

Optimization of Numerical Algorithms for Internal Coordinate Molecular Dynamics

VLADIMIR E. DOROFEYEV AND ALEXEY K. MAZUR*

Pacific Institute of Bioorganic Chemistry, Russian Academy of Sciences, Vladivostok 690022, Russia

Received June 20, 1991; revised October 8, 1992

Different computational procedures are compared for numerical solution of equations of motion for molecular dynamics of semi-rigid polymeric molecules with internal coordinates as general variables and an optimal method is proposed. The method uses forced conservation of momentum and angular momentum of the system and a predictor–corrector scheme with several iterations of correction for only generalized velocities at each time step. Variation of the accuracy of numerical solution with the time step and the order of the predictor–corrector algorithm is studied on three different partially fixed models of the oligopeptide (Ala)₉. The maximum possible time step gradually increases as fast degrees of freedom are frozen. For dynamics in torsion angles the method provides a reasonable accuracy for the time step as large as 0.02 ps. The difference between the results obtained here and the earlier estimates made by using Newtonian molecular dynamics with constraints is discussed. © 1993 Academic Press, Inc.

1. INTRODUCTION

Internal coordinate molecular dynamics (ICMD) is a method for molecular dynamics (MD) simulations of semi-rigid polymeric molecules [1–3]. Unlike the conventional MD method it uses internal coordinates, i.e., bond lengths, bond angles, and torsions, as generalized variables of the Lagrangian equations to describe the motion of the system being studied. Thus holonomic constraints on the fast degrees of freedom can be imposed implicitly by the general Lagrange–Hamilton formalism. This results in a theoretical possibility to “break through” the femtosecond oscillations, which normally limit the time step in MD, and to simulate large scale motions by freezing relatively rigid parts of molecules, which is particularly important for MD of biopolymers such as domain proteins and DNA.

Numerical algorithms that integrate Newtonian equations have been thoroughly studied during the recent decades [4, 5]. Usually the algorithms of Verlet [6], Beeman [7], or the leapfrog method [5] are considered as optimal for traditional MD because of their stability and relatively high accuracy. This fact, however, does not

necessarily mean that the same methods can be applied for ICMD because there are important differences in the general form of equations used in these two alternative approaches. It is clear, however, that since algorithms for ICMD should calculate forces only once per time step, only predictor–corrector (or simply predictor) methods can be used. In the few papers where generalized coordinates were used for MD simulations [3, 8, 9] predictor–corrector methods were applied. However, the problem of the optimal choice of the algorithms for non-Newtonian MD was never considered.

In this paper we present a comparative study of some linear multistep methods for integrating the second-order differential equations for ICMD of polymers. In Sections 2 and 3 we consider how certain intrinsic properties of the equations of motion used in ICMD can be used in numerical algorithms. Section 4 presents some characteristic results of numerical tests with several molecular models. The main factors that affect the accuracy of the calculations are discussed in this section. In Section 5 we briefly discuss general problems connected with MD simulations of constrained molecules and formulate the main conclusions. Reference information about algorithms used is given in the Appendix.

2. CONSERVATION LAWS

The conservation laws of mechanics play an important role in MD simulations because it is implied that average properties of the system being studied are determined by the values of the conserved parameters. Exact trajectories and those calculated by a computer can diverge significantly but normally this can be ignored if both trajectories are sampling the same hypersurface of the phase space determined by the conservation laws [4]. That is why these laws are customarily used for testing the accuracy and stability of the numerical algorithms for integration of equations of motion for MD.

Consider a predictor–corrector method written as $P_r^k P_v^k (E_r E_a C_r^q C_v^q)_m$, where P_r^k and P_v^k denote predictors of

* To whom correspondence should be addressed.

the k th order for generalized coordinates and velocities, respectively, given by the formulas:

$$\theta_i^{n+1} = \theta_i^n + h\dot{\theta}_i^n + h^2 \sum_{l=1}^{k-1} \alpha_l \ddot{\theta}_i^{n-l+1} \quad (1a)$$

and

$$\dot{\theta}_i^{n+1} = \dot{\theta}_i^n + h \sum_{l=1}^k \beta_l \ddot{\theta}_i^{n-l+1}. \quad (1b)$$

Here h is the step size, while θ_i^n , $\dot{\theta}_i^n$, and $\ddot{\theta}_i^n$ are generalized coordinates, their first and second time derivatives, respectively, at the n th time step. The subscript m denotes the number of correction iterations at each time step which include computations enclosed in the brackets. Here E_r denotes evaluation of coordinates, forces, and all other functions of coordinates, E_a denotes evaluation of generalized accelerations, C_r^q and C_v^q denote correctors of the q th order for generalized coordinates and velocities, respectively, which can be expressed as

$$\theta_i^{n+1} = \theta_i^n + h\dot{\theta}_i^n + h^2 \sum_{l=0}^{q-2} \gamma_l \ddot{\theta}_i^{n-l+1} \quad (2a)$$

and

$$\dot{\theta}_i^{n+1} = \dot{\theta}_i^n + h \sum_{l=0}^{q-1} \lambda_l \ddot{\theta}_i^{n-l+1}. \quad (2b)$$

The coefficients α_l , β_l , γ_l , λ_l for the algorithms used in this study are given in the Appendix for reference.

