JOURNAL OF COMPUTATIONAL PHYSICS **92**, 261–272 (1991)

Derivation and Testing of Explicit Equations of Motion for Polymers Described by Internal Coordinates

A. K. MAZUR AND V. E. DOROFEEV

Pacific Institute of Bioorganic Chemistry, USSR Academy of Sciences,
Vladivostok 690022, USSR

AND

R. A. ABAGYAN

Engelgardt Institute of Molecular Biology, USSR Academy of Sciences,
Vavilov str. 32, Moscow 117984, USSR

Received March 28, 1989; revised November 14, 1989

General Lagrange's equations of motion for a system of polymeric molecules are obtained in an explicit form. They can be used for simulating molecular dynamics of large molecules. The molecular conformations are described by internal coordinates, i.e., bond lengths, valence angles, and torsion angles. The equations derived permit any internal degrees of freedom to be frozen. The method is applied to an oligopeptide in an α-helical conformation. Three models of the molecule with different degrees of fixation are compared. It is shown that the method permits one to increase significantly the time step in molecular dynamics calculations.

1. INTRODUCTION

The method of molecular dynamics (MD), i.e., computer simulations of molecular motions, has been widely used in the recent decade for studying complex polymeric molecules such as proteins [1, 2] or nucleic acids [2, 3]. This method has also been used effectively as a tool for determining the three-dimensional structure of a macromolecule from NMR data [4]. Now it is widely accepted that MD simulation of complex polymeric molecules, biopolymers in particular, is very promising [1, 2].

There are two aspects of the MD method which form its basis: (i) the choice of a suitable mechanical model, and (ii) the derivation of the equations of motion of the system.

At the present MD method for polymers uses Cartesian coordinates of atoms to represent molecular structure and, correspondingly, Newton's equations to simulate its dynamical behavior. Therefore, at least from the methodological point of

0021-9991/91 $3.00

view it is "atomic dynamics" rather than molecular dynamics. By choosing the independent atoms as a mechanical model one can easily write the explicit equations of motion which can be directly integrated by a computer.

However, the Cartesian coordinates as variables inevitably introduce some limitations to the MD simulations. First, many of the 3N degrees of freedom of an *N*-atomic molecule are weakly exited at room temperature [5]. At the same time in this "atomic" representation such degrees of freedom are considered as classical oscillators, which is not physically justified. On the other hand the empirical potentials imposed on these degrees of freedom force the atoms to oscillate at very high frequencies. This in turn limits the time step in the numerical integration of Newtonian equations and reduces the time scale of the motions studied to subnanosecond range. Meanwhile, a number of "interesting events" in polyatomic molecules take rather long time intervals and are related to intramolecular rotations around single bonds. A straightforward solution of these problems obviously consists in freezing the covalent structure of the molecule, i.e., bond length, valence angles, aromatic rings, etc. Within the frame of the "atomic" representation this can be done by imposing explicit constraints on the Cartesian coordinates of certain atoms. A few types of constraints have been proposed for this purpose, with the fixation of atom–atom distances the simplest and most frequently used [6–9]. However, the computational problems connected with this approach increase rapidly as new constraints are added, and only the fixation of bond lengths appears to be practically efficient for large molecules [10].

It would be preferable, evidently, to freeze all the fast internal vibrations by introducing an appropriate number of rigid bonds and angles directly into the description of the molecule. The use of generalized coordinates and Lagrange–Hamilton formalism is a classical way to treat such systems. However, for a long time this approach was applied only to small molecules of several atoms, for example, to *n*-butane [11]. The Lagrange equations were derived for every particular molecule whereas the derivation of equations for any polymer was considered prohibitively complicated [6]. Recently we proposed a new methodology for modelling complex molecular systems such as biomacromolecules [12, 13]. Among other things it allows one to freeze any set of internal coordinates (i.e., bond lengths, bond angles, and dihedral angles) while describing the molecular motion by a universal set of differential equations. In this paper we present a complete derivation of the equations and the results of their testing on a small fragment of a protein molecule.

2. THE EQUATIONS OF MOTION FOR A SYSTEM OF BRANCHED POLYMER MOLECULES

In this section we derive the general equations of motion of one or several interacting branched polymer noncyclic molecules with their conformation specified by the internal coordinates.

