}^{N} \frac{P_{i,k}^2}{2M_i} + \sum_{i>j=1}^{N} V(Q_{i,k}, Q_{j,k}) \right). \tag{11}$$

Lowercase variables refer to the sub-system of interest while uppercase variables refer to the bath. The constants C, N, and S represent the number of copies, the number of particles in the bath, and the number of particles in the sub-system of interest, respectively. Upper and lowercase q's indicate positions, p's indicate momenta, and m's indicate masses. The first two terms are the energy due to the copied subsystem, the middle term is the interaction energy between the copies and the bath, and the final two terms are the energy due to the bath. This Hamiltonian, for the sake of simplicity, assumes a pair form of the potential. We then make a point transformation defined by

$$Q_i = \frac{1}{C}\left(\sum_{k=1}^{C} Q_{i,k}\right) \qquad P_i = \frac{1}{C}\left(\sum_{k=1}^{C} P_{i,k}\right)$$

$$Q'_{i,l} = \frac{1}{C}\left(Q_{i,1} + \sum_{k=2}^{C} c_{l,k} Q_{i,k}\right)$$

$$P'_{i,l} = \frac{1}{C}\left(P_{i,1} + \sum_{k=2}^{C} c_{l,k} P_{i,k}\right), \tag{12}$$

where $\{Q_i\}_{i=1}^{N}$ and $\{P_i\}_{i=1}^{N}$ are the "major" variables, $\{Q'_{i,l}\}_{i=1,l=2}^{N,C}$ and $\{P'_{i,l}\}_{i=1,l=2}^{N,C}$ are "minor" variables, and $\{c_{i,j}\}$ are the transform coefficients. The constant, C, is equal to the number of copies. The "major" variables denote the average coordinates of all copies of the bath particles, and the "minor" variables complete the description of the exact dynamics.

For future use, the transform coefficients possess three useful properties—orthonormality, Hermiticity, and zero-average

$$1 + \sum_{k=2}^{C} c_{n,k} c_{m,k} = C\delta_{m,n}, \tag{13a}$$

$$c_{i,j} = c^*_{j,i}, \tag{13b}$$

and

$$1 + \sum_{k=2}^{C} c_{n,k} = 0. \tag{13c}$$

Eq. (13b) is actually a more stringent requirement than necessary, as we expect all the coefficients to be real. This transformation was used in a previous work [22] attempting to explain LES, but its solution was limited to the harmonic approximation.

For a general potential, the transformed Hamiltonian can be written as [19].

$$\tilde{H} = \sum_{i=1}^{S}\sum_{k=1}^{C} \frac{p_{i,k}^2}{2m_i} + \sum_{i>j=1}^{S}\sum_{k=1}^{C} V(q_{i,k}, q_{j,k}) + \sum_{j=1}^{N} \frac{CP_j^2}{2M_j} + \sum_{i>j=1}^{N}\sum_{k=1}^{C} V\left(\left(Q_i + \sum_{m=2}^{C} c_{m,k} Q'_{i,m}\right), \left(Q_j + \sum_{m=2}^{C} c_{m,k} Q'_{j,m}\right)\right) + \sum_{j=1}^{C}\sum_{k=2}^{C} \frac{CP'^2_{j,k}}{2M_j} + \sum_{i=1}^{S}\sum_{j=1}^{N}\sum_{k=1}^{C} V\left(q_{i,k}\left(Q_j + \sum_{m=2}^{C} c_{m,k} Q'_{j,m}\right)\right). \tag{14}$$

The first two terms represent the Hamiltonian of the copied particles. The next three terms in Eq. (14) are the kinetic energy of the major variables of the bath, the potential energy

due to the bath, and the kinetic energy due to the minor variables of the bath. The final term contains the interaction energy between the two subsystems. The equations of motion for the major and minor variables can then be obtained [19]

$$\dot{q}_{i,k} = \frac{\partial \tilde{H}}{\partial p_{i,k}}, \quad \dot{p}_{i,k} = -\frac{\partial \tilde{H}}{\partial q_{i,k}}; \qquad \dot{Q}_i = \frac{1}{C}\frac{\partial \tilde{H}}{\partial P_i},$$
$$\dot{P}_i = -\frac{1}{C}\frac{\partial \tilde{H}}{\partial Q_i}; \qquad \dot{P}'_{i,k} = -\frac{1}{C}\frac{\partial \tilde{H}}{\partial Q'_{i,k}}, \quad \dot{Q}'_{i,k} = \frac{1}{C}\frac{\partial \tilde{H}}{\partial P'_{i,k}} \tag{15}$$

If the bath particles are initially at the same phase-space point and all of the coordinates of the minor variables are neglected, the LES equations of motion are recovered.

3.2. Liouville operator approach

A complimentary view of the results from the previous section can also be gained by using a phase-space density approach [19]. In fact, most theoretical work done regarding mean-field methods depends on a phase-space density approach [10,18,20–22], and often employs the Liouville equation.