If a microcanonical ensemble is considered the energy of the system must be constant as well as the total momentum and angular momentum. Suppose the system is described by Cartesian coordinates and Newton's equations are used,

$$m_i \dot{v}_i = \sum_j f_{ij}, \quad (3)$$

where m_i is the mass of the i th particle and f_{ij} is the force exerted by the j th particle (for simplicity only pairwise interactions are considered).

On the $(n+1)$ th step prediction for the momentum of the system gives

$$\sum_i m_i v_i^{n+1} = \sum_i m_i v_i^n + h \sum_i \sum_{l=1}^k \beta_l m_i \dot{v}_i^{n-l+1}. \quad (4)$$

If now we substitute Eq. (3) into the right-hand part of Eq. (4) and take into account the third Newton's law, we obtain

$$\sum_i m_i v_i^{n+1} = \sum_i m_i v_i^n. \quad (5)$$

The same considerations are clearly valid for the correction by Eq. (2a). Thus any of the algorithms defined by Eqs. (1)–(2) regardless of its stability and order conserves the momentum with the accuracy of a computer. The energy, however, even in this simple example is conserved only with the accuracy of the algorithm used, and it is sensitive to accumulation of errors, non-physical parasitic solutions in finite difference equations of high order, etc.

Now consider the same ensemble described by the set of internal coordinates θ_i^n . The general form of equations of motion for ICMD is [1, 3]:

$$\sum_{i=1}^n a_{ki} \ddot{\theta}_i = -\frac{\partial U}{\partial \theta_k} - \sum_{i=1}^n \sum_{m=1}^n b_{kim} \dot{\theta}_i \dot{\theta}_m, \quad k=1, \dots, n, \quad (6)$$

where n is the number of variables, U is the generalized potential, a_{ki} and b_{kim} are coefficients which depend only upon coordinates. It is easy to see that calculation of generalized accelerations from Eq. (6) and substitution into Eqs. (1)–(2) do not result in simple expressions like Eqs. (3)–(5) because of the second term in the right-hand side of Eq. (6). Therefore it can be expected (and it was observed in test calculations) that in the case of Eq. (6) momentum and angular momentum are conserved much worse than when Newton's equations are integrated.

The properties described can be illustrated by the following simple example. Consider a three-atom molecule shown in Fig. 1. It has one harmonic internal degree of freedom introduced by the bond angle θ_7 . According to the general approach described earlier [1] the molecule is integrated into a tree (BKS-tree) which is anchored to the origin of the global coordinate frame. The six generalized variables $\theta_{1, \dots, 6}$ are responsible for the free rotational and translational degrees of freedom.

Figure 2 presents the results of numerical integration of Eqs. (6) for this molecule with zero initial velocities. The time step for this example is chosen so that nearly 10 steps cover one period of the internal oscillation. Such a large

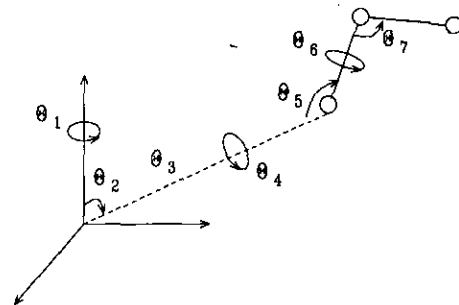


FIG. 1. Three atom molecule in the BKS-tree representation. The first six generalized coordinates of the molecule $\theta_1 - \theta_6$ are indicated. The virtual bond that corresponds to the variable bond length θ_3 is shown by the dash line. The variable bond angle θ_7 is the only internal degree of freedom.

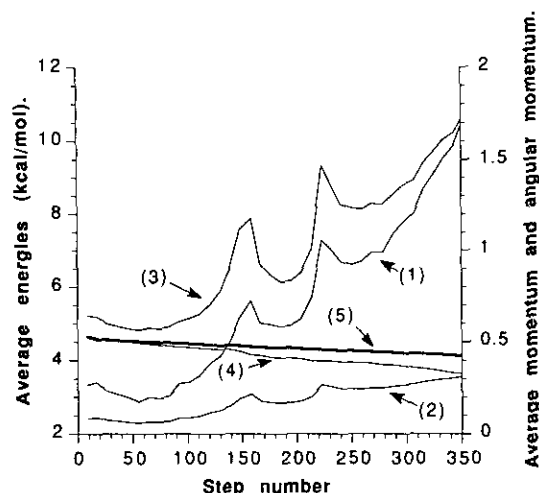


FIG. 2. The results of two ICMD runs of the three atom molecule from the same starting point. In the first run (curves (1)–(4)) all generalized velocities were found from Eq. (6). In the second run (curve (5)) the first generalized velocities $\dot{\theta}_1, \dots, \dot{\theta}_6$ were calculated from the conservation laws. (1) the total momentum; (2) the total angular momentum; (3) the total energy; (4) the total energy of the molecule in the coordinate frame connected with its center of mass; (5) the total energy of the molecule in the second run. Each point corresponds to the average computed over a period equal to one cycle of internal vibration. Atomic masses and the bond bending energy potential were those of the water molecule in the CHARMM force field [19].