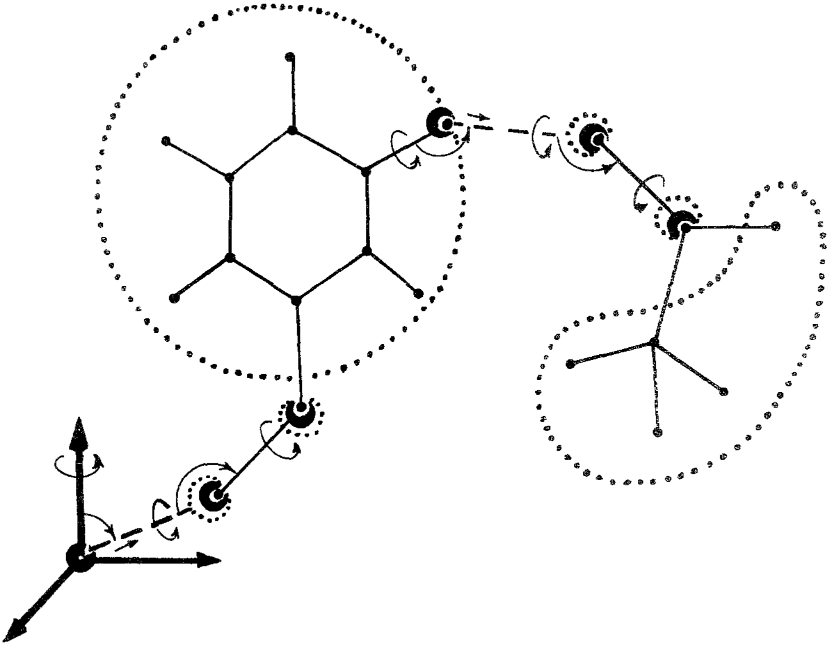

FIG. 1. A unified representation of a molecular system in the form of a BKS-tree [12]. Rigid bodies (encircled) are defined as groups of atoms depending on the same set of internal variables. The tree is constructed from its origin, which coincides with the origin of the gobal coordinate frame. The tree's structure is determined by internal coordinates, i.e., bond lengths and planar and dihedral angles, that can be arbitrarily frozen (in the example the two molecules with fixed internal structure can move independently relative to the coordinate system). Rigid bodies are connected in hinges or nodes (small heavy black circles) to which the internal coordinates are attributed. Virtual bonds are shown by broken lines.

2.1. *Formal Description and Original Equations*

Earlier we demonstrated that any number of molecules with any set of frozen internal coordinates (bond lengths inclusive) can be described as a unified tree [12]. The tree is composed of atomic groups called rigid bodies as far as their internal structure is kept fixed. The description implies the disconnection of cycles formed by rigid bodies and special numeration of rigid bodies, and it introduces virtual atoms and virtual bonds to connect molecules within the system and impose tree topology on it. We refer to such a representation as a BKS-tree [12]. Figure 1 shows an example of a BKS-tree, with the necessary definitions given in the legend.

The non-fixed internal coordinates of the BKS-tree form a set of the generalized coordinates (θ_k) of the system. Each variable is attributed to a particular node where the unit vector that determines the infinitesimal displacements in the structure due to this variable is defined (see Fig. 2). The proper ordering of the internal coordinates is extremely important for our purposes. There exists a natural order

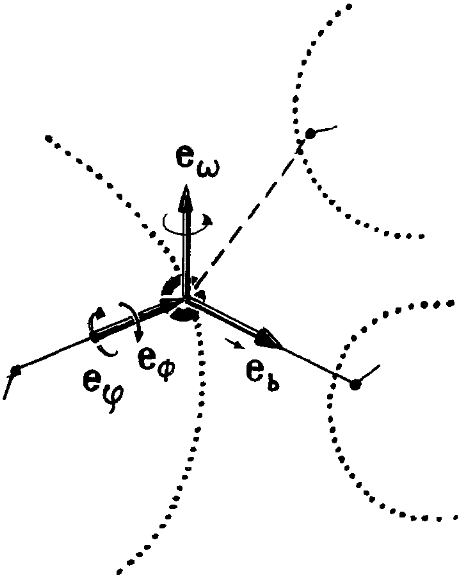

FIG. 2. Connection of rigid bodies in a BKS-tree. The unit vectors of the variables moving one of the two branches are shown. Their order is defined according to the influence they exert on the unit vectors (see the text). Therefore, the first are the torsion and phase angles (ϕ and Φ), then the planar angle (ω), and finally the bond length (b). The phase angle Φ is just a dihedral angle added to the torsion angle ϕ to construct various branches coming from the same node.

of variables owing to the relations between variables and their unit vectors. If the influence between two variables exists, only one of them can move the unit vector of another but not vice versa. It is natural that the former would always have a smaller index than the latter. The appropriate order of variables attributed to the same node is shown in Fig. 2.

Let us define a set of variables attributed to each atom in the following way. The position of the particular atom α (i.e., its radius-vector $\mathbf{r}_\alpha$) is determined by the chain V_α of generalized coordinates θ_i^α (Fig. 3a). They have the order described above at each node with the nodes ordered according to the rigid bodies which they define [12]. Let the lower index i run from 1 to n_α in succession. Define also a set of atoms d_k defined for the variable θ_k, which comprises the atoms having positions depending on θ_k (Fig. 3b).

