Maintain rigid structures in Verlet based Cartesian molecular dynamics simulations

Peng Tao, Xiongwu Wu, and Bernard R. Brooks

Laboratory of Computational Biology, National Heart, Lung and Blood Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

An algorithm is presented to maintain rigid structures in Verlet based Cartesian molecular dynamics (MD) simulations. After each unconstrained MD step, the coordinates of selected particles are corrected to maintain rigid structures through an iterative procedure of rotation matrix computation. This algorithm, named as SHAPE and implemented in CHARMM program suite, avoids the calculations of Lagrange multipliers, so that the complexity of computation does not increase with the number of particles in a rigid structure. The implementation of this algorithm does not require significant modification of propagation integrator, and can be plugged into any Cartesian based MD integration scheme. A unique feature of the SHAPE method is that it is interchangeable with SHAKE for any object that can be constrained as a rigid structure using multiple SHAKE constraints. Unlike SHAKE, the SHAPE method can be applied to large linear (with three or more centers) and planar (with four or more centers) rigid bodies. Numerical tests with four model systems including two proteins demonstrate that the accuracy and reliability of the SHAPE method are comparable to the SHAKE method, but with much more applicability and efficiency. [http://dx.doi.org/10.1063/1.4756796]

I. INTRODUCTION

Rigid body molecular dynamics is of increasing importance in computational chemistry and physics.[1–4] Many rigid body dynamics methods have been developed.[5–11] Generally, there are three categories of rigid body dynamics methods (Fig. 1). In the first category, the non-constrained dynamics integration is applied in each step before applying constrained corrections. Two widely applied methods, SHAKE[5] and RATTLE,[6] belong to this category. In the second category, the rigid body constraint forces are calculated to propagate each particle in rigid bodies.[7] In the third category, the rigid body motion is divided into the translation of center of mass and rotation about the center of mass. The translation and rotation motions of rigid body are propagated based on the force on the center of mass and torque about the center of mass.[12]

Each of the reported rigid body algorithms has some undesirable features, limitations, or restrictions. In many applications, certain parts of systems need to maintain a rigid structure, and should be treated as a single rigid body. For a nonlinear rigid structure containing M particles, $3M - 6$ Lagrange equations are typically solved to implement rigid structure constraints. Therefore, this rigid structure constraint problem is not convenient with methods described above for large M. In this work, we present an efficient algorithm, named as SHAPE, for simulation of rigid structures composed by an arbitrary number of particles. This rigid structure integration method is fully consistent and interchangeable with the SHAKE method.[5] It can also be combined with SHAKE in such a way rigid body constraints are solved consistently with SHAKE for atoms that involved in more than one type of constraints. For constant pressure simulation, this method can use the same virial correction scheme as in SHAKE.[13]

II. THEORY

Let us assume that M atoms compose a rigid structure in an N atoms system. During each MD step, the coordinates of whole system with N atoms are integrated according to the equation of motion. The coordinates of the M atoms are then corrected to maintain the rigid structure.

For an MD integration at time t with Δt as step size and without any constraint, the coordinates of the M atoms change from $\mathbf{r}_i(t)$ to $\mathbf{r}_i^{non}(t + \Delta t)$. Either with or without the rigid structure constraint, the structure's momentum and angular momentum from t to $t + \Delta t$ should be the same, which lays the theoretical ground for maintaining rigid structures.

The centers of mass (COM) for two sets of coordinates are

$$\mathbf{r}_{COM}(t) = \frac{\sum_{i=1}^{M} m_i \mathbf{r}_i(t)}{\sum_{i=1}^{M} m_i}, \tag{1}$$

$$\mathbf{r}_{COM}^{non}(t + \Delta t) = \frac{\sum_{i=1}^{M} m_i \mathbf{r}_i^{non}(t + \Delta t)}{\sum_{i=1}^{M} m_i}, \tag{2}$$

where m_i is the mass of atom i. The rigid structure motion of M atoms can be separated into translational and rotational parts. The translational motion is represented by COM of these M atoms. The more difficult part is the rotational motion, which can be treated in body-fixed local coordinate with its origin at the COM of the M atoms and XYZ orientation
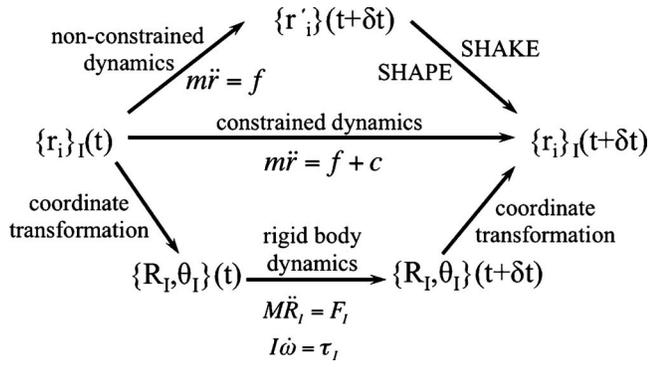
FIG. 1. Three propagation schemes for rigid structures in dynamics. In SHAKE and SHAPE schemes, the unconstrained positions of individual particles are iteratively corrected to reach a rigid structure. In constrained dynamics, the constrained forces, c, are calculated to maintain a rigid structure. Alternatively, the Cartesian coordinates of a rigid body can be represented by position and rotation variables, which can be used to propagate the rigid structure to the next time step.