time step is taken intentionally to make inaccuracies easily observable. Curves 1–3 demonstrate that the energy, momentum, and angular momentum deviate from initial values, and similar shapes of these curves evidence that the nature of their deviations is the same. The nearly straight line 4 represents the energy of the molecule calculated in the coordinate frame connected with its center of mass, i.e., the energy of the internal oscillation. It is clear that the internal energy is conserved better but it leaks gradually to the external degrees of freedom.

The effects shown in Fig. 2 illustrate in an exaggerated form the distinctive property of generalized internal coordinates: the total momentum, angular momentum, and the energy of a system in the numerical solution behave similarly. This fact is not so much important for statistical properties of the ensemble since with the time step sufficiently small to assure satisfactory conservation of energy all above effects could be ignored. There are, however, certain technical problems connected with non-conservation of momentum and angular momentum. In the case of a long simulation run a molecule or a system of molecules studied as a whole acquires non-zero velocity and moves away from the origin of the global coordinate frame; i.e., the virtual bond θ_3 in Fig. 1 is lengthened. Since the atomic coordinates are calculated by constructing the BKS-tree from the bottom to the top the calculation of interatomic distances and some other values in the system encounter the problem of

roundoff errors when θ_3 is much greater than the smallest interatomic distance. Similar difficulties are involved in calculations of atomic velocities.

Thus it is highly desirable to find a possibility to fix the momentum and angular momentum of the system while integrating Eq. (6). The tree-like topology of the BKS-tree model presents a relatively simple solution of the problem because there are six generalized velocities θ_1 – θ_6 that correspond to six external degrees of freedom. These velocities can be found from the laws of conservation of the moment and angular moment when all other generalized coordinates and velocities responsible for internal motions have been already determined. One can propose therefore the following scheme of calculations: (i) Eqs. (1a) and (2a) are used for predicting and correcting generalized coordinates; (ii) Eqs. (1b) and (2b) are used for predicting and correcting $\dot{\theta}_7$ – $\dot{\theta}_n$ only, while the first six velocities are calculated at each iteration from the conservation laws after $\dot{\theta}_7$ – $\dot{\theta}_n$ have been predicted (corrected).

Note that this approach is not equivalent to the transition to the coordinate frame connected with the center of mass of the molecule and its principal axes. In fact it is similar to substitution of the first six equations in system (6) for six scalar algebraic equations given by the conservation laws. Since the first six generalized velocities are involved in the second term in the right-hand side of Eq. (6) their values influence all generalized accelerations. Therefore the trajectories of the internal motion calculated in this way will differ from those computed by a straightforward integration. We have found that this has a considerable positive effect on the accuracy and stability of the numerical integration.

Curve 5 in Fig. 2 presents the time dependence of the total energy of the three-atom molecule computed by the above method. It does not coincide with curve 4 for the internal energy in the previous run and exhibits much less drift. The root-mean-square fluctuation (rmsf) of the total energy is two times less for curve 5 than for curve 4.

Figure 3 demonstrates the same effect in a more realistic case of an ICMD simulation of an α -helical peptide with fixed covalent structure (the model is described in Section 4). Here fast atomic collisions give the main source of errors in the solution. One can see in Fig. 3 that a relatively fast collision was encountered between steps 48 and 58 of the trajectory, which causes a temporary loss of accuracy. This results in a small discontinuity in the internal energy computed by the straightforward method but the total energy computed by the modified approach is conserved much better. It was found in many tests that the forced conservation of momentum and angular momentum by this method usually makes possible a significant increase in the time step.

The effect considered above is not obvious and can be interpreted in the following way. It is known that inaccuracies that appear during the numerical integration of

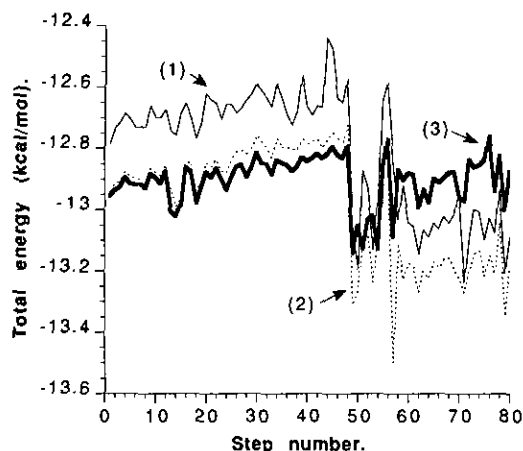


FIG. 3. Results of the two ICMD runs of model 3 of the α -helix (see Section 4) from the same starting point. In the first run (curves (1) and (2)) all generalized velocities have been found from Eq. (6). In the second run (curve (3)) the first generalized velocities $\theta_1, \dots, \theta_6$ were calculated from the conservation laws: (1) the total energy of the molecule; (2) the total energy of the molecule in the coordinate frame connected with its center of mass; (3) the total energy of the molecule in the second run.