To derive the general equations for any BKS-tree we start from the Lagrangian equations of motion [14]

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}_k}\right) - \frac{\partial L}{\partial \theta_k} = 0, \qquad k = 1, 2, ..., n. \tag{1}$$

The Lagrangian function L is the difference between the kinetic energy $T(\theta, \dot{\theta})$

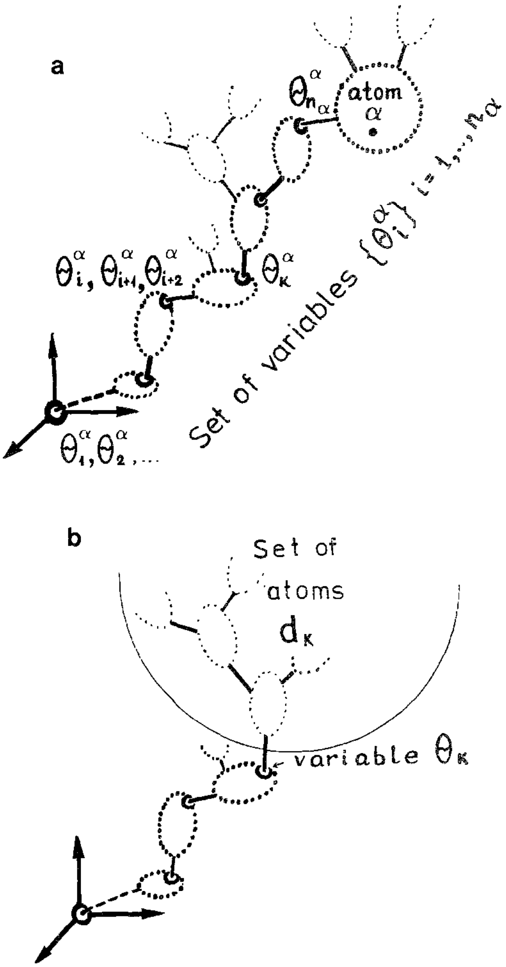

FIG. 3. Two sets attributed to each atom and each variable, respectively. (a) The chain of variables V_α determining the position of a particular atom α. From one to four variables of the chain are defined in each node. (b) The set d_k of atoms influenced by the generalized variable θ_k.

depending on vectors of generalized variables and velocities and the potential energy $U(\bar{\theta})$. Substituting $\sum_\alpha [m_\alpha \dot{\mathbf{r}}_\alpha^2/ - U(\bar{\theta})]$ for L in the Lagrange equations we obtain

$$\sum_{\alpha \in d_k} m_\alpha \left[\frac{d}{dt} \left(\dot{\mathbf{r}}_\alpha \frac{\partial \dot{\mathbf{r}}_\alpha}{\partial \dot{\theta}_k} \right) - \dot{\mathbf{r}}_\alpha \left(\frac{\partial \dot{\mathbf{r}}_\alpha}{\partial \theta_k} \right) \right] = -\frac{\partial U}{\partial \theta_k}. \tag{2}$$

In recent papers [12, 13, 15] the algorithm for rapid calculation of the derivatives of the conformational energy in the right-hand part of Eq. (2) have been described for any molecular system and any set of variables.

Our aim is to obtain the derivatives in the left-hand part of Eq. (2).

The infinitesimal displacement $d\mathbf{r}_\alpha$ is expressed by the formula

$$d\mathbf{r}_\alpha = \sum_{i=1}^{n_\alpha} [S_i \mathbf{e}_i^\alpha \times (\mathbf{r}_\alpha - \mathbf{r}_i^\theta)\, d\theta_i^\alpha + (1 - S_i)\, \mathbf{e}_i^\alpha\, d\theta_i^\alpha], \tag{3}$$

where $\mathbf{e}_i^\alpha$ are the unit vectors corresponding to the variable θ_i^α; while $\mathbf{r}_i^\theta$ is the radius-vector of its node. The upper index α indicates that the variables are taken from the chain V_α. S_i is the indicator of the type of variable: it is equal to unity or zero for the variable angles or bond lengths, respectively. The first term in the right-hand part of the formula describes the elementary rotation of the atom α around the axis of rotation defined by the unit vector $\mathbf{e}_i^\alpha$. The second term describes just the translation along the unit vector of the variable bond.

2.2. *Computation of* $(d/dt)[\dot{\mathbf{r}}_\alpha(\partial \dot{\mathbf{r}}_\alpha / \partial \dot{\theta}_k)]$

From Eq. (3) we obtain

$$\dot{\mathbf{r}}_\alpha = \sum_{i=1}^{n_\alpha} [S_i \mathbf{e}_i \times (\mathbf{r}_\alpha - \mathbf{r}_i^\theta) \cdot \dot{\theta}_i + (1 - S_i)\, \mathbf{e}_i \cdot \dot{\theta}_i] \tag{4}$$