The start of the analysis requires a density that satisfies the Liouville equation [34],

$$\rho(\bar{X}, t) = \mathrm{e}^{-\hat{L}t}\rho(\bar{X}, 0) \tag{16}$$

where $\hat{L}$ is the Liouville operator,

$$\hat{L} = \sum_{i=1}^{S}\sum_{k=1}^{C}\left(\frac{\partial H}{\partial p_{i,k}}\frac{\partial}{\partial q_{i,k}} - \frac{\partial H}{\partial q_{i,k}}\frac{\partial}{\partial p_{i,k}}\right) + \sum_{i=1}^{N}\sum_{k=1}^{C}\left(\frac{\partial H}{\partial P_{i,k}}\frac{\partial}{\partial Q_{i,k}} - \frac{\partial H}{\partial Q_{i,k}}\frac{\partial}{\partial P_{i,k}}\right), \tag{17}$$

$\rho(\bar{X}, 0)$ is the initial phase-space density:

$$\rho(\bar{X}, 0) = \left(\prod_{i=1}^{S}\prod_{k=1}^{C}\delta(q_{i,k} - q_{i,k,0})\delta(p_{i,k} - p_{i,k,0})\right) \times \left(\prod_{i=1}^{N}\prod_{k=1}^{C}\delta(Q_{i,k} - Q_{i,k,0})\delta(P_{i,k} - P_{i,k,0})\right) \tag{18}$$

and $\bar{X}$ is a point in phase-space. After applying the transformation defined by Eq. (12) to the Liouville operator and the initial probability density, we obtain

$$\hat{\tilde{L}} = \sum_{i=1}^{S}\sum_{k=1}^{C}\left(\frac{\partial \tilde{H}}{\partial p_{i,k}}\frac{\partial}{\partial q_{i,k}} - \frac{\partial \tilde{H}}{\partial q_{i,k}}\frac{\partial}{\partial p_{i,k}}\right) + \frac{C}{n^2}\left[\sum_{i=1}^{N}\left(\frac{\partial \tilde{H}}{\partial P_i}\frac{\partial}{\partial Q_i} - \frac{\partial \tilde{H}}{\partial Q_i}\frac{\partial}{\partial P_i}\right) + \sum_{i=1}^{N}\sum_{k=2}^{C}\left(\frac{\partial \tilde{H}}{\partial P'_{i,k}}\frac{\partial}{\partial Q'_{i,k}} - \frac{\partial \tilde{H}}{\partial Q'_{i,k}}\frac{\partial}{\partial P'_{i,k}}\right)\right]. \tag{19}$$

With the assumption that the bath particles are in the same locations in each of the copies, the transformed density can be written as:

$$\tilde{\rho}(\bar{\bar{X}}, 0) = \frac{1}{|J(\bar{X}, \bar{\bar{X}})|}\left(\prod_{i=1}^{S}\prod_{k=1}^{C}\delta(q_{i,k} - q_{i,k,0})\delta(p_{i,k} - p_{i,k,0})\right) \times \left(\prod_{i=1}^{N}\delta(Q_i - Q_{i,0})\delta(P_i - P_{i,0})\prod_{k=2}^{C}\delta(Q'_{i,k})\delta(P'_{i,k})\right). \tag{20}$$

Here $J(\bar{X}, \bar{\bar{X}})$ is the Jacobian of the transformation and $\bar{\bar{X}}$ is a point in the transformed phase-space. The transformed time dependent phase-space density, $\tilde{\rho}(\bar{\bar{X}}, t)$, is written using Eqs. (19) and (20) in the Liouville equation. Following Zheng and Zheng's work [22] the reduced density, $\tilde{\rho}'(\bar{\bar{X}}', t)$, that can be used to generate LES-type equations of motion is found when the minor variables are integrated out, as in

$$\tilde{\rho}'(\bar{\bar{X}}, t) = \int \mathrm{d}\gamma' |J(\bar{X}, \bar{\bar{X}})| e^{\hat{\bar{L}}t}\tilde{\rho}(\bar{\bar{X}}, 0), \tag{21}$$

where $\mathrm{d}\gamma'$ is the minor variables' volume element. Solving Eq. (21) serves the purpose of generating a density that can be used to obtain the LES equations of motion shown in Eq. (15). It does so by enforcing the holonomic constraints shown to result in LES.

3.3. Explanation of the "temperature disparity" problem

In the canonical ensemble, the average, $<q_i F_i>$, for a bound system can be written as

$$< q_i F_i >= -\frac{1}{Q}\int \mathrm{d}\Gamma \delta q_i q_i \frac{\partial H}{\partial q_i}\mathrm{e}^{-\beta H(q_i, \Gamma)}. \tag{22}$$

Here, Q is the partition function, q_i a general coordinate of the ith atom, and F_i is the force on the ith atom. The integral's exact value is $-1/\beta$, where β is $1/k_\mathrm{B}T$ (k_B is the Boltzmann constant). Therefore, after applying the virial theorem the average kinetic energy for C copies of a structure is simply

$$\bar{T} = -\frac{1}{2}\sum_{i=1}^{S}\sum_{k=1}^{C} < q_{i,k}\frac{\partial H}{\partial q_{i,k}} > -\frac{1}{2}\sum_{i=1}^{N}\sum_{k=1}^{C} < Q_{i,k}\frac{\partial H}{\partial Q_{i,k}} > = \frac{(SC + NC)}{2\beta}. \tag{23}$$