dynamic equations with large time steps, and also in the course of a long simulation are largely connected with non-physical parasitic solutions of finite difference equations [5]. Their influence is particularly significant when the order of the finite difference method is higher than the order of the corresponding dynamic equations. The trajectory calculated is a superposition of all fundamental solutions of the numerical algorithm, with the relative weights of non-physical contributions being small initially due to correct initial conditions. Correction of the first six generalized velocities introduces a perturbation such that the relative weights of the parasitic solutions that badly conserve momentum, angular momentum, and total energy, are decreased. Consequently the relative weight of the "physical" solution is increased, which results in better conservation of the total energy.

3. OPTIMAL STRATEGY OF CORRECTIONS

The numerical algorithms used in Newtonian MD normally involve only predictor steps without corrections [4–7]. In this case, correction of velocities has little or no effect on the trajectory since forces and, therefore, accelerations do not depend upon them. Several corrections of coordinates require recalculation of forces, which is too computationally costly and cannot be compensated by a possible increase in the time step.

The situation is clearly different for ICMD. When generalized velocities are corrected by Eq. (2b) generalized accelerations must be recalculated from Eq. (6). This should not be very costly if coordinates remain the same because

TABLE I

Effect of Correction Steps of Numerical Algorithms on Calculated ICMD Trajectories

Method	h fs	δ^a	Drift ^b	Rmsf E_{tot}	Time ^c
$P_r^5 P_v^5 E, E_a$	15 58		0.0024	0.026	1
$P_r^5 P_v^5 E, (E_a C_v^5)_4$	15 287		-0.27×10^{-4}	0.005	1.2
$P_r^5 P_v^5 (E, E_a C_r^5 C_v^5)_4$	15 1244		-0.26×10^{-5}	0.0013	3.5
$P_r^5 P_v^5 E, E_a$	20 11		0.007	0.9	1
$P_r^5 P_v^5 E, (E_a C_v^5)_4$	20 116		-0.16×10^{-3}	0.014	1.2
$P_r^5 P_v^5 (E, E_a C_r^5 C_v^5)_4$	20 72		0.24×10^{-3}	0.021	3.5

^a Quality of integration δ is defined as $\delta = \langle \Delta E_{\text{kin}}^2 \rangle / \langle \Delta E_{\text{tot}}^2 \rangle$, where $\langle \Delta E_{\text{kin}}^2 \rangle$ denotes rmsf of the kinetic energy and $\langle \Delta E_{\text{tot}}^2 \rangle$ denotes rmsf of the total energy.

^b The drift is calculated by a least squares fit of the total energy to a linear function of time.

^c Relative time required for one step of integration in arbitrary units.

the matrix $\{a_{ij}\}$ and the gradients depend only on coordinates, whereas the calculation of the second term in the right-hand side of Eq. (6) takes a few percent of the total computer time. Although the use of only velocity corrections without correcting coordinates cannot be supported by strong theoretical arguments they turn out to improve significantly the accuracy of calculations.

Table I presents some characteristic results obtained by three techniques: with predictor only, with corrections of generalized velocities and with corrections of both coordinates and velocities. As in Fig. 3 the α -helical peptide with fixed covalent structure was used for these tests (details of the calculations are given in Section 4). Table I shows that corrections of generalized velocities decrease the rmsf of the total energy and increase the quality of integration δ while the computer time per time step is only slightly larger. When coordinates are also corrected the accuracy is further improved in the first case and remains similar in the second, but this costs much more computer time.

We tested several different routines for corrections: (i) $P^k (EC^k)_m$; (ii) $P^k (EC^k)_m E$; (iii) $P^k (EC^{k+1})_m$; (iv) $P^k (EC^{k+1})_m E$, which are recommended in the literature [10] and tried to optimize the number of iterations of correction. Normally 1–2 iterations are sufficient while 3–4 iterations guarantees the convergence. Since the computer time required for one iteration is relatively small, $m = 4$ can be recommended as a reasonable number. With this number of iterations the particular sequence of (i)–(iv) does not play a role.

4. COMPARISON OF MOLECULAR MODELS AND NUMERICAL ALGORITHMS

The conventional MD method operates essentially with two molecular models: (i) an unconstrained molecule with

$3N$ degrees of freedom, where N is the number of atoms, and (ii) a molecule with bond lengths fixed by imposing distance constraints [11, 12]. It was demonstrated that optimal numerical algorithms differ for these two models [13]. Since any polymeric molecule in the ICMD can be described by a large variety of models with widely different degree of fixation, one can expect to encounter a similar situation here. Because of this reason below we compare three models of one molecule.

As we wanted to cover a range of models that includes the representation of constrained Newtonian MD and because of the large number of tests required, we were limited in the size of the object. We took an oligopeptide fragment (Ala)₉ in the α -helical conformation as an example because this is a fragment of a common protein secondary structure which includes all sorts of interactions normally considered for proteins.