(here and below we omit the upper index α for the sake of brevity). The partial derivative $\sum \partial \dot{\mathbf{r}}_\alpha / \partial \dot{\theta}_k$ is then equal to

$$\frac{\partial \dot{\mathbf{r}}_\alpha}{\partial \dot{\theta}_k} = S_k \mathbf{e}_k \times \mathbf{r}_{\alpha/k} + (1 - S_k)\, \mathbf{e}_k, \tag{5}$$

where $\mathbf{r}_{\alpha/k}$ stands for $(\mathbf{r}_\alpha - \mathbf{r}_k^\theta)$. To differentiate Eqs. (4) and (5) with respect to time one should calculate the derivatives $d\mathbf{e}_i/dt$ and $d\mathbf{r}_{\alpha/i}/dt$. The increments $d\mathbf{e}_i$ and $d\mathbf{r}_i^\theta$ are the sums of elementary rotations and displacements (only rotations for $d\mathbf{e}_i$):

$$d\mathbf{e}_i = \sum_{m=1}^{i-1} S_m \mathbf{e}_m \times \mathbf{e}_i \cdot d\theta_m \tag{6}$$

$$d\mathbf{r}_i^\theta = \sum_{m=1}^{i-1} [S_m \mathbf{e}_m \times (\mathbf{r}_i^\theta - \mathbf{r}_m^\theta) \cdot d\theta_m + (1 - S_m)\, \mathbf{e}_m \cdot d\theta_m]. \tag{7}$$

By subtracting relation (7) from (3) we obtain also

$$d\mathbf{r}_{\alpha/i} = \sum_{m=1}^{i-1} S_m \mathbf{e}_m \times \mathbf{r}_{\alpha/i} \cdot d\theta_m + \sum_{m=i}^{n_\alpha} [S_m \mathbf{e}_m \times \mathbf{r}_{\alpha/m} \cdot d\theta_m + (1 - S_m)\, \mathbf{e}_m \cdot d\theta_m]. \tag{8}$$

The corresponding time derivatives are given by

$$\dot{\mathbf{e}}_i = \sum_{m=1}^{i-1} S_m \mathbf{e}_m \times \mathbf{e}_i \cdot \dot{\theta}_m \tag{9}$$

$$\dot{\mathbf{r}}_i^{\theta} = \sum_{m=1}^{i-1} [S_m \mathbf{e}_m \times (\mathbf{r}_i^{\theta} - \mathbf{r}_m^{\theta}) \cdot \dot{\theta}_m + (1 - S_m)\, \mathbf{e}_m \cdot \dot{\theta}_m] \tag{10}$$

$$\dot{\mathbf{r}}_{\alpha;i} = \sum_{m=1}^{i-1} S_m \mathbf{e}_m \times \mathbf{r}_{\alpha;i} \cdot \dot{\theta}_m + \sum_{m=i}^{n_\alpha} [S_m \mathbf{e}_m \times \mathbf{r}_{\alpha;m} \cdot \dot{\theta}_m + (1 - S_m)\, \mathbf{e}_m \cdot \dot{\theta}_m]. \tag{11}$$

Using these time derivatives one can differentiate Eqs. (4) and (5). The required time derivative can be then expressed by 4 single sums and 16 double sums ($a \times b \times c$ stands for $a \times (b \times c)$; the intervening computations are omitted):

$$\begin{aligned}
\frac{d}{dt}\left(\dot{\mathbf{r}}_\alpha \frac{\partial \dot{\mathbf{r}}_\alpha}{\partial \dot{\theta}_k}\right) = {} & \sum_{i=1}^{n_\alpha} (1 - S_k)\, S_i \mathbf{e}_k (\mathbf{e}_i \times \mathbf{r}_{\alpha;i})\, \ddot{\theta}_i + \sum_{i=1}^{n_\alpha} (1 - S_k)(1 - S_i)\, \mathbf{e}_k \mathbf{e}_i \cdot \ddot{\theta}_i \\
& + \sum_{i=1}^{n_\alpha} S_k S_i (\mathbf{e}_k \times \mathbf{r}_{\alpha;k})(\mathbf{e}_i \times \mathbf{r}_{\alpha;i})\, \ddot{\theta}_i + \sum_{i=1}^{n_\alpha} S_k (1 - S_i)(\mathbf{e}_k \times \mathbf{r}_{\alpha;k})\, \mathbf{e}_i \cdot \ddot{\theta}_i \\
& + \sum_{i=1}^{n_\alpha} \sum_{m=1}^{k-1} S_k S_m S_i (\mathbf{e}_m \times \mathbf{e}_k \times \mathbf{r}_{\alpha;k})(\mathbf{e}_i \times \mathbf{r}_{\alpha;i}) \cdot \dot{\theta}_m \dot{\theta}_i \\
& + \sum_{i=1}^{n_\alpha} \sum_{m=k}^{n_\alpha} S_k S_m S_i (\mathbf{e}_k \times \mathbf{e}_m \times \mathbf{r}_{\alpha;m})(\mathbf{e}_i \times \mathbf{r}_{\alpha;i}) \cdot \dot{\theta}_m \dot{\theta}_i \\
& + \sum_{i=1}^{n_\alpha} \sum_{m=k}^{n_\alpha} S_k (1 - S_m)\, S_i (\mathbf{e}_k \times \mathbf{e}_m)(\mathbf{e}_i \times \mathbf{r}_{\alpha;i}) \cdot \dot{\theta}_m \dot{\theta}_i \\
& + \sum_{i=1}^{n_\alpha} \sum_{m=1}^{i-1} S_k S_m S_i (\mathbf{e}_k \times \mathbf{r}_{\alpha;k})(\mathbf{e}_m \times \mathbf{e}_i \times \mathbf{r}_{\alpha;i}) \cdot \dot{\theta}_m \dot{\theta}_i \\
& + \sum_{i=1}^{n_\alpha} \sum_{m=i}^{n_\alpha} S_k S_m S_i (\mathbf{e}_k \times \mathbf{r}_{\alpha;k})(\mathbf{e}_i \times \mathbf{e}_m \times \mathbf{r}_{\alpha;m}) \cdot \dot{\theta}_m \dot{\theta}_i \\
& + \sum_{i=1}^{n_\alpha} \sum_{m=i}^{n_\alpha} S_k (1 - S_m)\, S_i (\mathbf{e}_k \times \mathbf{r}_{\alpha;k})(\mathbf{e}_i \times \mathbf{e}_m) \cdot \dot{\theta}_m \dot{\theta}_i \\
& + \sum_{i=1}^{n_\alpha} \sum_{m=1}^{k-1} S_k S_m (1 - S_i)(\mathbf{e}_m \times \mathbf{e}_k \times \mathbf{r}_{\alpha;k})\, \mathbf{e}_i \cdot \dot{\theta}_m \dot{\theta}_i \\
& + \sum_{i=1}^{n_\alpha} \sum_{m=k}^{n_\alpha} S_k (1 - S_m)(1 - S_i)(\mathbf{e}_k \times \mathbf{e}_m)\, \mathbf{e}_i \cdot \dot{\theta}_m \dot{\theta}_i \\
& + \sum_{i=1}^{n_\alpha} \sum_{m=k}^{n_\alpha} S_k S_m (1 - S_i)(\mathbf{e}_k \times \mathbf{e}_m \times \mathbf{r}_{\alpha;m})\, \mathbf{e}_i \cdot \dot{\theta}_m \dot{\theta}_i
\end{aligned}$$