the same as the global coordinate. In other words, the body-fixed local coordinate for the rigid body is constructed by transforming the origin of global Cartesian coordinate to the COM without any rotation. Therefore, the coordinates of the M atoms at time t and $t + \Delta t$ in body-fixed local coordinate can be calculated with Eqs. (3) and (4), respectively,

$$\mathbf{r}_i^b(t) = \mathbf{r}_i(t) - \mathbf{r}_{COM}(t), \tag{3}$$

$$\mathbf{r}_i^{non, \, b}(t + \Delta t) = \mathbf{r}_i^{non}(t + \Delta t) - \mathbf{r}_{COM}^{non}(t + \Delta t). \tag{4}$$

The angular momentum $\mathbf{L}^{non}$ of the M atoms at time $t + \Delta t/2$ without rigid body constraints is

$$
\begin{aligned}
\mathbf{L}^{non} & \left(t + \frac{\Delta t}{2} \right) \\
&= \sum_{i=1}^{M} \mathbf{r}_i^{non, \, b} \left(t + \frac{\Delta t}{2} \right) \times \mathbf{p}_i^{non, \, b} \left(t + \frac{\Delta t}{2} \right) \\
&= \sum_{i=1}^{M} \mathbf{r}_i^{non, \, b} \left(t + \frac{\Delta t}{2} \right) \times \left(m_i \mathbf{v}_i^{non, \, b} \left(t + \frac{\Delta t}{2} \right) \right) \\
&= \sum_{i=1}^{M} \mathbf{r}_i^{non, \, b} \left(t + \frac{\Delta t}{2} \right) \times \left(m_i \frac{\mathbf{r}_i^{non, \, b}(t + \Delta t) - \mathbf{r}_i^b(t)}{\Delta t} \right),
\end{aligned}
\tag{5}
$$

where $\mathbf{p}_i^b$ is the momentum, and $\mathbf{v}_i^b$ is the velocity for atom i in body-fixed local coordinate at time $t + \Delta t/2$. The velocities in body-fixed local coordinate are finite difference of coordinates calculated from Eqs. (3) and (4).

The rigid body motion of M atoms is achieved by correcting the coordinates of the M atoms at time $t + \Delta t$ from $\mathbf{r}_i^{non, \, b}(t + \Delta t)$, to maintain the rigid structure, and to have the angular momentum same to $\mathbf{L}^{non}(t + \Delta t/2)$. The coordinate correction is calculated in the body-fixed local coordinate through an iterative process described as the following.

Assume that the moment of inertia $\mathbf{I}$, a 3×3 matrix, is for M atoms in body-fixed local coordinates $\mathbf{r}_i^b(t)$:

$$
\mathbf{I} = \begin{bmatrix} I_{xx} & I_{xy} & I_{yz} \\ I_{yx} & I_{yy} & I_{yz} \\ I_{zx} & I_{zy} & I_{zz} \end{bmatrix}. \tag{6}
$$

The angular momentum of rigid structure composed of the M atoms can be calculated from $\mathbf{I}$ and angular velocity vector $\omega^{rig}(t + \Delta t/2)$ as

$$\mathbf{L}^{rig} \left(t + \frac{\Delta t}{2} \right) = \mathbf{I} \cdot \omega^{rig} \left(t + \frac{\Delta t}{2} \right). \tag{7}$$

ω is a three components vector, $\omega = (\omega_x, \omega_y, \omega_z)$. Based on the equality of angular momentum, we have

$$\mathbf{L}^{non} \left(t + \frac{\Delta t}{2} \right) = \mathbf{L}^{rig} \left(t + \frac{\Delta t}{2} \right). \tag{8}$$

Plug Eqs. (5) and (7) to Eq. (8), we have

$$
\begin{aligned}
\mathbf{I} \cdot \omega^{rig,(1)} \left(t + \frac{\Delta t}{2} \right) &= \sum_{i=1}^{M} \mathbf{r}_i^{non, \, b} \left(t + \frac{\Delta t}{2} \right) \\
&\times \left(m_i \frac{\mathbf{r}_i^{non, \, b}(t + \Delta t) - \mathbf{r}_i^b(t)}{\Delta t} \right),
\end{aligned}
\tag{9}
$$

where the superscript (1) indicates the first estimation. Rearrangement of Eq. (9) with inverse of $\mathbf{I}$ gives us $\omega^{rig,(1)} (t + \Delta t/2)$ as

$$
\begin{aligned}
\omega^{rig,(1)} \left(t + \frac{\Delta t}{2} \right) &= \mathbf{I}^{-1} \cdot \sum_{i=1}^{M} \mathbf{r}_i^{non, \, b} \left(t + \frac{\Delta t}{2} \right) \\
&\times \left(m_i \frac{\mathbf{r}_i^{non, \, b}(t + \Delta t) - \mathbf{r}_i^b(t)}{\Delta t} \right). \tag{10}
\end{aligned}
$$

Once vector ω is obtained, a skew-symmetric matrix $\hat{\omega}$ is constructed as

$$
\hat{\omega} = \begin{bmatrix} 0 & -\omega_z & \omega_y \\ \omega_z & 0 & -\omega_x \\ -\omega_y & \omega_x & 0 \end{bmatrix}. \tag{11}
$$

