



Published in final edited form as:

*J Phys Chem B*. 2005 July 21; 109(28): 13785–13797. doi:10.1021/jp058008o.

## Molecular Dynamics with the United-Residue Model of Polypeptide Chains. I. Lagrange Equations of Motion and Tests of Numerical Stability in the Microcanonical Mode

Mey Khalili<sup>†</sup>, Adam Liwo<sup>†,‡</sup>, Franciszek Rakowski<sup>§,||</sup>, Paweł Grochowski<sup>§</sup>, and Harold A. Scheraga<sup>\*,†</sup>

<sup>†</sup> Baker Laboratory of Chemistry and Chemical Biology, Cornell University, Ithaca, New York 14853-130

<sup>‡</sup> Faculty of Chemistry, University of Gdańsk, Sobieskiego Str. 18, 80-952 Gdańsk, Poland <sup>§</sup> Interdisciplinary Centre for Mathematical and Computational Modelling, University of Warsaw, Pawińskiego Str. 5a, 02-106 Warsaw, Poland <sup>||</sup> Faculty of Physics, Warsaw University of Technology, Koszykowa Str. 75, 00-662 Warsaw, Poland

### Abstract

The Lagrange formalism was implemented to derive the equations of motion for the physics-based united-residue (UNRES) force field developed in our laboratory. The  $C^{\alpha}\cdots C^{\alpha}$  and  $C^{\alpha}\cdots SC$  (SC denoting a side-chain center) virtual-bond vectors were chosen as variables. The velocity Verlet algorithm was adopted to integrate the equations of motion. Tests on the unblocked Ala<sub>10</sub> polypeptide showed that the algorithm is stable in short periods of time up to the time step of 1.467 fs; however, even with the shorter time step of 0.489 fs, some drift of the total energy occurs because of momentary jumps of the acceleration. These jumps are caused by numerical instability of the forces arising from the  $U_{\text{rot}}$  component of UNRES that describes the energetics of side-chain-rotameric states. Test runs on the Gly<sub>10</sub> sequence (in which  $U_{\text{rot}}$  is not present) and on the Ala<sub>10</sub> sequence with  $U_{\text{rot}}$  replaced by a simple numerically stable harmonic potential confirmed this observation; oscillations of the total energy were observed only up to the time step of 7.335 fs, and some drift in the total energy or instability of the trajectories started to appear in long-time (2 ns and longer) trajectories only for the time step of 9.78 fs. These results demonstrate that the present  $U_{\text{rot}}$  components (which are statistical potentials derived from the Protein Data Bank) must be replaced with more numerically stable functions; this work is under way in our laboratory. For the purpose of our present work, a nonsymplectic variable-time-step algorithm was introduced to reduce the energy drift for regular polypeptide sequences. The algorithm scales down the time step at a given point of a trajectory if the maximum change of acceleration exceeds a selected cutoff value. With this algorithm, the total energy is reasonably conserved up to a time step of 2.445 fs, as tested on the unblocked Ala<sub>10</sub> polypeptide. We also tried a symplectic multiple-time-step reversible RESPA algorithm and achieved satisfactory energy conservation for time steps up to 7.335 fs. However, at present, it appears that the reversible RESPA algorithm is several times more expensive than the variable-time-step algorithm because of the necessity to perform additional matrix multiplications. We also observed that, because Ala<sub>10</sub> folds and unfolds within picoseconds in the microcanonical mode, this suggests that the effective (event-based) time unit in UNRES dynamics is much larger than that of all-atom dynamics because of averaging over the fast-moving degrees of freedom in deriving the UNRES potential.

\* Author to whom correspondence should be addressed. Phone: (607) 255-4034. Fax: (607) 254-4700. E-mail: has5@cornell.edu.