$$
\begin{aligned}
&+\sum_{i=1}^{n_\alpha}\sum_{m=1}^{i-1} S_k S_m(1-S_i)(\mathbf{e}_k\times\mathbf{r}_{\alpha/k})(\mathbf{e}_m\times\mathbf{e}_i)\cdot\dot\theta_m\dot\theta_i\\
&+\sum_{i=1}^{n_\alpha}\sum_{m=1}^{k-1} (1-S_k)\,S_m S_i(\mathbf{e}_m\times\mathbf{e}_k)(\mathbf{e}_i\times\mathbf{r}_{\alpha/i})\cdot\dot\theta_m\dot\theta_i\\
&+\sum_{i=1}^{n_\alpha}\sum_{m=1}^{i-1} (1-S_k)\,S_m S_i\mathbf{e}_k(\mathbf{e}_m\times\mathbf{e}_i\times\mathbf{r}_{\alpha/i})\cdot\dot\theta_m\dot\theta_i\\
&+\sum_{i=1}^{n_\alpha}\sum_{m=i}^{n_\alpha} (1-S_k)\,S_m S_i\mathbf{e}_k(\mathbf{e}_i\times\mathbf{e}_m\times\mathbf{r}_{\alpha/m})\cdot\dot\theta_m\dot\theta_i\\
&+\sum_{i=1}^{n_\alpha}\sum_{m=i}^{n_\alpha} (1-S_k)(1-S_m)\,S_i\mathbf{e}_k(\mathbf{e}_i\times\mathbf{e}_m)\cdot\dot\theta_m\dot\theta_i\\
&+\sum_{i=1}^{n_\alpha}\sum_{m=1}^{k-1} (1-S_k)\,S_m(1-S_i)(\mathbf{e}_m\times\mathbf{e}_k)\,\mathbf{e}_i\cdot\dot\theta_m\dot\theta_i\\
&+\sum_{i=1}^{n_\alpha}\sum_{m=1}^{i-1} (1-S_k)\,S_m(1-S_i)\,\mathbf{e}_k(\mathbf{e}_m\times\mathbf{e}_i)\cdot\dot\theta_m\dot\theta_i.
\end{aligned}
\tag{12}
$$

2.3. *Computation of* $\frac{1}{2}(\partial\dot{\mathbf{r}}_\alpha^2/\partial\theta_k)$.

The derivatives of $\mathbf{e}_m$ and $\mathbf{r}_{\alpha/m}$ with respect to θ_k necessary for further computations can be derived easily from Eqs. (6) and (8):

$$
\frac{\partial\mathbf{e}_m}{\partial\theta_k}=\begin{cases} S_k(\mathbf{e}_k\times\mathbf{e}_m) & \text{for } m>k\\ 0 & \text{for } m\leqslant k\end{cases}
\tag{13}
$$

$$
\frac{\partial\mathbf{r}_{\alpha/m}}{\partial\theta_k}=\begin{cases} S_k(\mathbf{e}_k\times\mathbf{e}_{\alpha/m}) & \text{for } m>k\\ S_k(\mathbf{e}_k\times\mathbf{r}_{\alpha/k})+(1-S_k)\,\mathbf{e}_k & \text{for } m\leqslant k.\end{cases}
\tag{14}
$$

Now one can obtain Eq. (4) squared and differentate it applying Eqs. (13) and (14).