The angle θ undertaken by the rigid body in time interval Δt is $\|\omega\| \Delta t$. The corresponding skew-symmetric matrix $\hat{\theta}$ is $\hat{\omega} \Delta t$. The rotation matrix corresponding to the angle θ can be expressed as matrix exponential

$$\mathbf{R} = e^{\hat{\theta}}. \tag{12}$$

Following the Rodrigues' formula,[14] the matrix exponential is given by

$$e^{\hat{\theta}} = \mathbf{1} + \frac{\hat{\theta}}{\|\theta\|} \sin(\|\theta\|) + \frac{\hat{\theta}^2}{\|\theta\|^2} (1 - \cos(\|\theta\|)), \tag{13}$$

where $\mathbf{1}$ is the identity matrix, $\|\theta\|$ is the Euclidean norm of vector θ, and also is the rotational angle undertaken by the rigid structure. Equation (13) faces the problem of numerical

instability when $\|\theta\|$ is small. In such situation, the Taylor expansion is applied to construct rotation matrix

$$e^{\hat{\theta}} = \mathbf{1} + \left(1 - \frac{\|\theta\|^2}{3!} + \frac{\|\theta\|^4}{5!} - \cdots\right)\hat{\theta}$$

$$+ \left(\frac{1}{2!} - \frac{\|\theta\|^2}{4!} + \frac{\|\theta\|^4}{6!} \cdots\right)\hat{\theta}^2. \tag{14}$$

After obtaining rotation matrix $\mathbf{R}$, the body-fixed local coordinate of the M atoms after the current MD step are updated through rotational operation on the coordinates in body-fixed local coordinate at time t

$$\mathbf{r}_i^{rig,\,b(n)}(t + \Delta t) = \mathbf{R}^{(n)} \cdot \mathbf{r}_i^b(t), \tag{15}$$

where the superscript (n) specifies the number of iteration. For the first iteration, $n = 1$. It should be noted that the coordinates $\mathbf{r}_i^{rig,\,b(n)}(t + \Delta t)$ is in the body-fixed local coordinates for rigid structure at time $t + \Delta t$, and $\mathbf{r}_i^b(t)$ is in the body-fixed local coordinates at time t. Therefore, the rotational operation $\mathbf{R}^{(n)}$ was applied on M atoms around the center of mass.

From the updated coordinates, the angular momentum in the iteration, $\mathbf{L}^{rig,(n)}$, is calculated as

$$\mathbf{L}^{rig,\,(n)}\left(t + \frac{\Delta t}{2}\right)$$

$$= \sum_{i=1}^{M} \mathbf{r}_i^{rig,\,b,(n)}\left(t + \frac{\Delta t}{2}\right) \times \left(m_i \dot{\mathbf{r}}_i^{rig,\,b,(n)}\left(t + \frac{\Delta t}{2}\right)\right)$$

$$= \sum_{i=1}^{M} \mathbf{r}_i^{rig,\,b,(n)}\left(t + \frac{\Delta t}{2}\right) \times \left(m_i \frac{\mathbf{r}_i^{rig,\,b,(n)}(t + \Delta t) - \mathbf{r}_i^b(t)}{\Delta t}\right), \tag{16}$$

where $\dot{\mathbf{r}}_i^{rig,\,b,(n)}$ is the velocity of atom i calculated from the updated coordinate in body-fixed local coordinate during the iteration.

If $\mathbf{L}^{rig,(n)}(t + \Delta t/2)$ differs from $\mathbf{L}^{non}(t + \Delta t/2)$ by amount less than a predefined tolerance, the procedure is considered converged. Otherwise, angular velocity vector $\omega^{rig,(n)}(t + \Delta t/2)$ is updated to the next iteration.

Since the coordinates at time $t + \Delta t/2$ are not readily available from the simulation, it can be approximated using the square root of rotation matrix, $\mathbf{R}^{(n)}$,

$$\mathbf{r}_i^{rig,\,b,(n)}\left(t + \frac{\Delta t}{2}\right) = (\mathbf{R}^{(n)})^{\frac{1}{2}}\mathbf{r}_i^b(t). \tag{17}$$

Consequently, the angular momentum for both non-rigid and rigid structure can be computed as the following two equations:

$$\mathbf{L}^{non}\left(t + \frac{\Delta t}{2}\right)$$

$$= \sum_{i=1}^{M} \mathbf{r}_i^{non,\,b}\left(t + \frac{\Delta t}{2}\right) \times \left(m_i \frac{\mathbf{r}_i^{non,\,b}(t + \Delta t) - \mathbf{r}_i^b(t)}{\Delta t}\right)$$

$$\approx \sum_{i=1}^{M} (\mathbf{R}^{(n)})^{\frac{1}{2}} \cdot \mathbf{r}_i^{non,\,b}(t) \times \left(m_i \frac{\mathbf{r}_i^{non,\,b}(t + \Delta t) - \mathbf{r}_i^b(t)}{\Delta t}\right)$$

$$\approx (\mathbf{R}^{(n)})^{\frac{1}{2}} \cdot \sum_{i=1}^{M} \mathbf{r}_i^{non,\,b}(t) \times \left(m_i \frac{\mathbf{r}_i^{non,\,b}(t + \Delta t) - \mathbf{r}_i^b(t)}{\Delta t}\right)$$

$$= (\mathbf{R}^{(n)})^{\frac{1}{2}} \cdot \mathbf{L}'^{non} \tag{18}$$

$$\mathbf{L}^{rig,(n)}\left(t + \frac{\Delta t}{2}\right)$$

$$= \sum_{i=1}^{M} \mathbf{r}_i^{rig,(n),\,b}\left(t + \frac{\Delta t}{2}\right) \times \left(m_i \dot{\mathbf{r}}_i^{rig,(n),\,b}\left(t + \frac{\Delta t}{2}\right)\right)$$

$$= \sum_{i=1}^{M} (\mathbf{R}^{(n)})^{\frac{1}{2}} \cdot \mathbf{r}_i^{rig,(n),\,b}(t) \times \left(m_i \frac{\mathbf{r}_i^{rig,(n),\,b}(t + \Delta t) - \mathbf{r}_i^b(t)}{\Delta t}\right)$$