After making the transformation,

$$< Q_{i,k}\frac{\partial H}{\partial Q_{i,k}} > = < \frac{C}{n}\left(Q_i + \sum_{l=2}^{C} c_{l,k} Q'_{i,l}\right) \times \frac{C}{n^2}\left(\frac{\partial H}{\partial Q_i} + \sum_{m=2}^{C} c_{m,k}\frac{\partial H}{\partial Q'_{i,m}}\right) >, \tag{24}$$

and using the result in the virial theorem, the exact result $\bar{T} = (\mathrm{SC} + \mathrm{NC})/2\beta$ is generated.

However, if the minor variables are constrained to vanish, the average kinetic energy is $\bar{T} = (N + \mathrm{CS})/2\beta$. Consequently, the LES virial is incorrect because mean-field systems have fewer degrees of freedom [19], and this manifests itself as the temperature disparity problem.

3.4. Analysis of validity of the LES approximation

It should be noted that the work shown above presents a purely classical mechanical description of the approximations inherent in mean-field methods [19]. It also allows some comments to be made regarding the accuracy of mean-field trajectories. For example, mean-field methods can be viewed as a collection of separate systems coupled together through the requirement that the particles labeled as a "bath" are required to occupy the same positions in each system.

Knowing this allows one to consider the approximation's range of validity. From this work it should be expected that mean-field methods can accurately reproduce the correct classical trajectories when the force on the "minor" variables are small. Two conditions might give rise to this. The first occurs when the interactions between the copied particles and the bath particles are small, as might be expected in a low density simulation. The second condition is that the simulation is of short enough duration that the dispersal of the copies of the bath particles is not important. If either condition is true, the trajectory generated with a mean-field method would be expected to be accurate. Unfortunately, useful simulations are frequently impossible to perform under these conditions, because long trajectories need to be analyzed or because the system of interest is in a condensed phase. Our contribution described in the next section [24] is designed to alleviate this problem, and allow mean-field methods to be used to simulate condensed-phase systems more accurately.

4. Practical contributions

4.1. The EXACT approximation

4.1.1. Description of the algorithm

The main result from Eqs. (14) and (15) is that the LES approximation can only be accurately applied where the minor variables are small enough to be ignored. We developed a method that accounts for the minor variable's influence called the EXACT approximation [19,24]. Eq. (14) along with an initial holonomic constraint set to force the minor variables to vanish is the basis, but corrections are applied when the forces on the minor variable at each step are large when compared to the force on the corresponding major variable. The ratio $\phi_{i,l} = |\dot{P}'_{i,l}/\dot{P}_i|$, the force on the minor variable divided by the force on the major variable, is calculated and then compared with a user-defined tolerance, Φ. For any coordinate associated with a particle in which the ratio, $\phi_{i,l}$, is greater than the tolerance ($\Phi < \phi_{i,l}$), the particle is removed from the bath and treated exactly (as a copied particle) for a user-defined number of time steps, and then returned to the bath in a manner that ensures that the energy is conserved. As $\Phi \to 0$, conventional dynamics for an ensemble of system copies is produced; when $\Phi \to \infty$, LES results.

4.1.2. Calculation of the "minor" forces

It is necessary to calculate the force on the minor coordinates to use the EXACT approximation. This is accomplished economically by first writing the forces as derivatives of the potential energy:

$$\frac{\partial \tilde{H}}{\partial Q'_{i,m}} = \sum_{k=1}^{C} \left(\frac{\partial V_{B,k}}{\partial Q'_{i,m}} + \frac{\partial V_{X,k}}{\partial Q'_{i,m}} \right). \tag{25}$$

Here $V_{B,k}$ is the potential energy of the bath for copy k and $V_{X,k}$ is the interaction energy between the copied region and the bath for copy k.

We have earlier shown [24] that this can be rewritten as:

$$\frac{\partial \tilde{H}}{\partial Q'_{i,m}} = \frac{\partial V_B}{\partial Q_i} \left(1 + \sum_{k=2}^{C} c_{m,k} \right) + \frac{\partial V_{X,1}}{\partial Q_i} + \sum_{k=2}^{C} \frac{\partial V_{X,k}}{\partial Q_i} c_{m,k}. \tag{26}$$

Finally, applying Eqs. (13)–(26) yields:

$$\frac{\partial \tilde{H}}{\partial Q'_{i,m}} = \frac{\partial V_{X,1}}{\partial Q_i} + \sum_{k=2}^{C} \frac{\partial V_{X,k}}{\partial Q_i} c_{m,k}. \tag{27}$$

Eq. (27) is important because it shows that the forces on the minor variables can be calculated with trivial effort since $\partial V_{X,k}/\partial Q_i$, the forces between bath atoms and copies, is already calculated—it is the force on the corresponding major variable. This equation also requires that the transform coefficients be explicitly known, unlike in other mean-field methods. We have previously published a method to calculate them [24].