The ultimate formula for the second term in the right-hand part of Lagrange's equation (3) is given by

$$
\begin{aligned}
\frac{1}{2}\left(\frac{\partial\dot{\mathbf{r}}_\alpha^2}{\partial\theta_k}\right)=&\sum_{i=1}^{n_\alpha}\sum_{m=k+1}^{n_\alpha} S_k S_m S_i(\mathbf{e}_k\times\mathbf{e}_m\times\mathbf{r}_{\alpha/m})(\mathbf{e}_i\times\mathbf{r}_{\alpha/i})\cdot\dot\theta_m\dot\theta_i\\
&+\sum_{i=1}^{n_\alpha}\sum_{m=1}^{k} S_k S_m S_i(\mathbf{e}_m\times\mathbf{e}_k\times\mathbf{r}_{\alpha/k})(\mathbf{e}_i\times\mathbf{r}_{\alpha/i})\cdot\dot\theta_m\dot\theta_i\\
&+\sum_{i=1}^{n_\alpha}\sum_{m=1}^{k} (1-S_k)\,S_m S_i(\mathbf{e}_m\times\mathbf{e}_k)(\mathbf{e}_i\times\mathbf{r}_{\alpha/i})\cdot\dot\theta_m\dot\theta_i\\
&+\sum_{i=1}^{n_\alpha}\sum_{m=k+1}^{n_\alpha} S_k S_m(1-S_i)(\mathbf{e}_k\times\mathbf{e}_m\times\mathbf{r}_{\alpha/m})\,\mathbf{e}_i\cdot\dot\theta_m\dot\theta_i
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{i=1}^{n_\alpha}\sum_{m=1}^{k} S_k S_m(1-S_i)(\mathbf{e}_m\times\mathbf{e}_k\times\mathbf{r}_{\alpha,k})\,\mathbf{e}_i\cdot\dot{\theta}_m\dot{\theta}_i\\
&+\sum_{i=1}^{n_\alpha}\sum_{m=1}^{k}(1-S_k)\,S_m(1-S_i)(\mathbf{e}_m\times\mathbf{e}_k)\,\mathbf{e}_i\cdot\dot{\theta}_m\dot{\theta}_i\\
&+\sum_{i=k+1}^{n_\alpha}\sum_{m=1}^{n_\alpha} S_k(1-S_i)\,S_m(\mathbf{e}_m\times\mathbf{r}_{\alpha:m})(\mathbf{e}_k\times\mathbf{e}_i)\cdot\dot{\theta}_m\dot{\theta}_i\\
&+\sum_{i=1}^{n_\alpha}\sum_{m=k+1}^{n_\alpha} S_k(1-S_m)(1-S_i)(\mathbf{e}_k\times\mathbf{e}_m)\,\mathbf{e}_i\cdot\dot{\theta}_m\dot{\theta}_i.
\end{aligned}
\tag{15}
$$

2.4. *The Ultimate Form of the Equations*

If we substitute the right-hand parts of Eqs. (12) and (15) into Eq. (2) and simplify the expression, we obtain

$$
\begin{aligned}
&\sum_{\alpha\in d_k} m_\alpha \overset{V_\alpha}{\Bigg\{} \sum_{i=1}^{n_\alpha}\big[S_k S_i(\mathbf{e}_k\times\mathbf{r}_{\alpha/k})(\mathbf{e}_i\times\mathbf{r}_{\alpha:i}) + S_k(1-S_i)(\mathbf{e}_k\times\mathbf{r}_{\alpha\cdot k})\,\mathbf{e}_i\\
&\quad+(1-S_k)\,S_i\mathbf{e}_k(\mathbf{e}_i\times\mathbf{r}_{\alpha,i})+(1-S_k)(1-S_i)\,\mathbf{e}_k\mathbf{e}_i\big]\cdot\ddot{\theta}_i\\
&\quad+\sum_{i=1}^{n_\alpha}\big[S_k S_i^2(\mathbf{e}_k\times\mathbf{r}_{\alpha/k})(\mathbf{e}_i\times\mathbf{e}_i\times\mathbf{r}_{\alpha/i})+(1-S_k)\,S_i^2\mathbf{e}_k(\mathbf{e}_i\times\mathbf{e}_i\times\mathbf{r}_{\alpha/i})\big]\cdot\dot{\theta}_i^2\\
&\quad+2\cdot\sum_{i=2}^{n_\alpha}\sum_{m=1}^{i-1}\big[S_k S_i S_m(\mathbf{e}_k\times\mathbf{r}_{\alpha\cdot k})(\mathbf{e}_m\times\mathbf{e}_i\times\mathbf{r}_{\alpha/i})\\
&\quad+S_k(1-S_i)\,S_m(\mathbf{e}_k\times\mathbf{r}_{\alpha/k})(\mathbf{e}_m\times\mathbf{e}_i)+(1-S_k)\,S_i S_m\mathbf{e}_k(\mathbf{e}_m\times\mathbf{e}_i\times\mathbf{r}_{\alpha;i})\\
&\quad+(1-S_k)(1-S_i)\,S_m\mathbf{e}_k(\mathbf{e}_m\times\mathbf{e}_i)\big]\cdot\dot{\theta}_m\dot{\theta}_i\Bigg\} = -\frac{\partial U}{\partial\theta_k}.
\end{aligned}
\tag{16}
$$