## 1. Introduction

This paper describes an effort to increase the time scale of molecular dynamics (MD) so as to be able to simulate protein folding and unfolding<sup>1–3</sup> in real time. Atomistically detailed simulations with explicit inclusion of solvent are limited at present to studying unfolding pathways of native proteins starting from the native state followed by subsequent refolding<sup>2–4</sup> or by the umbrella-sampling method in which the change of selected reaction coordinates (usually the fraction of native contacts and the radius of gyration) is controlled along the folding pathway.<sup>4</sup> The reasons for these limitations are (i) the high computational cost of energy evaluation and (ii) the short values of the time step of molecular dynamics simulations, which are on the order of femtoseconds, while the shortest folding times of proteins are on the order of microseconds.<sup>5</sup> Nevertheless, in connection with experimental data, atomistically detailed unfolding simulations provide valuable insights into folding pathways.<sup>3</sup> A stochastic difference equation approach has also been devised recently to study folding pathways at the all-atom level; however, this method requires a priori knowledge of both the unfolded and the folded state.<sup>6,7</sup> Ab initio folding simulations of even small proteins are still rare. One example of such a successful simulation is that performed by Duan and Kollman on the 36-residue villin headpiece;<sup>8</sup> they observed short-lived folding intermediates in a 1- $\mu$ s-long run. However, the advent of distributed computing promises that this might improve soon.<sup>9</sup> When the solvent is considered at the mean-field level, larger systems can be treated in real time; a recent example is the ab initio folding simulation of the B-domain of staphylococcal protein A carried out by Jang et al.<sup>10</sup> with the AMBER force field<sup>11,12</sup> and the generalized Born (GB) model<sup>13,14</sup> of water that treats this solvent as a high-dielectric continuum.

Reduced (mesoscopic) models of proteins, in which each amino acid residue is represented by only a few interaction sites, in principle offer further extensions of the time scale of simulations. However, although simulations with reduced models have contributed significantly to our understanding of the events that occur in the folding process, the interaction potentials are not usually meant to reproduce the detailed features of the protein energy surface, but rather to study general properties of folding.<sup>15–18</sup> Quite often the interaction potentials are intentionally biased toward the experimental structure (the G $\bar{o}$ -like models).<sup>19–21</sup> Reduced models, applied with some success to folding simulations of real proteins by using molecular dynamics, can be termed semimezoscopic because the backbone is represented at the all-atom level, while the side chains are treated as united interaction sites.<sup>22–27</sup> Monte Carlo dynamics on high-resolution lattices is also used together with knowledge-based potentials for protein-structure predictions;<sup>28–31</sup> this method, however, does not implement Newton's (or Lagrange's) equations of motion. Recently, Rudnicki et al.<sup>32</sup> designed a molecular dynamics algorithm based on the Lagrange formalism<sup>33</sup> for their physics-based mesoscopic model of nucleic acids.<sup>34</sup>

For the past several years, we have been developing a physics-based united-residue force field, hereafter referred to as UNRES.<sup>35–49</sup> Each amino acid residue is represented by only two interaction sites, which makes the model simple enough to carry out large-scale simulations. In comparison to other mesoscopic protein models, UNRES offers two important advantages. First, the force field has been derived carefully as a potential of the mean force of the united-residue chain. This treatment enabled us to rationally introduce multibody terms that do not have counterparts in all-atom force fields and thus far were derived in a heuristic manner.<sup>28, 30,31</sup> The multibody terms are indispensable for reproducing the correct geometry of regular secondary structures, such as  $\alpha$ -helices and  $\beta$ -sheets in unrestricted simulations. Second, our recently developed method of optimization of the UNRES force field<sup>44,47–49</sup> is based on a hierarchical concept of the protein energy landscape<sup>25</sup> and, thereby, reflects the hierarchy of folding events of the training proteins. In connection with the efficient conformational space annealing (CSA) method<sup>50,51</sup> of global optimization, UNRES is able to predict the structures

of real-size proteins without ancillary information from structural databases.<sup>39,43,48,49,52,53</sup> For all these reasons, UNRES appears to be a good mesoscopic force field with which the folding pathways of proteins can be studied in real time. Therefore, our present work is aimed at implementation of the UNRES force field in a molecular dynamics algorithm. As did Rudnicki et al.,<sup>32</sup> we made use of the general Lagrange formalism. In this paper, we describe the extension of the UNRES model to molecular dynamics simulations and initial tests of the stability of the algorithm run in the microcanonical mode. In the accompanying paper,<sup>54</sup> we introduce Langevin dynamics and perform initial tests of folding the model Ala<sub>10</sub> polypeptide and the B-domain of staphylococcal protein A. On the basis of these tests, we relate the time scale of UNRES dynamics to that of all-atom dynamics. More extensive tests on more proteins have appeared in a separate paper.<sup>55</sup>

## 2. Theory

### 2.1. United-Residue Force Field

In the UNRES model,<sup>35–49</sup> a polypeptide chain is represented as a sequence of  $\alpha$ -carbon atoms ( $C^\alpha$ 's). The  $C^\alpha$ 's are linked together by virtual bonds (designated as dCs), which constitute the backbone. United side chains (SCs) are connected to the backbone by the virtual bonds (dXs). United peptide groups (p's) are in the center of the virtual bonds. The centers of mass of the side chains are at the ends of the dXs (Figure 1A).