4.2. Extension of LES and EXACT to constant temperature ensembles

Because the Nosé–Hoover chain (NHC) method [35] has been shown to be a useful method for constant temperature simulations, we showed [24] how to implement this algorithm for the mean-field methods LES and EXACT. An extended system designed to act as a heat bath is appended to the original Hamiltonian of a system. Thus, the Hamiltonian for an NHC system of C non-interacting systems can be written as:

$$H_{\mathrm{NHC}} = H + \sum_{k=1}^{C} \left(\sum_{i=1}^{2} \frac{p^2_{\eta_{i,k}}}{2\mu_i} + k_{\mathrm{B}} T [D\eta_{1,k} + \eta_{2,k}] \right). \tag{28}$$

In this equation, H is given by Eq. (11), $p_{\eta_{i,k}}$ and $\eta_{i,k}$ are the momenta and "position" of the heat bath's chain particles, and μ_i is the "mass" of the chain particle. The subscripts

on $p_{\eta_{i,k}}$ indicate that this is the momentum of chain particle i and D is the number of particles to be copied. From the above Hamiltonian the forces on the major variables can be expressed as

$$\dot{P}_i = \frac{1}{C}\sum_{k=1}^{C}\dot{P}_{i,k} \tag{29}$$

which can be rewritten using Eqs. (12) and (13) as

$$\dot{P}_i = -\frac{1}{C}\left(\frac{\partial H_{\mathrm{NHC}}}{\partial \mathrm{Q}_i} + P_i\sum_{k=1}^{C}\frac{\partial H_{\mathrm{NHC}}}{\partial p_{\eta_{1,k}}} + \sum_{l=2}^{C}P'_{i,l}\sum_{k=1}^{C}c_{l,k}\frac{\partial H_{\mathrm{NHC}}}{\partial p_{\eta_{1,k}}}\right). \tag{30}$$

An assumption of both LES (and EXACT when the minor variables of the bath are ignored) is that the minor variables expressing the constant temperature constraint can be assumed to vanish, so the final term can be ignored. Thus, the first two terms comprise the force needed to ensure a constant temperature simulation. Doing a transformation similar to the one used to derive Eq. (30) gives the temperature controlled force on the minor variables, after applying the holonomic constraint shown to yield the LES equations of motion:

$$\dot{P}'_{i,l} = -\frac{1}{C}\left[\frac{\partial H_{\mathrm{NHC}}}{\partial Q'_{i,l}} + P_i\left(\frac{\partial H_{\mathrm{NHC}}}{\partial p_{\eta_{1,1}}} + \sum_{k=2}^{C}c_{l,k}\frac{\partial H_{\mathrm{NHC}}}{\partial \mathrm{p}_{\eta_{1,k}}}\right)\right]. \tag{31}$$

With these two equations, both LES and EXACT algorithms can be designed to simulate constant temperature trajectories. Similar logic can be used for barostats [1,36].

4.3. Tests of the EXACT approximation

4.3.1. Effect on simple pair distribution functions

Fig. 1 is a collection of pair distribution functions of a 64 atom cluster of argon [24]. The curve generated from purely LES data contains peaks located in the correct places, but they are too broad and short compared to the MD result. This is evidence that LES samples regions of space forbidden to particles in a conventional MD simulation. The EXACT approximation's results, however, scale between these two extremes. When the threshold is set with $\Phi = 0.8$, the simulation produces a taller and narrower initial peak. When the threshold is set with $\Phi = 0.3$, the simulation produces a result that is only slightly different from the conventional MD result. By carefully selecting the threshold, Φ, the simulation's sampling properties can range from conventional MD for C non-interacting copies to the mean-field (LES) limit.

4.3.2. Effect on cooling behavior of copied particles

The same Ar cluster described above was used to perform a cooling simulation using conventional MD, LES, and the constant energy implementation of the EXACT algorithm [24]. The simulation consisted of choosing a particle near the center of the cluster, and raising its kinetic energy to 300 K, and observing its cooling. Results are shown in Fig. 2, and demonstrate the effect of the "temperature-disparity