The only trick sometimes used in the course of these algebraic computations is the change of the summation order in the double sums:

$$
\sum_{i=1}^{n_\alpha-1}\sum_{m=i+1}^{n_\alpha} F_{im}=\sum_{i=2}^{n_\alpha}\sum_{m=1}^{i-1} F_{mi}.
\tag{17}
$$

The symbol V_α over the braces indicates that within them the variables are numbered in succession according to their order in the chain V_α. One should note that the above computations are strongly dependent on the numeration of the generalized variables (the rule is discussed above and in the legend to Fig. 2). In fact, the indices i, k, m within the braces in Eq. (16) are functions of α. When the global numeration is used (i.e., each variable has its unique index) and i, j, k run from 1 to the total number of variables n_{var}, the equations are given in the general form

$$\sum_i a_{ki}\ddot{\theta}_i = -\frac{\partial U}{\partial \theta_k} - \sum_i b_{ki}\dot{\theta}_i^2 - \sum_i \sum_m c_{kim}\dot{\theta}_m\dot{\theta}_i. \tag{18}$$

In this form of the equations a_{ki}, b_{ki}, and c_{kim} are calculated by summation taking the tree topology into account. As a matter of fact it can be done in different ways, in particular, by a straightforward summation of the terms from Eqs. (16) or using a procedure which takes the tree topology into account [12]. In the last case these computations take only a minor part of the total computation time.

3. A TEST EXAMPLE: MOLECULAR DYNAMICS OF AN α-HELIX

To test the equations and the computational scheme we applied the method to an oligopeptide $(\mathrm{Ala})_9$, having the α-helical starting conformation. The solvent effects are neglected; i.e., one molecule in a vacuum is considered.

The molecular models tested had different sets of free variables. To check the validity of the equations we carried out several probe integrations in forward and backward time directions for each model. The conservation of energy, momentum, and angular momentum of the molecule has been checked. In addition, for a completely unfixed molecule we compared the trajectories calculated by Eqs. (16) and (18) and those obtained from a traditional MD simulation in Cartesian coordinates. All tests confirmed the correctness of the explicit Lagrangian equations derived in the previous section.

In all test calculations the trajectories of 110 time steps [7] are obtained by integrating Eqs. (16) and (18). Before this the molecule was slowly heated up to the temperature of $\cong 300$ K and equilibrated during 4 ps. This time was sufficient for the mean fluctuation of the kinetic energy $\langle \delta K \rangle$ to be stabilized. This value is large for a non-equilibrium system and decreases to a minimum as equilibrium is approached [16]. The temperature of the system was calculated as

$$T = \frac{2 \cdot \langle K \rangle}{N \cdot k_{\mathrm{B}}},$$

where N is the number of internal degrees of freedom, K is the instantaneous kinetic energy, and k_{B} is the Boltzmann constant.

Since the system of equations given by Eq. (18) is not resolved with respect to $\ddot{\theta}$ the corresponding linear algebraic system is solved at each time step by the Kholezki algorithm [17]. The equations are integrated by the method of Beeman [18] with the fourth-order prediction formulas for generalized velocities. In the calculations we use the empirical potentials compiled from those of ECEPP [19, 20] and CHARMM [21]. Such a mixing seems quite satisfactory in view of the limited purposes of our numerical tests.

Three models of the molecule were chosen for comparisons. The first one consists of 93 atoms and includes all hydrogen atoms explicitly. All the bond lengths in the molecule are fixed at their equilibrium values. In addition, the model has fixed phase angles (see Fig. 2), as far as at present there are no proper empirical potentials for these angles in the literature. With these fixations the molecule has 133 degrees of freedom.

The second model differs from the first one in that all its valence angles are fixed. As a result the number of variables describing the structure decreases to 42. In the third model we use united atoms instead of the alanine side chains, thus excluding the torsion angles of the methyl groups. In addition, the torsion angle rotating the hydroxyl group at the C-terminus is fixed. This model is described by 32 degrees of freedom.