The first model is similar to molecular models with fixed bond lengths often employed at present; it can be used for comparisons:

— *Model 1.* CH₃ groups are represented as united atoms while all other atoms are considered explicitly. Bond angles and torsion angles are used as variables (the total number of variables equals 101).

The other two models correspond to the degree of fixation which in practice cannot be obtained by the conventional MD with constraints:

— *Model 2.* All atoms are treated explicitly and only torsion angles are variable (42 variables).

— *Model 3.* Here unlike in the second one united atoms are used instead of CH₃ groups and the torsion angle that rotates the hydroxyl group at the C-terminus is fixed (32 variables). In this model the highest frequency components of the molecular motion in torsion angles are removed: the χ_1 angles of alanines and the angle ω that rotates the hydroxyl group at the C-terminus. In both cases the highest frequency is connected with rotation of hydrogen atoms. Note that the N-terminus amide group does not introduce permanent fast oscillations into the system because in the ECEPP force field there is no torsion potential on the ϕ angles of amino acids. This group, however, presents the main source of fast collisions that limit the accuracy of calculations for this model.

The computer time per one step integration differs as 1.7:1.1:0.8 s for those three models, respectively, on the computer ES1061. In the first case the solution of the linear algebraic system with respect to $\dot{\theta}_i$ by the Kholezki algorithm [14] is most time consuming, whereas for the other two models this time is smaller than the calculation of gradients.

The energy function is composed of the terms that represent bond-angle bending, sinusoidal dihedral torsions, and

Van der Waals, electrostatic (Coulomb), and hydrogen-bond interactions. We use empirical potentials compiled from those of ECEPP [15–17], UNIEPP [18], and CHARMM [19], which seem quite satisfactory for numerical tests of integration techniques. In particular, in the first model the bond angle bending terms were taken from the CHARMM force field and ECEPP parameters were used for other terms. For the last two models all the energy terms were from ECEPP potentials while the united atom parameters were taken from the UNIEPP system. The interaction between explicit and united atoms was described by the standard ECEPP expressions; i.e., no qualitative distinction between united and explicit atoms was made.

Equations of motion for ICMD involve a difficulty similar to that involved in the description of motion of a rigid body (see Fig. 1): when the first virtual bond $\theta_3 = 0$ or the planar angles θ_2 or θ_5 equal zero or π the set of the first six coordinates degenerates, which results in the zero value of determinant of the matrix $\{a_{ik}\}$ in Eq. (6). Although this problem is usually solved by using quaternions instead of polar coordinates [4] such an approach is too complicated for long flexible polymers. In our tests we simply checked the values of θ_2 , θ_3 , and θ_5 at each step and repositioned the origin of the global coordinate frame when necessary. Note that in combination with the forced fixation of momentum and angular momentum considered above this does not require any additional calculations.

All test runs start from an equilibrium configuration of the molecule at a temperature of about 300 K. An appropriate starting configuration was obtained as follows: We started with the approximate α -helical structure (ϕ , $\psi = -60^\circ$, -60°) and minimized total potential energy by use of both conjugate gradient method and the tunnel algorithm [20]. The energy was minimized until the decrease was less than 10 kcal/mole per step. Then the molecule was heated in a series of short MD runs with temperature fixation [21]. After that the system was equilibrated during 10 ps of the MD simulation with the temperature fixation used at the beginning. This time was sufficient for the mean fluctuation of the kinetic energy $\langle \Delta K \rangle$ to be stabilized. All methods were compared on the same trajectory which was 1 ps for the first model and 3 ps long for the other two. The duration for the first model was chosen for comparison with the data obtained by the procedure SHAKE [13, 22]. It was necessary to increase the time interval for the last two models because of larger time steps.

Figures 4 and 5 present some characteristic results obtained in calculations with models 2 and 3 of the α -helix with fixed bond angles. In all cases the rmsf of the total energy regularly grows with the time step. The algorithms of higher order give better results with small time steps but their accuracy rapidly falls and for maximum steps lower-order methods can give similar or even better results.

This is a normal pattern which is discussed in textbooks

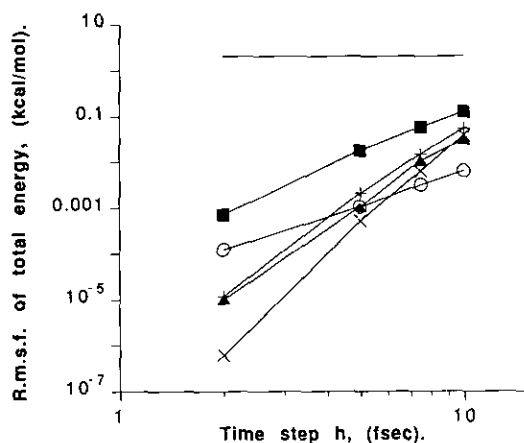


FIG. 4. Results of ICMD simulations of a 3-ps trajectory for model 2 of (Ala)₉ by different methods. The upper solid line indicates the level of r.m.s.f. of the kinetic energy. The algorithms applied are $P^k P^k E_r (E_a C^k)_4$ with $k=2$ (○), 3 (■), 4 (+), 5 (▲), 6 (×). The notation is explained in the text.