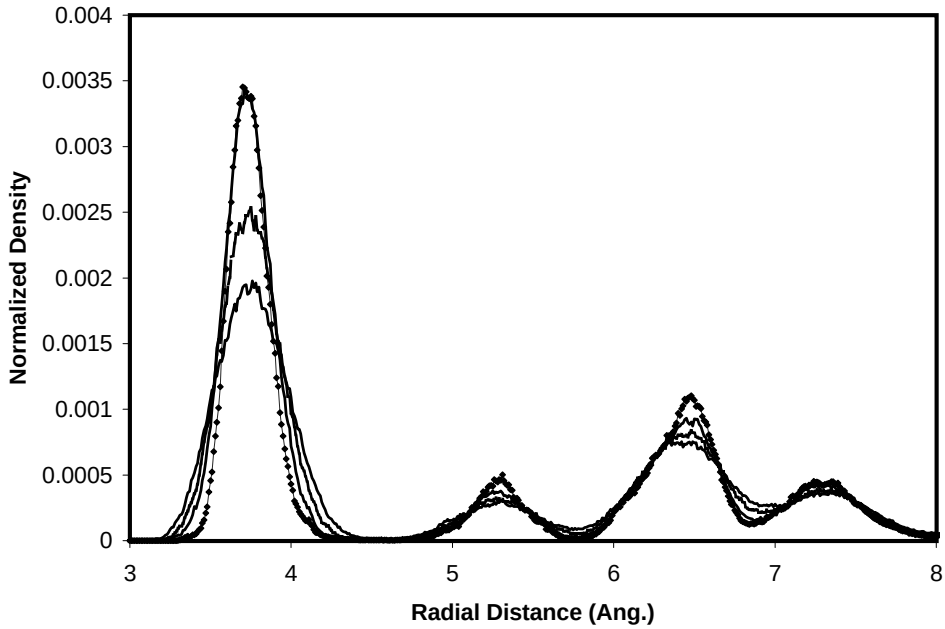

Fig. 1. Normalized density distribution for a single, copied Ar atom in a bath of Ar. Simulations correspond to conventional MD (highest, narrowest peak, marked by diamonds), LES (shortest, broadest peak), and intermediate cases, generated using the EXACT approximation. The EXACT approximation allows interpolation between conventional MD and the enhanced sampling of LES.

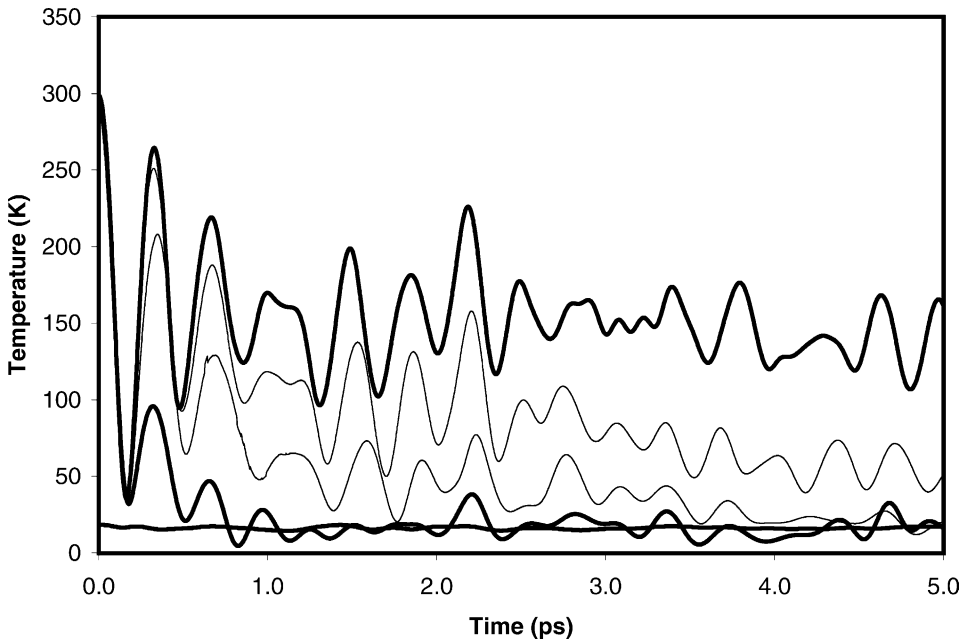

Fig. 2. Comparison of temperature vs. time for the bath of 64 Ar atoms (flattened curve near the bottom) with the temperature of one Ar atom in conventional MD (bold curve oscillating about the bath temperature), LES (bold curve at the top showing large oscillations), and intermediate cases, generated using the EXACT approximation. The EXACT approximation allows the hot Ar to cool to the temperature of the bath but LES does not.

problem." The problem has been described as the failure of a hot particle to cool properly during a mean-field simulation, and has been the focus of several theoretical investigations [19–22]. The crux of the problem is that although the Ar atom cools quickly to the bath temperature using conventional MD, LES simulations do not seem to demonstrate any cooling at all. Also shown in the figure are results generated by the EXACT approximation using thresholds $\Phi =$ 0.8 and 0.3. The 0.8 simulation improves the simulation, but not markedly. The $\Phi = 0.3$ simulation, though, relaxes to the correct temperature within 5 ps. This is a clear improvement over LES.

4.3.3. Effect on sampling a torsion angle of melatonin

Melatonin [37], a tryptophan derivative, is a molecule with seven flexible torsion angles, as shown in Fig. 3. There are three main sections of melatonin which include the 5-methoxy group, the indole ring, and the peptide-like side chain located on the C3 position of the ring.

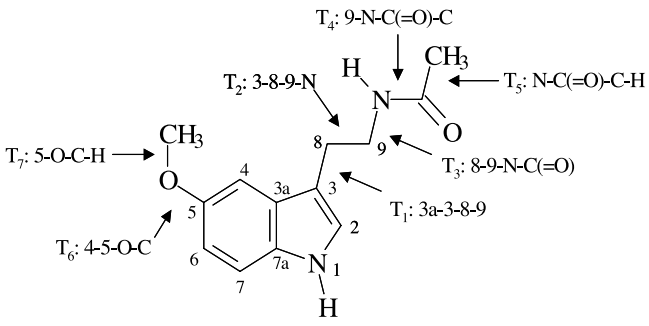

Fig. 3. Structure of melatonin, along with its standard numbering scheme. Interesting torsion angles are labeled as T_i.