The accuracy of the calculations is compared using the conservation of the total energy as an indicator. For each test trajectory the average total energy $\langle E\rangle$ is computed as well as the root mean square deviation $\sqrt{\langle \Delta E^2\rangle}$. The ratio

$$\delta_{\mathrm{E}}=\frac{\sqrt{\langle \Delta E^2\rangle}}{\langle E\rangle}$$

is used for comparisons because $\langle E\rangle$ varies strongly from one model to another. These differences arise first from the fact that the kinetic energy of the molecule at the same temperature depends on the number of degrees of freedom. The second reason is that the interactions within the rigid bodies composing the BKS-tree are always discarded; i.e., only the varying parts of the potential energy are calculated.

With the small time step $h=0.5$ fs the δ_{E} and $\langle E\rangle$ values computed from the 100-step trajectories (the first 10 steps of the 110-step trajectory are ignored as in [7]) were 0.4×10^{-2}, 0.4×10^{-5}, and 0.8×10^{-6} and 23.5, -22.6, -26.8 kcal/mol for the three models considered, respectively. When the time step increases the values of δ_{E} also increase, with the ratios between them being approximately constant. With the time step of 1 fs commonly used in MD simulations of proteins the δ_{E} value for the first model equals approximately 10^{-2}. This level of accuracy is reached with the h values 9 and 13 fs for the other two models, respectively.

In constructing the three models of an α-helix we tried in each succeeding model to freeze the degrees of freedom corresponding to the fastest oscillations in the previous model. Note that unlike the case of Newton's equations the additional freezing of the molecules makes the computational task less expensive. For instance, one step of integration for the three models used in this paper takes 6.0, 1.1, and 0.6 s, respectively, on the computer (EC1061). In the first case the solution of the linear algebraic system of 133 equations takes the major part of the computer time. In the other two cases this system is much smaller and the calculation of the energy gradient appears to be even more time consuming. These results show that the time step in MD simulations using Eqs. (18) can be increased significantly by excluding the fast degrees of freedom from the system.

A number of computational problems are still to be solved. Finding an effective method for solving the linear algebraic system given by Eqs. (18) is probably the most important one. Nevertheless it seems clear that by using Eqs. (16) and (18) one obtains useful tools for investigating the dynamic properties of complex polymeric molecules.

REFERENCES

1. M. KARPLUS AND J. A. MCCAMMON, *Ann. Rev. Biochem.* **53**, 263 (1984).
2. J. A. MCCAMMON AND S. C. HARVEY, *Dynamics of Proteins and Nucleic Acids* (Cambridge Univ. Press, Cambridge, 1987).
3. S. C. HARVEY, M. PRABHARAN, B. MAO, AND J. A. MCCAMMON, *Science* **223**, 1189 (1984).
4. G. M. CLORE AND A. M. GRONENBORN, *Protein Eng.* **1**, 275 (1987).
5. N. GO AND H. A. SCHERAGA, *J. Chem. Phys.* **51**, 4751 (1969).
6. J. P. RYCKAERT, G. CICCOTTI, AND H. J. C. BERENDSEN, *J. Comput. Phys.* **23**, 327 (1977).
7. W. F. VAN GUNSTEREN AND H. J. C. BERENDSEN, *Mol. Phys.* **34**, 1311 (1977).
8. J. P. RYCKAERT, *Mol. Phys.* **55**, 549 (1985).
9. J. W. PERRAM AND H. G. PETERSEN, *Mol. Phys.* **65**, 861 (1988).
10. W. F. VAN GUNSTEREN AND M. KARPLUS, *Macromolecules* **15**, 1528 (1982).
11. K. REFSON AND G. S. PAWLEY, *Mol. Phys.* **61**, 669 (1987).
12. A. K. MAZUR AND R. A. ABAGYAN, *J. Biomol. Struct. Dyn.* **6**, 815 (1989).
13. R. A. ABAGYAN AND A. K. MAZUR, *J. Biomol. Struct. Dyn.* **6**, 833 (1989).
14. L. D. LANDAU AND E. M. LIFSHITS, *Mechanics* (Nauka, Moscow, 1974).
15. H. ABE, W. BRAUN, T. NOGUTI, AND N. GO, *Comput. & Chem.* **8**, 239 (1984).
16. J. L. LEBOWITZ, J. K. PERCUS, AND L. VERLET, *Phys. Rev.* **153**, 250 (1967).
17. S. PISSARNEZKY, *Sparse Matrix Technology* (Academic Press, New York, 1984), Chap. 2.
18. D. BEEMAN, *J. Comput. Phys.* **20**, 130 (1976).
19. F. A. MOMANY, R. F. MCGUIRE, A. W. BURGESS, AND H. A. SCHERAGA, *J. Phys. Chem.* **79**, 2361 (1975).
20. G. NEMETHY, M. S. POTTIE, AND H. A. SCHERAGA, *J. Phys. Chem.* **87**, 1183 (1983).
21. B. R. BROOKS, R. E. BRUCCOLERY, B. D. OLAFSON, D. J. STATES, S. SWAMINATHAN, AND M. KARPLUS, *J. Comput. Chem.* **2**, 187 (1983).