The interaction sites are the united peptide groups in the middle of the dCs, and the united side chains at the ends of the dXs. The p centers represent only the C', O, N, and H atoms of the peptide groups, while the  $C^\alpha$  atoms are included in the SC centers. Consequently, the positions of the  $C^\alpha$  atoms are only geometric points. Until now, in the applications of the UNRES force field, only the side-chain angles ( $\alpha_{SC}$  and  $\beta_{SC}$ ), the virtual-bond angles ( $\theta$ ), and the virtual-bond dihedral angles ( $\gamma$ ) could vary. In this work, we also allow the virtual-bond lengths to vary, which requires introducing additional harmonic potentials (last term of eq 1). The SC centers of glycine residues are located on their  $C^\alpha$  atoms; consequently, there are no virtual  $C^\alpha \cdots SC$  bonds for glycine residues.

The first and the last  $C^\alpha$ 's correspond to the N- and the C-terminal blocking groups, respectively. If there is a peptide group between the first (last) full residue and a blocking group, then the first (last)  $C^\alpha$  is treated as a glycine SC center. To define the ends of a chain in a consistent manner, we introduce dummy blocking groups if the ends are unblocked and are not glycine. If one or both ends is blocked or is a glycine, then the first (last) UNRES residue is a glycine residue, because it has either a  $C^\alpha$  atom or a methyl carbon atom, e.g., of an acetyl group, which effectively serves as a  $C^\alpha$  atom. This is illustrated in Figure 1B and also in Figure 2 in ref 35.

UNRES is a physics-based force field, which is derived as a restricted free energy (RFE) function of a polypeptide chain. The RFE is defined as the free energy of a given coarse-grain conformation obtained by integrating the Boltzmann factor of the all-atom (i.e., the polypeptide chain plus solvent) energy over the degrees of freedom that are neglected in the united-residue model.<sup>38,40,45,46</sup> The complete UNRES potential-energy function is expressed by eq 1.

$$\begin{aligned}
 U = & \sum_j \sum_{i < j} U_{SC_i SC_j} + w_{SCp} \sum_j \sum_{i \neq j} U_{SC_i p_j} + w_{el} \sum_j \sum_{i < j-1} U_{p_i p_j} + \\
 & w_{tor} \sum_i U_{tor}(\gamma_i) + w_{tord} \sum_i U_{tord}(\gamma_i, \gamma_{i+1}) + w_b \sum_i U_b(\theta_i) + \\
 & w_{rot} \sum_i U_{rot}(\alpha_{SC_i}, \beta_{SC_i}) + \sum_{m=2}^{N_{corr}} w_{corr}^{(m)} U_{corr}^{(m)} + w_{vib} \sum_i U_{vib}(d_i)
 \end{aligned}
 \tag{1}$$

The terms  $U_{SC_i SC_j}$  correspond to the mean free energy of hydrophobic (hydrophilic) interactions between the side chains. These terms implicitly contain the contributions from the

interactions of the side chain with the solvent. The terms  $U_{SC_i p_j}$  correspond to the excluded-volume potential of the side chain–peptide group interactions. The terms  $U_{p_i p_j}$  represent the energy of average electrostatic interactions between backbone peptide groups. The terms  $U_{\text{tor}}$  and  $U_{\text{tor}d}$  are the torsional and the double-torsional potentials, respectively, for the rotation about a given virtual bond or two consecutive virtual bonds. These potentials account for intrinsic barriers to rotation about a single and about two consecutive virtual bonds, respectively, and originate in the coupling between the local conformational states of two or three consecutive residues, respectively.<sup>40,46</sup> The terms  $U_b$  and  $U_{\text{rot}}$  are the virtual-angle-bending and side-chain-rotamer potentials, respectively. The terms  $U_{\text{corr}}^{(m)}$  correspond to the correlations (of order  $m$ ) between peptide-group electrostatic and backbone-local interactions. The terms  $U_{\text{vib}}(d_i)$ ,  $d_i$  being the length of the  $i$ th virtual bond introduced in this work, are simple harmonic potentials defined by eq 2.

$$U_{\text{vib}}(d_i) = \frac{1}{2} k_{d_i} (d_i - d_i^{\circ})^2 \quad (2)$$

where  $k_{d_i}$  is the force constant of the  $i$ th virtual bond, currently set at 500 kcal/(mol Å<sup>2</sup>) and  $d_i^{\circ}$  is the average length (corresponding to that used in the fixed-bond UNRES) of the  $i$ th virtual bond, e.g.,  $d_i^{\circ} = 3.8$  Å for a C<sup>α</sup>–C<sup>α</sup> virtual bond corresponding to a trans peptide group.