The dihedral angles of the methoxy group and side chain determine the molecule's spatial conformation, while the indole ring remains essentially planar. The methoxy group has two torsion angles—one represents the nearly free rotation of the hydrogens around the methyl group (T_7) and the other determines the position of the methyl carbon group (T_6). The peptide-like side chain contains the balance of the flexible torsion angles, of which only T_1, T_2, and T_3 are both interesting and free to rotate, as T_4 is a peptide bond, and T_5 determines the conformation of a nearly free methyl rotor.

Melatonin proves an excellent test system to illustrate the ability of a method to enhance conformational sampling, as its T_1 torsion angle must overcome a relatively high energy barrier to rotate the side chain from one face of the indole ring to the other. As a consequence, T_1 is expected to be relatively fixed with an angle near $\pm 90^\circ$. It is around this torsion angle that the boundary is made between the copied region and the bath. Everything in the indole ring is designated bath, every atom in the sidechain is copied four times. Thus, the T_1 torsion angle is expected to enjoy the bulk of the enhanced sampling in EXACT simulations.

For this test, a gas phase conventional MD simulation of melatonin was performed as described in Section 2.1 to observe the degree of sampling normally found in molecular dynamics simulations of this molecule. This was followed by a gas phase EXACT simulation expected to benefit from enhanced sampling, using a threshold expected to be near the LES limit ($\Phi \to \infty$). Each of these tests were allowed 300 ps of simulation time.

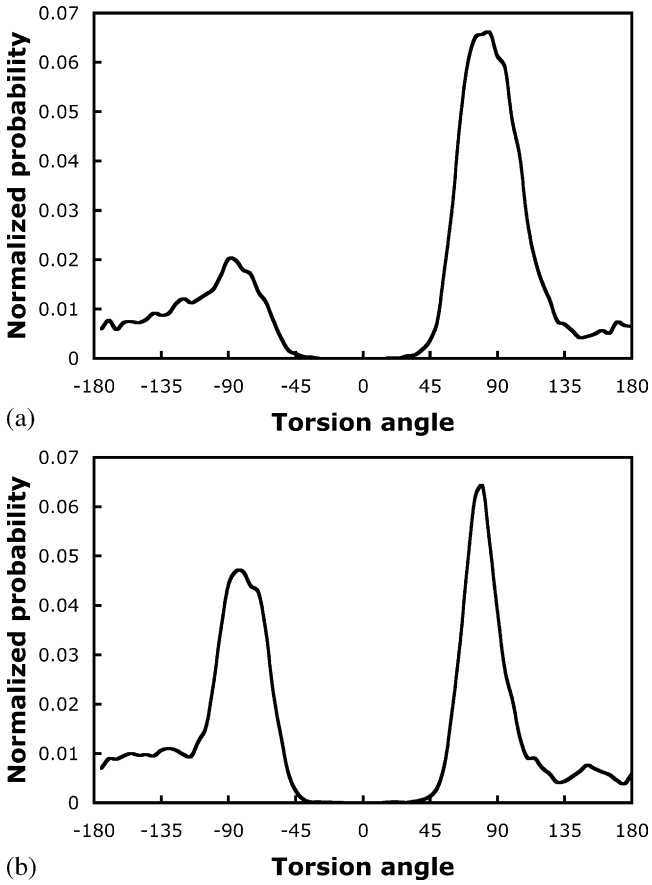

Fig. 4. (a) Normalized histogram recording the number of times melatonin's T1 torsional angle visited a given value during a 300 ps gas-phase conventional MD simulation. The asymmetry indicates preferential sampling on one side of the indole ring. (b) Normalized histogram recording the number of times melatonin's T1 torsional angle visited a given value during a 300 ps gas-phase EXACT simulation. More balanced sampling on both sides of the indole ring is achieved, as indicated by the more symmetrical distribution of torsion angles.

Histograms of the T_1 dihedral are presented for both the gas phase conventional MD and EXACT simulation of 300 ps in Fig. 4a and b, respectively. The conventional simulation does not provide balanced sampling on both faces of melatonin's indole ring (defined by $T_1 = \pm 90^\circ$). The positive dihedral region is greatly favored in the conventional simulation, because there was not adequate time for equal sampling to be achieved. For the EXACT simulation, both faces of melatonin's indole ring are sampled nearly equally. The centers of the peaks are at $\pm 90^\circ$, and are almost equal in area.

5. Conclusions

LES and other mean-field methods have been successfully used in a variety of applications [25–27,29–33] and have been the subject of a great deal of theoretical scrutiny [19–24]. Two new routes to obtain the mean-field equations of motion shown here add interesting insights to the understanding of mean-field methods. We show that mean-field methods can be viewed as a type of constrained dynamics [19]. These constraints reduce the number of degrees of freedom in the system, which affects the temperature equilibration properties of the system. Further, the classical mechanical view demonstrates the degree of approximation in mean-field methods and indicates that these methods can be truly accurate under certain conditions. Finally, we use this knowledge to create a new method, called the EXACT approximation [24], that is designed to allow a researcher to decide the degree of accuracy or enhanced sampling a simulation requires.