[4, 10]. A careful reader, however, may note that earlier it was observed only for unconstrained molecules. It was shown that when bond lengths are constrained the accuracy does not increase with the order of the algorithm, even for small step size, supposedly because of the different nature of the forces that limit the accuracy of calculations for free and constrained molecules [13]. In the first case these forces are regular and predictable, whereas in the second they are more stochastic and, consequently, less predictable. The increased order of the algorithm obviously cannot give better accuracy if the forces are stochastic [13].

The above interpretation evidently disagrees with the results presented in Figs. 4 and 5 because one can expect that the forces in models 2 and 3 are much more "stochastic"

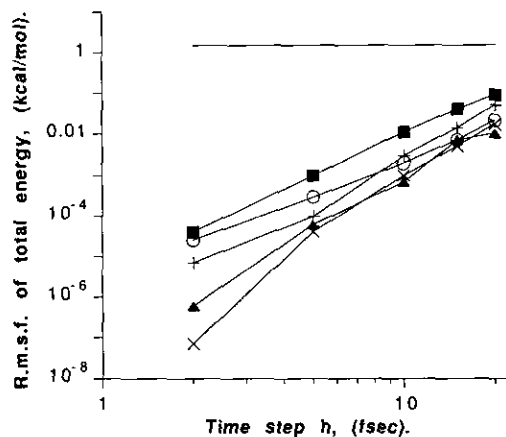


FIG. 5. Results of ICMD simulations of a 3-ps trajectory for model 3 of (Ala)₉ by different methods. All symbols are same as in Fig. 4.

than in models with only bond lengths fixed. Moreover, we have obtained a pattern similar to Fig. 4 for model 1 (results not shown). It is not easy to explain this discrepancy because the ICMD and constrained Newtonian dynamics should give the same trajectory for an equally constrained system. We believe that it is not connected with the difference in the empirical potentials and the smaller size of the molecule we took for our tests. It may be suggested that with small time steps the total accuracy of integration for an iterative algorithm like SHAKE can be limited by the convergence of this routine which in turn may vary from one integration algorithm to another. In this case the total accuracy would not increase with the time step decreased. Although the possible step of numerical integration is generally determined by the characteristics of forces in the system, one may note that strictly speaking all forces given by empirical potentials are equally predictable except artificial perturbations caused by truncation of long-range potentials.

We suppose that the patterns presented in Figs. 4 and 5 simply reflect the order of the truncation error in the corresponding finite difference approximations. Thus the plots for higher-order methods start lower but then they intersect the plots for lower-order methods and the only question is whether this occurs within the range of stability or required accuracy of calculations.

Another interesting aspect of these results is an apparent difference in the course of the curves in Figs. 4 and 5. In Fig. 4 the second-order algorithm yields the best accuracy even for medium step size while in Fig. 5 the fifth-order method looks better. This difference is caused by the intrinsic properties of the two models.

In the first case (Fig. 4) the highest-frequency components of the motion are presented by the fast oscillations of the torsion angles listed above. They cover all of the trajectory and each of its parts equally contributes to the total error. In the second case (Fig. 5) all the high frequency oscillations in molecular motion are frozen and the highest frequency component of the motion is connected with rapid collisions. One example of such a collision can be found in Fig. 3 above. The dynamics of a particular collision depends on the masses and moments of inertia of the colliding rigid bodies and its probability depends on the conformation of the molecule. Usually the probability of very fast collisions is relatively small and they cover only a small part of the trajectory. Thus for most of the trajectory for model 3 even the maximal step size can be considered as relatively small and, therefore, higher-order algorithms yield better results.

One may note that these properties of model 3 are similar to those of Lennard-Jones (LJ) liquids. However, unlike LJ liquid of a large number of similar particles, a polymer in ICMD consists of different rigid bodies and for a particular conformation only few of them can produce critically fast collisions.

As a result of the above qualitative picture the accuracy of ICMD simulation of not-very-large molecules sometimes is trajectory dependent. The same algorithm with the same time step can yield excellent conservation of energy on one trajectory and significantly worse on another. Moreover, sometimes better results can be obtained with larger time steps on one and the same trajectory. One such example is presented in Fig. 6. Slightly better conservation of E_{tot} is obtained with the time step 20 fs than with 15 fs for model 3. This results from small shifts of the collision times with respect to the grid points of the corresponding finite difference equation.

From the plots presented in Figs. 4 and 5 the possible time steps for ICMD simulations of models 2 and 3 of the α -helix can be estimated as 10 fs and 20 fs, respectively with the quality of integration $\delta = 75$ in the first case and $\delta = 400$ in the second. For two times longer trajectory the value of δ is only 20% lower and the algorithms do not exhibit instabilities on trajectories of tens of picoseconds.