$$\approx (\mathbf{R}^{(n)})^{\frac{1}{2}} \cdot \sum_{i=1}^{M} \mathbf{r}_i^{rig,(n),\,b}(t) \times \left(m_i \frac{\mathbf{r}_i^{rig,(n),\,b}(t + \Delta t) - \mathbf{r}_i^b(t)}{\Delta t}\right)$$

$$= (\mathbf{R}^{(n)})^{\frac{1}{2}} \cdot \mathbf{L}'^{rig,(n)} \tag{19}$$

The updated angular velocity vector $\omega^{rig,(n+1)}(t + \Delta t/2)$ is expected to satisfy

$$\mathbf{I} \cdot \omega^{rig,(n+1)}\left(t + \frac{\Delta t}{2}\right) = \mathbf{L}^{non}\left(t + \frac{\Delta t}{2}\right). \tag{20}$$

Combining Eqs. (20) and (7), we have

$$\mathbf{I} \cdot \omega^{rig,(n+1)}\left(t + \frac{\Delta t}{2}\right)$$

$$= \mathbf{I} \cdot \omega^{rig,(n)}\left(t + \frac{\Delta t}{2}\right)$$

$$+ \mathbf{L}^{non}\left(t + \frac{\Delta t}{2}\right) - \mathbf{L}^{rig,(n)}\left(t + \frac{\Delta t}{2}\right). \tag{21}$$

Multiply both sides of Eq. (21) by the inverse of moment of inertia, $\mathbf{I}^{-1}$, and plug Eqs. (18) and (19) into Eq. (21) with rearrangement we have

$$\omega^{rig,(n+1)}\left(t + \frac{\Delta t}{2}\right) = \omega^{rig,(n)}\left(t + \frac{\Delta t}{2}\right)$$

$$+ \mathbf{I}^{-1} \cdot (\mathbf{R}^{(n)})^{\frac{1}{2}} \cdot \left(\mathbf{L}'^{non} - \mathbf{L}'^{rig,(n)}\right). \tag{22}$$

Since $\mathbf{R}^{(n)}$ is close to the identity matrix for small time step Δt, we can write

$$\mathbf{R}^{(n)} = \mathbf{1} + \delta\mathbf{R}, \tag{23}$$

where $\delta\mathbf{R} = \mathbf{R}^{(n)} - \mathbf{1}$ is small. Hence

$$(\mathbf{R}^{(n)})^{\frac{1}{2}} = \mathbf{1} + \frac{\delta\mathbf{R}}{2} + O\left((\delta\mathbf{R})^2\right)$$

$$= \mathbf{1} + \frac{\mathbf{R}^{(n)} - \mathbf{1}}{2} + O\left((\delta\mathbf{R})^2\right)$$

$$= \frac{\mathbf{1} + \mathbf{R}^{(n)}}{2} + O\left((\delta\mathbf{R})^2\right). \tag{24}$$

By neglecting the quadratic term, we have approximation of square root of matrix $\boldsymbol{R}^{(n)}$ as

$$(\boldsymbol{R}^{(n)})^{\frac{1}{2}} \approx \frac{1}{2}(\boldsymbol{1} + \boldsymbol{R}^{(n)}). \tag{25}$$

Plug Eq. (25) into Eq. (22), we have updating scheme for angular velocity as

$$\omega^{rig,(n+1)}\left(t + \frac{\Delta t}{2}\right) = \omega^{rig,(n)}\left(t + \frac{\Delta t}{2}\right)$$
$$+ \boldsymbol{I}^{-1} \cdot \frac{(\boldsymbol{1} + \boldsymbol{R}^{(n)})}{2} \cdot (\boldsymbol{L}'^{non} - \boldsymbol{L}'^{rig,(n)}), \tag{26}$$

where $\omega^{rig,(n+1)}$ is the updated angular velocity vector for the $(n + 1)$th iteration.

Using the updated angular velocity vector $\omega^{rig,(n+1)}$, the procedure starting from Eq. (10) is repeated until the angular momentum $\boldsymbol{L}^{rig,(n)}(t + \Delta t/2)$ converges to $\boldsymbol{L}^{non}(t + \Delta t/2)$. Typically, three iterations ($n = 3$) are sufficient to converge to a double precision accuracy. This iterative approach is far more computationally efficient than any analytic approach.

After convergence, the coordinates in global reference for rigid body M atoms are calculated as

$$\mathbf{r}_i^{rig,\,(n)}(t + \Delta t) = \mathbf{r}_i^{rig,\,b(n)}(t + \Delta t) + \mathbf{r}_{COM}^{non}(t + \Delta t). \tag{27}$$

The final coordinates $\mathbf{r}_i^{rig,\,(n)}(t + \Delta t)$ are the rigid structure coordinates for the desired MD trajectory.

The procedure described above iteratively solves Eq. (8), or more explicitly, the following equation:

$$\sum_{i=1}^{M} \mathbf{r}_i^b\left(t + \frac{\Delta t}{2}\right) \times \left(m_i \frac{\mathbf{r}_i^{non,\,b}(t + \Delta t) - \mathbf{r}_i^b(t)}{\Delta t}\right)$$
$$= \sum_{i=1}^{M} \mathbf{r}_i^b\left(t + \frac{\Delta t}{2}\right) \times \left(m_i \frac{\mathbf{r}_i^{rig,\,b}(t + \Delta t) - \mathbf{r}_i^b(t)}{\Delta t}\right), \tag{28}$$

where $\mathbf{r}_i^{non,\,b}(t + \Delta t)$ and $\mathbf{r}_i^{rig,\,b}(t + \Delta t)$ are the body-fixed local coordinates without and with rigid body constraints at time $t + \Delta t$, respectively.