Acknowledgements

This work was supported by the Oklahoma Center for the Advancement of Science and Technology through OCAST award no. HR01-148, by the Chemical Sciences, Geosciences and Bioscience Division, Office of Basic Energy Sciences, Office of Science, US Department of Energy through Grant no. DE-FG03-01ER15164, by the NSF/NRAC through super-computer time award no. MCA96-N019, and by supercomputer time from the Oklahoma Super-computing Center for Education and Research (OSCER). We are grateful to the University of Oklahoma's Graduate College for supporting CAH through an OU Alumni Graduate Fellowship and to the US Air Force and the US Air Force Academy for supporting JC. RW is grateful to Peter Kollman's research group, especially Prof. Carlos Simmerling, for their hospitality and helpful discussions during his sabbatical, when this project was conceived. CAH would like to thank Prof. Kieran Mullen for many helpful, stimulating discussions, and we would like to thank Dr. Henry Neeman for programming advice. We would also like to thank Drs. Gina M. Florio and Timothy S. Zwier for proofs of their melatonin paper. (J. Phys. Chem. A, 107 (7), 974–983, 2003. 10.1021/jp027053i S1089-5639(02)07053—Solvation of a Flexible Biomolecule in the Gas Phase: The Ultraviolet and Infrared Spectroscopy of Melatonin–Water Clusters).

Appendix A. Details of melatonin simulation

Simulations on melatonin were performed to ensure that EXACT could properly allow for the enhanced sampling that mean-field methods allow [28]. The melatonin simulations used the AMBER94 all-atom force field [38] with charges obtained using electrostatic fitting from electron densities obtained using Gaussian 94 [39], at the B3LYP/6-31G* level of theory. Gas-phase simulations of melatonin were performed using our own implementation, using the molecular modeling toolkit (MMTK) [40] to prepare the initial conditions. Temperature was controlled with the Nosé–Hoover chain algorithm [35]. Covalent bond distances involving

hydrogen were constrained using the SHAKE algorithm [41] for conventional MD, and with a slightly modified SHAKE (to account for the common bath) in the EXACT approximation simulations. Atoms involving the side-chain of melatonin (which is the functional group attached to carbon 3 in the indole ring, as depicted in Fig. 3) were copied four times in the EXACT simulations.

References

[1] M.E. Tuckerman, G.J. Martyna, Understanding modern molecular dynamics: techniques and applications, J. Phys. Chem. B 3 (2000) 159–178, and refs. therein.

[2] P. Kollman, Free-energy calculations—applications to chemical and biochemical phenomena, Chem. Rev. 3 (1993) 2395–2417, and refs. therein.

[3] M.P. Allen, D.J. Tildesley, Computer Simulations of Liquids, Clarendon Press, Oxford, 1987.

[4] R.P. Feynman, Statistical Mechanics: A Set of Lectures, W.A. Benjamin, Inc., Reading, MA, 1972.

[5] N. Makri, Time-dependent quantum methods for large systems, Annu. Rev. Phys. Chem. 3 (1999) 167–191.

[6] S. Kirkpatrick, C.D. Gelatt Jr., M.P. Vecchi, Optimization by simulated annealing, Science 3 (1983) 671–680.

[7] L. Piela, J. Kostrowicki, H.A. Scheraga, The multiple-minima problem in the conformational-analysis of molecules—deformation of the potential-energy hypersurface by the diffusion equation method, J. Phys. Chem. 3 (1989) 3339–3346.

[8] J. Ma, D. Hsu, J.E. Straub, Approximate solutions of the classical Liouville equation using Gaussian phase packet dynamics—applications to enhanced equilibrium averaging and global optimization, J. Chem. Phys. 3 (1993) 4024–4035.

[9] B.J. Berne, J.E. Straub, Novel methods of sampling phase space in the simulation of biological systems, Curr. Op. Struct. Biol. 3 (1997) 181–189.

[10] P. Koehl, M. Delarue, Mean-field minimization methods for biological macro-molecules, Curr. Op. Struct. Biol. 3 (1996) 222–226.

[11] T. Huber, W.F. van Gunsteren, SWARM-MD: searching conformation space by cooperative molecular dynamics, J. Phys. Chem. A 3 (1998) 5937–5943.

[12] C. Tsallis, D.A. Stariolo, Generalized simulated annealing, Phys. A 3 (1996) 395–406.

[13] T. Huber, A.E. Torda, W.F. van Gunsteren, Optimization methods for conformational sampling using a Boltzmann-weighted mean field approach, Biopolymers 3 (1996) 103–114.

[14] C. Maranas, C.A. Floudas, A deterministic global optimization approach for molecular-structure determination, J. Chem. Phys. 3 (1994) 1247–1261.

[15] G.A. Huber, J.A. McCammon, Weighted-ensemble simulated annealing: faster optimiaztion on heirachical energy surfaces, Phys. Rev. E 3 (1997) 4822–4825.