In Fig. 6 the results obtained by the method $P_r^5 P_v^5 E_r (E_a C_v^5)_2$ for all three models are compared. The main goal of this study was to find optimal numerical algorithm for ICMD. The results shown in Fig. 6 allow us to believe that the accuracy obtained with this method is close to the possible maximum. On the one hand, for model 1 with fixed bond lengths and with the time step 4 fs this algorithm results in the quality of integration $\delta = 15$. An approximately similar value was reported for constrained Newtonian MD of a protein with the same time step [22]. On the other hand, for model 3 the maximal time step is 15–20 fs which is similar to the values reported for LJ liquids with comparable molecular weight of particles [7]. It is clear that only an approximate comparison of this kind is possible because of the differences in molecular models and sets of potentials used.

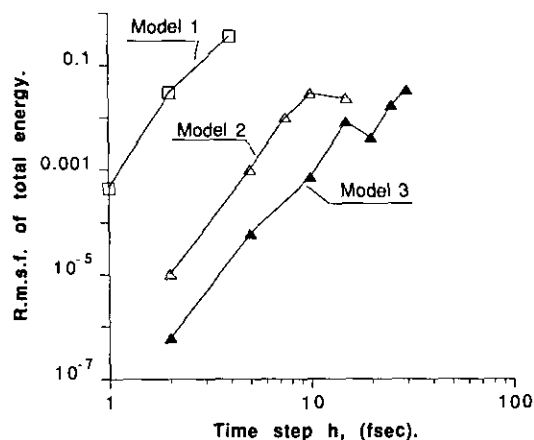


FIG. 6. The results of ICMD simulations for three models of (Ala)_n obtained by the algorithm $P_r^5 P_v^5 E_r (E_a C_v^5)_2$. The lengths of the trajectories are 1-ps for model 1 and 3-ps for the other two.

5. DISCUSSION

In recent decades MD has become a common method for simulation of flexible molecules. The problem of the elimination of fast degrees of freedom introduced by the empirical potentials that maintain the covalent structure of molecules was considered many times in two different aspects.

The first aspect is a theoretical one. It is known that constraints exert an influence on the statistical properties of the ensemble [23] and there are different points of view on which the mode of constraints should be preferred [22, 24–25]. In many cases, however, even a qualitative simulation of slow concerted movements of large molecules presents great interest. This explains the growing interest in the methods which can “freeze” large portions of molecular structures [1, 26, 27].

The second aspect is a practical one. Imposing constraints implies a trade between the computer time spent on additional computations and a possible increase in the step size. When constraints are imposed explicitly by the methods used in Newtonian MD [11, 12], more constraints require more computations. Therefore bond length constraints are used routinely but the same approach requires much more effort and its algorithms converge poorly when the bond angles should be fixed [22]. It was argued also that only bond-length constraints are efficient while additional fixations cannot give an increase in the step size because only frequencies of bond length oscillations are well separated from the frequencies of other vibrations in molecules [22].

The method of ICMD used in this study leaves the first of the above practical considerations aside because in this method constraints are imposed implicitly and more constraints means much less calculation. In addition, Fig. 6 clearly demonstrates the possibility of increasing the time step of ICMD simulations by changing the model of the molecule. This conclusion apparently disagrees with earlier estimates [22] but the disagreement is explained easily. Fixation of bond angles really does not give a great increase in the step size compared to fixation of only bond lengths because many torsion angles rotate with comparable frequencies. But specific selection and freezing of fast degrees of freedom, which is easy in ICMD, can give a significant increase in the step size.

The data presented in Section 4 include only characteristic results of the best algorithms given by Eqs. (1)–(2) and coefficients listed in the Appendix. Other computational schemes, in particular, Gear predictor–corrector algorithms [28], were also studied and similar results were obtained. We suppose that the most important factors that influence the accuracy and stability of computations are the forced conservation of momentum and angular momentum and the optimal strategy of correction discussed in

Sections 2 and 3, respectively. The choice of a particular computational method plays a minor role.

We believe that the above conclusions are determined by the general form of equations of motion (6) and they should be generally valid for non-Newtonian MD. The estimates of possible time steps obtained in this study probably can be extrapolated to larger molecular systems although additional tests are required to take into account specific effects introduced by the truncation of non-bonded potentials.

APPENDIX

Tables A1–A4 include the coefficients of predictor–corrector algorithms used in this study.

TABLE A1
Prediction of Generalized Coordinates [7]

Order ^a	α_1	α_2	α_3	α_4	α_5
2	$\frac{1}{2}$				
3	$\frac{2}{3}$	$-\frac{1}{6}$			
4	$\frac{19}{24}$	$-\frac{5}{12}$	$\frac{1}{8}$		
5	$\frac{323}{360}$	$-\frac{11}{15}$	$\frac{53}{120}$	$-\frac{19}{180}$	
6	$\frac{1427}{1440}$	$-\frac{133}{120}$	$\frac{241}{340}$	$-\frac{173}{360}$	$\frac{3}{32}$

^a The order of the method is defined as the order of the truncation.