The terms  $U_{SC_i SC_j}$ ,  $U_b$ , and  $U_{\text{rot}}$  were parametrized<sup>36,37</sup> from the distribution and correlation functions determined from the Protein Data Bank (PDB).  $U_{\text{tor}}$ ,  $U_{\text{tor}d}$ , and  $U_{\text{corr}}$  were based on the cumulant expansion of the RFE of the polypeptide chains<sup>38,40</sup> and parametrized from the RFEs of model systems obtained by high-level MP2/6-31G(d,p) ab initio calculations of model systems.<sup>45,46</sup> Finally, the  $w$ 's are weights of the various energy terms. All of them except  $w_{\text{vib}}$  were obtained in our earlier work<sup>48</sup> by optimization of the potential-energy function to obtain a funnel-like energy landscape<sup>44</sup> of 1IGD (an  $\alpha + \beta$  protein). The resulting force field reproduces the structure of  $\alpha$  and  $\alpha + \beta$  proteins reasonably well.<sup>48</sup> The weight  $w_{\text{vib}}$  was arbitrarily set at 1.

## 2.2. Lagrange Equations of Motion with the UNRES Model

The Cartesian coordinates of the interacting sites in the UNRES model (which do not include C<sup>α</sup>'s) are not sufficient to define the UNRES conformations. Therefore, we adopt the Lagrange formulation of the equations of motion,<sup>33</sup> which is frequently used to describe the time evolution of a classical system characterized by a set of generalized coordinates. For the UNRES model, we chose the virtual-bond vectors  $\mathbf{dC}$ s and  $\mathbf{dX}$ s. The choice of virtual-bond vectors instead of the positions of geometric points (C<sup>α</sup>'s and SC's) was motivated by the fact that the analytical expressions<sup>35,38,40</sup> for the  $U_{\text{pp}}$  and  $U_{\text{corr}}^{(m)}$  terms (eq 1) contain not only distances between the interacting sites but also orientations of the virtual bonds containing the interacting sites and their neighbors. The internal coordinates (the angles  $\theta$ ,  $\gamma$ ,  $\alpha$ , and  $\beta$ ) needed to calculate the torsional ( $U_{\text{tor}}$ ), double-torsional ( $U_{\text{tor}d}$ ), virtual-bond angle-bending ( $U_b$ ), and side-chain-rotamer ( $U_{\text{rot}}$ ) energy terms can also be readily calculated from these vectors. We denote the generalized coordinates by  $\mathbf{q}$  and the corresponding generalized velocities by  $\dot{\mathbf{q}}$ . The vectors  $\mathbf{q}$  and  $\dot{\mathbf{q}}$  are composed of virtual-bond vectors and the associated velocities, respectively, as defined by eq 3.

$$q = \begin{pmatrix} dC_0 \\ dC_s \\ \vdots \\ dC_{e-1} \\ dX_{i_1} \\ dX_{i_2} \\ \vdots \\ dX_{i_m} \end{pmatrix} \quad \text{and} \quad \dot{q} = \begin{pmatrix} d\dot{C}_0 \\ d\dot{C}_s \\ \vdots \\ d\dot{C}_{e-1} \\ d\dot{X}_{i_1} \\ d\dot{X}_{i_2} \\ \vdots \\ d\dot{X}_{i_m} \end{pmatrix} \quad (3)$$