[16] Z. Zhu, M.E. Tuckerman, S.O. Samuelson, G.J. Martyna, Using novel variable transformations to enhance conformational sampling in molecular dynamics, Phys. Rev. Lett. 3 (2002) 100201.

[17] R.B. Gerber, V. Buch, M.A. Ratner, Time-dependent self-consistent field approximation for intramolecular energy-transfer. 1. Formulation and application to dissociation of van der Waals molecules, J. Chem. Phys. 3 (1982) 3022–3030.

[18] R. Elber, M. Karplus, Enhanced sampling in molecular dynamics—use of the time-dependent Hartree approximation for a simulation of carbon monoxide diffusion through myoglobin, J. Am. Chem. Soc. 3 (1990) 9161–9175.

[19] C.A. Hixson, R.A. Wheeler, Rigorous classical mechanical derivation of a multiple-copy algorithm for sampling statistical mechanical ensembles, Phys. Rev. E 3 (2001) 026701.

[20] J.E. Straub, M. Karplus, Energy equipartitioning in the classical time-dependent Hartree approximation, J. Chem. Phys. 3 (1991) 6737–6739.

[21] A. Ulitsky, R. Elber, The thermal-equilibrium aspects of the time-dependent Hartree and the locally enhanced sampling approximations—formal properties, a correction, and computational examples for rare-gas clusters, J. Chem. Phys. 3 (1993) 3380–3388.

[22] W. Zheng, Q. Zheng, An analytical derivation of the locally enhanced sampling approximation, J. Chem. Phys. 3 (1997) 1191–1194.

[23] C. Stultz, M. Karplus, On the potential surface of the locally enhanced sampling approximation, J. Chem. Phys. 3 (1998) 8809–8815.

[24] C.A. Hixson, R.A. Wheeler, Practical multiple-copy methods for sampling classical statistical mechanical ensembles, Chem. Phys. Lett. (2004) in press.

[25] C. Simmerling, et al., Combining MONSSTER and LES/PME to predict protein structure from amino acid sequence: application to the small protein CMTI-1, J. Am. Chem. Soc. 3 (2000) 8392–8402.

[26] A. Roitberg, R. Elber, Modeling side-chains in peptides and proteins—application of the locally enhanced sampling and the simulated annealing methods to finnd minimum energy conformations, J. Chem. Phys. 3 (1991) 9277–9287.

[27] C. Simmerling, T. Fox, P.A. Kollman, Use of locally enhanced sampling in free energy calculations: testing and application to the alpha beta anomerization of glucose, J. Am. Chem. Soc. 3 (1998) 5771–5782.

[28] J. Chen, Masters Thesis, University of Oklahoma, 2003.

[29] M.L. Quillin, et al., Structural and functional-effects of apolar mutations of the distal valine in myoglobin, J. Mol. Biol. 3 (1995) 416–436.

[30] A. Ulitsky, R. Elber, Application of the locally enhanced sampling (LES) and a mean-field with a binary collision correction (cLES) to the simulation of Ar diffusion and NO recombination in myoglobin, J. Phys. Chem. 3 (1994) 1034–1043.

[31] A. Miranker, M. Karplus, Functionality maps of binding-site—a multiple copy simultaneous search method, Proteins Struct. Funct. Gen. 3 (1991) 29–34.

[32] H.A. Carlson, K.M. Masukawa, J.A. McCammon, Method for including the dynamic fluctuations of a protein in computer-aided drug design, J. Phys. Chem. A 3 (1999) 10213–10219.

[33] A. Caflisch, A. Miranker, M. Karplus, Multiple copy simultaneous search and construction of ligands in binding sites—application to inhibitors of HIV-1 aspartic proteinase, J. Med. Chem. 3 (1993) 2142–2167.

[34] R.L. Liboff, Kinetic Theory: Classical, Quantum, and Relativistic Descriptions, Wiley, New York, 1998;
I. Prigogine, Non-Equilibrium Statistical Mechanics, Interscience, New York, 1962.

[35] G.J. Martyna, M.L. Klein, M. Tuckerman, Nose–Hoover chains—the canonical ensemble via continuous dynamics, J. Chem. Phys. 3 (1992) 2635–2643.

[36] G.J. Martyna, D.J. Tobias, M.L. Klein, Constant-pressure molecular dynamics algorithms, J. Chem. Phys. 3 (1994) 4177–4189.

[37] M. Marco, et al., Melatonin Curr. Med. Chem. 3 (1999) 501–518.

[38] W.D. Cornell, et al., A second generation force-field for the simulation of proteins, nucleic-acids, and organic-molecules, J. Am. Chem. Soc. 3 (1995) 5179–5197.

[39] M.J. Frisch, et al., Gaussian 94 (Revision B.2), Gaussian, Inc., Pittsburgh, PA, 1995.

[40] K. Hinsen, The molecular modeling toolkit: a new approach to molecular simulations, J. Comp. Chem. 3 (2000) 79–85.

[41] B.J. Palmer, Direct application of SHAKE to the velocity-Verlet algorithm, J. Comp. Phys. 3 (1993) 470–472.