One important aspect of this rigid body integration method is that different rigid bodies may share one common atomic center. A rigid body may also share atoms that are involved in SHAKE constraints. This is possible because, as with SHAKE, applying the constraint to partial forces in multiple steps yields the same solution as applying the full force once, to within numerical precision. When a given atom is involved in two or more SHAKE constraints, the constraints are iterated in a cyclic fashion until convergence. The same approach works for SHAPE, and if an atom is involved in multiple SHAPE constraints, these must also be applied in a cyclic fashion until convergence. As with SHAKE, each application of the constraint conserves angular and linear momentum in a holonomic fashion. An NVE simulation converged with this approach would also approximately conserve energy. An efficient implementation of this approach would converge both SHAKE and SHAPE constraints simultaneously within the same outer iteration loop. A prospective use for this would be in making peptide bonds rigid and planar in a protein. In such

application, each C_α would be involved in two constraints that would need to be solved to consistency. A test case of such simulation is presented in this paper.

When using this method on a large-scale parallel machine where different atom centers are integrated on different processors, large rigid bodies may span across more than one processor. It is necessary to communicate components of $\mathbf{r}_{COM}^{non}$ and $\boldsymbol{L}^{non}$ at the beginning of the cycle and distribute the $\boldsymbol{R}$ matrix at the end. It is never necessary to distribute coordinates, since each processor can apply the final $\boldsymbol{R}$ matrix to atoms in each rigid body it contains. Thus the method can be efficiently applied even if whole proteins are made rigid.

The SHAPE rigid structure algorithm is implemented in CHARMM[15] program suite within the SHAPE module. The key subroutine of this algorithm is included in the supplementary material.[16]

The SHAPE method is based on the conservation of linear and angular momentum. The use of body-fixed local coordinate system guarantees the conservation of linear momentum after rigid body correction of each MD step. Therefore, the major error of SHAPE rigid body integration method comes from the angular momentum calculation. Through an iterative approach, a matrix is determined for rigid structure rotation to conserve the angular momentum obtained through non-rigid body integration. When using the leapfrog integration scheme, once a convergence is reached, we have

$$\boldsymbol{L}^{rig}\left(t + \frac{\Delta t}{2}\right) = \boldsymbol{L}^{non}\left(t + \frac{\Delta t}{2}\right). \tag{29}$$

Therefore, the accuracy of integration is determined by the angular momentum.

$$\boldsymbol{L}^{rig}\left(t + \frac{\Delta t}{2}\right)$$
$$= \boldsymbol{L}^{rig}(t) + \frac{1}{2}\tau\Delta t + O(\Delta t^2)$$
$$= \sum_i r_i^b(t) \times m_i v_i^b(t) + \frac{\Delta t}{2}\sum_i r_i^b(t) \times \mathbf{f}_i(t) + O(\Delta t^2)$$
$$= \sum_i r_i^b(t) \times \left(m_i v_i^b(t) + \frac{\Delta t}{2} \cdot \mathbf{f}_i(t)\right) + O(\Delta t^2)$$
$$= \sum_i r_i^b(t) \times m_i v_i^b\left(t + \frac{\Delta t}{2}\right) + O(\Delta t^2)$$
$$= \sum_i r_i^b(t) \times m_i \frac{\mathbf{r}_i^{rig,\,b}(t + \Delta t) - \mathbf{r}_i^b(t)}{2} + O\left(\Delta t^2\right)$$
$$= \boldsymbol{L}'^{rig}\left(t + \frac{\Delta t}{2}\right) + O(\Delta t^2), \tag{30}$$

where τ is the torque on the rigid structure. The angular velocity at time $t + \Delta t$ is