TABLE A2
Correction of Generalized Coordinates [7]

Order	γ_0	γ_1	γ_2	γ_3
2	$\frac{1}{2}$			
3	$\frac{1}{6}$	$\frac{2}{6}$		
4	$\frac{3}{24}$	$\frac{10}{24}$	$-\frac{1}{24}$	
5	$\frac{38}{360}$	$\frac{171}{360}$	$-\frac{36}{360}$	$\frac{7}{360}$

TABLE A3
Prediction of Generalized Velocities [28]

Order	β_1	β_2	β_3	β_4	β_5	β_6
2	$\frac{3}{5}$	$-\frac{1}{2}$				
3	$\frac{23}{12}$	$-\frac{4}{3}$	$\frac{5}{12}$			
4	$\frac{55}{24}$	$-\frac{59}{24}$	$\frac{37}{24}$	$-\frac{9}{24}$		
5	$\frac{1901}{720}$	$-\frac{2774}{720}$	$\frac{2616}{720}$	$-\frac{1274}{720}$	$\frac{251}{720}$	
6	$\frac{4277}{1440}$	$-\frac{7929}{1440}$	$\frac{9982}{1440}$	$-\frac{7298}{1440}$	$\frac{2877}{1440}$	$-\frac{475}{1440}$

TABLE A4

Correction of Generalized Velocities [28]

Order	λ_0	λ_1	λ_2	λ_3	λ_4	λ_5
2	$\frac{1}{2}$	$\frac{1}{2}$				
3	$\frac{5}{12}$	$\frac{8}{12}$	$-\frac{1}{12}$			
4	$\frac{9}{24}$	$\frac{19}{24}$	$-\frac{5}{24}$	$\frac{1}{24}$		
5	$\frac{251}{720}$	$\frac{646}{720}$	$-\frac{254}{720}$	$\frac{106}{720}$	$-\frac{19}{720}$	
6	$\frac{475}{1440}$	$\frac{1427}{1440}$	$-\frac{798}{1440}$	$\frac{482}{1440}$	$-\frac{173}{1440}$	$\frac{27}{1440}$

REFERENCES

1. A. K. Mazur and R. A. Abagyan, *J. Biomol. Struct. Dyn.* **6**, 815 (1989).
2. R. A. Abagyan and A. K. Mazur, *J. Biomol. Struct. Dyn.* **6**, 833 (1989).
3. A. K. Mazur, V. E. Dorofeyev, and R. A. Abagyan, *J. Comput. Phys.* **92**, 261 (1991).
4. M. P. Allen and D. T. Tildesley, *Computer Simulation of Liquids* (Clarendon Press, Oxford, 1987).
5. R. W. Hockney and J. W. Eastwood, *Computer Simulation Using Particles* (McGraw-Hill, New York, 1981).
6. L. Verlet, *Phys. Rev.* **159**, 98 (1967).
7. D. Beeman, *J. Comput. Phys.* **20**, 130 (1976).
8. M. Gottlieb, *Comput. & Chem.* **1**, 155 (1977).
9. K. Refson and G. S. Pawley, *Mol. Phys.* **61**, 669 (1987).
10. G. Hall and J. M. Watt, *Modern Numerical Methods for Ordinary Differential Equations* (Clarendon Press, Oxford, 1976).
11. J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, *J. Comput. Phys.* **23**, 327 (1977).
12. G. Ciccotti and J. P. Ryckaert, *Comput. Phys. Rep.* **4**, 345 (1986).
13. W. F. van Gunsteren and H. J. C. Berendsen, *Mol. Phys.* **34**, 1311 (1977).
14. S. Pissarnezky, *Sparse Matrix Technology* (Academic Press, New York, 1984).
15. F. A. Momany, R. F. McGuire, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.* **79**, 2361 (1975).
16. G. Nemethy, M. S. Pottle, and H. A. Scheraga, *J. Phys. Chem.* **87**, 1883 (1983).
17. M. J. Sippl, G. Nemethy, and H. A. Scheraga, *J. Phys. Chem.* **88**, 6231 (1984).
18. L. G. Dunfield, A. W. Burgess, and H. A. Scheraga, *J. Phys. Chem.* **82**, 2609 (1978).
19. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4**, 187 (1983).
20. A. V. Levy and A. Montalvo, *SIAM J. Sci. Stat. Comput.* **6**, 15 (1985).
21. J. Aqvist, W. F. van Gunsteren, M. Leijonmarck, and O. Tapia, *J. Mol. Biol.* **183**, 461 (1985).
22. W. F. van Gunsteren and M. Karplus, *Macromolecules* **15**, 1528 (1982).
23. M. Fixman, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 3050 (1974).
24. N. Go and H. A. Scheraga, *J. Chem. Phys.* **51**, 4751 (1969).
25. N. G. van Kampen, *Phys. Rep.* **124**, 69 (1985).
26. J. W. Perram and H. G. Petersen, *Mol. Phys.* **65**, 861 (1988).
27. K. D. Gibson and H. A. Scheraga, *J. Comput. Chem.* **11**, 468 (1990).
28. C. W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations* (Prentice-Hall, New York, 1971).