where  $dC_0$  and  $d\dot{C}_0$  specify the position and velocity of the first  $C^\alpha$  atom of the chain, respectively,  $dC_i$  and  $d\dot{C}_i$  specify the  $C_i^{\alpha} \cdots C_{i+1}^{\alpha}$  virtual-bond vector and the associated velocity, respectively, while  $dX_i$  and  $d\dot{X}_i$  specify the  $C_i^{\alpha} \cdots SC_i$  virtual-bond vector and the associated velocity, respectively,  $s$  and  $e$  are the number of the first and the last nondummy  $C^\alpha$  atom, respectively (if the first residue is a dummy, then  $s = 2$ ; otherwise  $s = 1$ ; if the last residue is a dummy, then  $e = n - 1$ ; otherwise  $e = n$ ), while  $i_1, i_2, \dots, i_m$  denote the numbers of non-glycine residues, at which the  $C^\alpha \cdots SC$  virtual-bond vectors are anchored ( $m$  being the number of non-glycine residues). The SCs of glycine residues and nondummy blocking groups are located on the corresponding  $C^\alpha$  atoms, and those sites do not possess a  $dX$  vector (cf. section 2.1). Consequently, the number of  $C^\alpha \cdots SC$  vectors is equal to  $m = n - n_{\text{Gly}}$ , where  $n$  is the number of residues and nondummy blocking groups in the chain and  $n_{\text{Gly}}$  is the number of glycine residues and dummy blocking groups in the chain.

From eq 3, the Cartesian coordinates  $r_{p_i}$  and  $r_{SC_i}$  and velocities  $v_{p_i}$  and  $v_{SC_i}$  of the peptide groups and side chains, respectively, are defined by eqs 4–7.

$$r_{p_i} = \sum_{j=0}^{i-1} dC_j + \frac{1}{2} dC_i \quad (4)$$

$$v_{p_i} = \sum_{j=0}^{i-1} d\dot{C}_j + \frac{1}{2} d\dot{C}_i \quad (5)$$

$$r_{sc_i} = \begin{cases} \sum_{j=0}^{i-1} dC_j & \text{if residue } i \text{ is glycine} \\ \sum_{j=0}^{i-1} dC_j + dX_i & \text{otherwise} \end{cases} \quad (6)$$

$$v_{sc_i} = \begin{cases} \sum_{j=0}^{i-1} d\dot{C}_j & \text{if residue } i \text{ is glycine} \\ \sum_{j=0}^{i-1} d\dot{C}_j + d\dot{X}_i & \text{otherwise} \end{cases} \quad (7)$$

The factor of  $1/2$  in front of the last  $dC_i$  in eq 4 and  $d\dot{C}_i$  in eq 5 comes from the fact that the masses of the peptide groups are located in the middle of the respective virtual bonds.

The definition of the virtual-bond vectors and their relation to the peptide-group and side-chain Cartesian coordinates is illustrated in Figure 1C.

It should be noted that we consider only the UNRES degrees of freedom here, neglecting those that are integrated out when passing from the all-atom to the UNRES potential-energy function.<sup>38,40</sup> Therefore, we assume that the neglected degrees of freedom undergo much faster

motions than the UNRES degrees of freedom.<sup>56</sup> Consequently, the forces acting on the system come from the potential of mean force, i.e., the UNRES energy function defined by eq 1 plus dissipative and friction forces, coming mainly from kinetic-energy exchange with the solvent.

The Lagrangian is defined by eq 8<sup>33</sup>

$$L(\mathbf{q}, \dot{\mathbf{q}}) \equiv K(\mathbf{q}, \dot{\mathbf{q}}) - V(\mathbf{q}) \quad (8)$$

where  $K$  and  $V$  are the kinetic and potential energy of the system, respectively.

For each generalized coordinate, the Lagrangian equations of motion can be written as a system of second-order ordinary differential equations<sup>33</sup> (eq 9).

$$\frac{d}{dt}(\nabla_{\dot{\mathbf{q}}} L) - \nabla_{\mathbf{q}} L = \mathbf{f}^{\text{frict}} + \mathbf{f}^{\text{rand}} \quad (9)$$

where  $\mathbf{f}^{\text{frict}}$  and  $\mathbf{f}^{\text{rand}}$  are the friction (dissipative) and random forces, respectively, acting on the system and  $\nabla_{\mathbf{a}} = (\partial/\partial a_1, \partial/\partial a_2, \dots, \partial/\partial a_M)$ , with  $\mathbf{a}$  being a vector of variables and  $M$  its dimension.

It should be stressed that eq 9 should be considered an assumption, because it is not the exact form that describes the motion of a united-residue chain. Eliminating fast motions from a system of differential equations often results in equations of quite different form than the original ones.<sup>56</sup> However, for example, a system of rigid bodies, which can be considered as a collection of point masses, can still be described by Lagrange equations of motion. Therefore, the assumption seems to be reasonable.