$$\omega^{rig}\left(t + \frac{\Delta t}{2}\right) = \boldsymbol{I}^{-1}\boldsymbol{L}^{rig}\left(t + \frac{\Delta t}{2}\right)$$
$$= \boldsymbol{I}^{-1}\boldsymbol{L}'^{rig}\left(t + \frac{\Delta t}{2}\right) + O(\Delta t^2)$$
$$= \omega'^{rig}\left(t + \frac{\Delta t}{2}\right) + O(\Delta t^2), \tag{31}$$
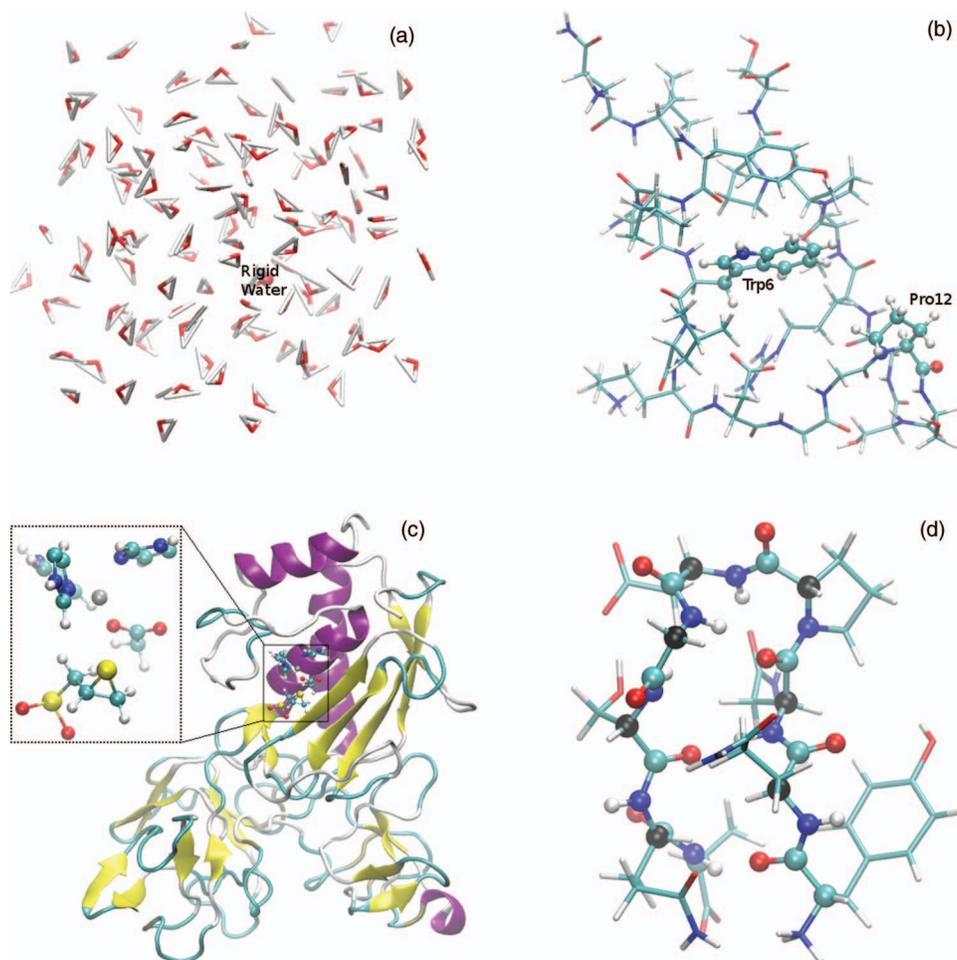
FIG. 2. Model systems, the rigid bodies are shown in ball and stick mode: (a) 126 water molecules box, the rigid water molecule is labeled; (b) Trp-case protein, two separate rigid bodies are labeled; (c) matrix metalloproteinase 2 with substrate, the single rigid body is enlarged in inset; (d) nine-residue β-hairpin peptide, peptide bonds as treated as individual rigid structures, C_α carbons shared by adjacent rigid structures are shown in black. The solvent molecules are not shown in (b), (c), or (d).

where $\omega'^{rig}(t + \Delta t/2)$ is the computed angular velocity in the simulation. Consequently, the rotational angle at time $t + \Delta t$ is

$$\theta = \omega^{rig}\left(t + \frac{\Delta t}{2}\right)\Delta t$$

$$= \left(\omega'^{rig}\left(t + \frac{\Delta t}{2}\right) + O(\Delta t^2)\right)\Delta t$$

$$= \omega'^{rig}\left(t + \frac{\Delta t}{2}\right)\Delta t + O(\Delta t^3)$$

$$= \theta'^{rig} + O(\Delta t^3), \tag{32}$$

where θ'^{rig} is the computed rotation angle in the simulation. Finally, the coordinates at time $t + \Delta t$ is

$$\mathbf{r}^{rig,\,b}(t + \Delta t) = \boldsymbol{R} \cdot \mathbf{r}^b(t) = e^\theta \mathbf{r}^b(t) = e^{\theta'^{rig}}\mathbf{r}^b(t) + O(\Delta t^3). \tag{33}$$

TABLE I. Total energies (in kcal/mol) and standard deviation of model systems from simulations using different time steps.

| Model systems | Time step (fs) | | | | | |
| | 1.0 | | 1.5 | | 2.0 | |
	Total energy	Standard deviation	Total energy	Standard deviation	Total energy	Standard deviation
a (SHAPE)	−665.49	0.017	−661.78	0.104	−661.44	0.86
a (SHAKE)	−665.49	0.017	−661.72	0.092	−651.14	0.58
a (no constraint)	−664.12	0.016	−660.22	0.081	−651.05	0.61
b (SHAPE)	−7681.02	0.051	−7631.59	0.26	−7505.21	6.6
b (no constraint)	−7783.08	0.040	−7733.09	0.22	−7607.02	3.8
c (SHAPE)	−111 021.16	2.8	−110 294.37	17.0	−108 361.57	192.8
c (no constraint)	−111 078.77	2.9	−110 341.24	20.2	−108 436.01	165.6

Therefore, for a single time step this algorithm has the third order of accuracy. Over a given time period, t, the number of the steps is in the order of Δt^{-1}. Therefore, the global error of the algorithm is in the order of $O(\Delta t^2)$.

III. NUMERICAL TESTS

Four model systems were applied as test cases for this rigid structure algorithm with the fourth system described at the end of this section. The model system a is a water box with 126 TIP3P water molecules (Fig. 2(a)).[17] Only one water molecule in the water box was treated as a rigid structure using SHAPE method. As a comparison, SHAKE for holonomic constraint was also applied to implement rigid structure for the same water molecule, which is treated by SHAPE. Simulation for the same water box without any constraints was also carried out for benchmark purpose. The model system b is a Trp-cage protein (PDB code: 1L2Y) in a box of 1169 water molecules (Fig. 2(b)). The side chain of residue Trp6 and the whole residue of Pro20 were treated as separate rigid structures. The model system c is protein matrix metalloproteinase 2 (MMP2) with its substrate in a box of 12 636 water molecules (Fig. 2(c)).[18] The active site, including side chains of residues His288, His 292, His298, and Glu289, the thiirane ring, methylene, and sulfone groups in substrate, and zinc ion, are combined as a single rigid structure. The time step of 1 fs was used for all there systems.

For system a, the NVE simulation using SHAPE method provided results that are essentially interchangeable to that using SHAKE method in terms of total energy of the whole systems and standard deviation along the 1 ns trajectory (Table I). The total energies presented in Table I include high frequency correction based on the expectation value of the symplectic shadow Hamiltonian.[15, 19] Both simulations using SHAPE and SHAKE methods approximately conserve the total energies at the level of accuracy very close to the simulation without any constraints (Table I). The RMSD of whole water box between two trajectories are calculated for the first 1000 steps (Fig. 3). The whole water box is very close to each other between two trajectories, demonstrating the same level of reliability for SHAPE algorithm as to the SHAKE method. The average and standard deviation of RMSD for the rigid water molecule of 1 ns trajectories using SHAPE method show
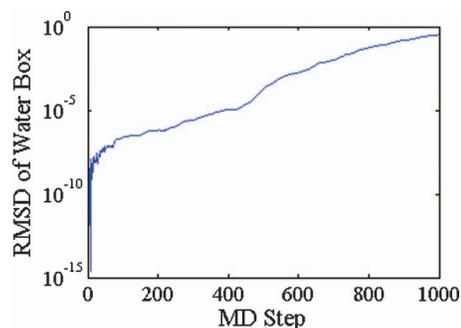
FIG. 3. The RMSD of whole water box of system a between two trajectories comparing SHAKE to SHAPE to implement rigid water constraints are plotted for the first 1000 MD steps. The unit of RMSD value is in Å.

even higher accuracy than the results of 1 ns trajectory using SHAKE method (Table II).

For system b, the dynamics using SHAPE method for two separate rigid bodies demonstrated the same level of conservation of total energy throughout the 1 ns trajectory as to the free MD simulation without any constraints (Table I). Both rigid structures maintain high accuracy during the simulation (Table II). In the system c, the rigid structure is composed of 43 atoms from 6 residues including substrate and zinc ion. The dynamics using SHAPE method for the defined rigid structure demonstrated the same level of conservation of total energy throughout the 1 ns trajectory as to the free MD simulation without any constraints (Table I). The rigid structure in this large protein system maintains the same level of accuracy (Table II) as to the systems a and b. For both systems b and c, it is not practical to implement rigid structure constraints using SHAKE.

To investigate the effect of time step on the SHAPE method, all the simulations presented above were repeated using both 1.5 and 2.0 fs as time step, respectively. It should be noted that these time steps are not suitable for the simulation setup in this study, and were employed only for the test and demonstration purposes. For system a, the simulation using SHAPE showed very similar accuracy in the conservation of total energies to the one using SHAKE when applying 1.5 fs as time step (Table I), and only slightly worse when applying 2.0 fs as time step (Table I). The SHAPE method maintained the rigid structure of the water molecule with slightly

TABLE II. RMSD (Å) of rigid bodies in model systems from simulations using different time steps.[a]

| Model systems | Time step (fs) | | | | | |
| | 1.0 | | 1.5 | | 2.0 | |
	Average	Standard deviation	Average	Standard deviation	Average	Standard deviation
a (SHAPE)	2.3×10^{-7}	2.2×10^{-7}	1.1×10^{-6}	9.6×10^{-7}	1.0×10^{-6}	7.5×10^{-7}
a (SHAKE)	4.2×10^{-7}	3.6×10^{-7}	7.7×10^{-7}	7.2×10^{-7}	1.3×10^{-6}	9.4×10^{-7}
b Trp6 side chain[b] (SHAPE)	2.3×10^{-7}	9.2×10^{-8}	3.1×10^{-7}	1.2×10^{-7}	5.1×10^{-7}	2.8×10^{-7}
b Pro12[b] (SHAPE)	2.8×10^{-7}	5.5×10^{-8}	3.0×10^{-7}	4.8×10^{-8}	6.9×10^{-7}	3.4×10^{-7}
c (SHAPE)	6.7×10^{-7}	5.8×10^{-8}	6.6×10^{-7}	6.2×10^{-8}	6.9×10^{-7}	6.2×10^{-8}

[a]Only the rigid parts are used to calculate RMSD. The RMSD of rigid body in each frame from MD trajectories is calculated with reference to the first frame.
[b]In system b, the RMSDs of two separate rigid bodies are calculated separately.

TABLE III. A nine-residue β-hairpin peptide as model system d for outer loop of SHAPE algorithm.

Methods	Energy fluctuation[a] (kcal/mol)	Energy drift (kcal/mol ps)[b]	Average iteration steps	Maximum iteration steps
SHAPE[c,d]	0.10	0.0034	18.1[e]	27[e]
SHAKE[f,d]	0.08	0.0028	13.3	20
SHAPE non-sharing[g,d]	0.06	0.0020	1[e]	1[e]

[a]Standard deviation of total energy with high frequency correction.

[b]Overall energy drift is measured by the least squares slope of energy change with respect to time.

[c]For SHAPE method, all eight peptide bonds are treated as individual planar rigid structures which share one C_α carbon with adjacent structures.

[d]All simulations are 100 ps with 1 fs time step. The convergence tolerance for outer loop of SHAPE is 10^{-7}, which is also applied for SHAKE for fair comparison.

[e]Iteration steps for SHAPE are outer loop iterations. When there is no sharing atom among rigid structures in SHAPE, only one outer loop iteration is necessary.

[f]For SHAKE method, the main chain chemical bonds in nine residues are constrained. But peptide planes are not treated as individual rigid structures.

[g]In this simulation using SHAPE method, each C_α carbon belongs to previous peptide plane only. No atoms are shared by different rigid planar structures.

larger RMSD and standard deviation when using 1.5 fs as time step, and smaller RMSD and standard deviation when using 2.0 fs time step comparing to SHAKE method (Table II). For systems b and c, the simulations using SHAPE showed comparable and slightly better accuracy in conservation of total energies, respectively, comparing to simulations without rigid body when using 1.5 fs time step (Table I). When using 2.0 fs time step, the simulations using SHAPE showed marginally worse conservation of total energies (Table I). For systems b and c, the time step showed very little effect on maintaining the rigid structure throughout the simulation using SHAPE method (Table II). The time step analysis indicates that the SHAPE method shows very little effect on the conservation of total energies, and is robust when using different time steps.

The model system d is a nine-residue peptide folded as β-hairpin[20] in a box of 290 water molecules to test the outer loop of SHAPE algorithm to handle the case in which one atom is shared by two rigid structures. All eight peptide planes are treated as individual rigid structures. Therefore, seven C_α carbons are shared by adjacent peptide bonds as rigid structures (Fig. 2(d)). Note that SHAKE cannot impose planar constraints, so in a separate simulation as comparison, SHAKE is applied to constrain all the main chain chemical bonds in the nine-residue peptide without treating peptide planes as individual rigid structures. Both simulations are run for 100 ps with 1 fs time step. The same convergence tolerance (10^{-7}) is applied for both SHAKE and outer loop of SHAPE algorithm. Two simulations are compared in Table III. The fluctuation of total energy with high frequency correction from the simulation using SHAPE algorithm is slightly higher but still comparable to that using SHAKE algorithm. The least squares slope of energy change with respect to time shows that both SHAKE and SHAPE display observable overall energy drift, with SHAKE slightly better than SHAPE. On average, it takes five more iterations for outer loop in SHAPE algorithm to converge than iterations of SHAKE algorithm. In an addition SHAPE simulation, eight peptide planes are treated as individual rigid structures, but each C_α carbon belongs to previous peptide plane only. Therefore, no atoms are shared by different rigid structures. This simulation displays smaller total energy fluctuation and overall energy drift than both SHAPE and SHAKE simulations above (Table III).

IV. CONCLUDING REMARKS

We presented an algorithm to implement rigid body constraints in MD simulations through iterative procedure of constructing rotation matrix for the rigid structure to preserve angular momentum. This algorithm avoids computing Lagrange multipliers of individual holonomic constraints. Therefore, an arbitrary number of particles can be selected to form a single rigid structure, and is more efficient than SHAKE for any systems with more than three atoms. An arbitrary number of such rigid structures can be implemented in simulation. The iterative procedure is independent from and conducted after each MD free integration step. Therefore, no significant modification of existing MD integrator is necessary to implement this algorithm. This can be easily plugged into any simulation package that already includes SHAKE. In addition, multiple rigid structures can be defined in a simulation, even those sharing one particle with neighboring rigid structures. The simulation of four test cases, including a pure water box and three solvated systems, demonstrated the reliability and utility of this algorithm.

Since this method is interchangeable with SHAKE, it suffers from the same failures that SHAKE exhibits at very large time steps or very high temperatures where the equations cannot be solved. In this instance, non-energy conserving approximations are employed while temperatures are too large, as an alternative to program termination.

In summary, the SHAPE method introduced here conserves linear and angular momentum, and is the only rigid body integration algorithm that is fully consistent and interchangeable with the SHAKE method. It shares with SHAKE both its holonomic and symplectic nature, and well conserves total energy for long simulations with minimal energy drift. It extends the ability of SHAKE like constraints to linear systems with three or more atoms, planar systems with four or more atoms, and to larger rigid structures where SHAKE is intractable.

ACKNOWLEDGMENTS

[1] I. G. Tironi, R. M. Brunne, and W. F. van Gunsteren, Chem. Phys. Lett. **250**, 19 (1996).

[2] S. Karl, E. T. Robert, G. S. Bobby, and W. N. Donald, Nanotechnology **8**, 103 (1997).

[3] A. V. Borisov and I. S. Mamaev, Regul. Chaotic Dyn. **7**, 177 (2002).

[4] J. B. Klauda, M. F. Roberts, A. G. Redfield, B. R. Brooks, and R. W. Pastor, Biophys. J. **94**, 3074 (2008).

[5] J.-P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).

[6] H. C. Andersen, J. Comput. Phys. **52**, 24 (1983).

[7] X.-W. Wu and S.-S. Sung, J. Comput. Chem. **19**, 1555 (1998).

[8] T. R. Forester and W. Smith, J. Comput. Chem. **19**, 102 (1998).

[9] H. M. Chun *et al.*, J. Comput. Chem. **21**, 159 (2000).

[10] R. Kutteh and R. B. Jones, Phys. Rev. E **61**, 3186 (2000).

[11] N. Neto and L. Bellucci, Chem. Phys. **328**, 259 (2006).

[12] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, 1987).

[13] S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks, J. Chem. Phys. **103**, 4613 (1995).

[14] Y. Ma, S. Soatto, J. Kosecka, and S. S. Sastry, *An Invitation to 3-D Vision* (Springer, 2003).

[15] B. R. Brooks *et al.*, J. Comput. Chem. **30**, 1545 (2009).

[16] See supplementary material at http://dx.doi.org/10.1063/1.4756796 for the subroutine implementing SHAPE rigid structure algorithm.

[17] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).

[18] P. Tao, J. F. Fisher, Q. Shi, T. Vreven, S. Mobashery, and H. B. Schlegel, Biochemistry **48**, 9839 (2009).

[19] J. Zhou, S. Reich, and B. R. Brooks, J. Chem. Phys. **112**, 7919 (2000).

[20] X. Wu and B. R. Brooks, Biophys. J. **86**, 1946 (2004).