The nonconservative forces are zero when a closed system is treated at the atomistic level (in other words, the total energy of the system is conserved). However, for a united-residue chain immersed in water, there will be significant friction and random forces coming from the exchange of energy and momentum with the solvent molecules as well as random forces arising from the internal degrees of freedom that are integrated out. We leave the introduction of these forces until the accompanying paper.<sup>54</sup> In this initial work, we assume that the right-hand side of eq 9 is zero; i.e., there are no friction or random forces. This certainly does not correspond to a physical situation, but we need an ensemble with a constant energy to test the numerical stability of the algorithm. It should also be noted that, because we average over the fast-moving degrees of freedom, the time unit of the UNRES dynamics is larger than that of the all-atom dynamics as we will demonstrate in the accompanying paper.<sup>54</sup>

The kinetic energy of a polypeptide chain in the UNRES model will consist of two parts: the part arising from the UNRES degrees of freedom ( $K_x$ ) and the average kinetic energy of the hidden degrees of freedom due to averaging corresponding to a given UNRES conformation and momenta ( $K_y$ ). In general, both parts will depend on the coordinates and momenta corresponding to the UNRES degrees of freedom. However, the treatment of the part corresponding to the hidden degrees of freedom is not straightforward, and therefore, in this paper we assume that this kinetic energy,  $K_y$ , is constant, as given by eq 10. Consequently, we consider only  $K_x$ .

$$K = K_x(\mathbf{q}, \dot{\mathbf{q}}) + K_y(\mathbf{q}, \dot{\mathbf{q}}) \approx K_x(\mathbf{q}, \dot{\mathbf{q}}) + \text{constant} \quad (10)$$

The right-hand side of eq 10 is exact if it is assumed (i) that the transformation from one UNRES conformation to another one is infinitely slow compared to the motion of the “fast” degrees of freedom and (ii) that, at each point of the UNRES trajectory, the UNRES degrees of freedom act as constraints.<sup>57</sup> However, while condition i can be considered to be reasonably satisfied,

the UNRES degrees of freedom should be considered as average values of the respective coordinates rather than exact constraints. Such a case belongs to the third category of variable separations according to van Kampen,<sup>56</sup> where the slow variables correspond to the coordinates of “guiding centers”. We, therefore, treat eq 10 as an assumption.

Because the degrees of freedom corresponding to the rotation about the  $C^{\alpha}\cdots C^{\alpha}$  and  $C^{\alpha}\cdots SC$  axes are averaged out, we consider only stretching of the virtual bonds along their axes,  $d\mathbf{C}$  and  $d\mathbf{X}$ , and rotations of the virtual bonds about axes perpendicular to them and passing through the center of mass. Rotations about the virtual-bond axes and vibrations along the axes perpendicular to the virtual-bond axes correspond to the hidden degrees of freedom. The average potential energy corresponding to these degrees of freedom is included in the UNRES energy function (eq 1), while the average kinetic energy is assumed to be constant (cf. eq 10). In other words, we treat the virtual bonds as thin rods. We further assume that the density ( $\rho$ ) of the mass along a rod is a function of its relative position along the rod, as given by eq 11, and that the center of mass is in the middle of a rod.

$$\rho = \rho(z/d) \quad (11)$$

where  $d$  is the length of a rod and  $z$  is the position of a point on the rod taking the center of mass as the origin. For the  $i$ th peptide group, the length of the rod is equal to the length of the corresponding virtual bond, while for a side chain except glycine the length is twice the length of the  $C^{\alpha}\cdots SC$  virtual bond; for glycine, the side chain is a point mass located on the  $C^{\alpha}$  atom. The kinetic energy  $K_x$  consists of the contributions arising from the translational motions as well as the internal motions of all interaction sites, the latter including rotation and vibration.

$$K_x = \sum_i [K_{p_i}^{(trans)} + K_{p_i}^{(int)} + K_{sc_i}^{(trans)} + K_{sc_i}^{(int)}] \quad (12)$$

In the above equation,  $K_{p_i}^{(trans)}$  is the translational kinetic energy of the  $i$ th peptide group,  $K_{sc_i}^{(trans)}$  is the translational kinetic energy of the  $i$ th side chain group,  $K_{p_i}^{(int)}$  is the internal kinetic energy of the  $i$ th peptide group (which is the kinetic energy of the distribution of the mass of the virtual bond given a fixed position of the center of the mass), and  $K_{sc_i}^{(int)}$  is the internal kinetic energy of the  $i$ th side-chain group, respectively.

On the basis of eqs 5 and 7, we can express  $K_{p_i}^{(trans)}$ ,  $K_{p_i}^{(int)}$ ,  $K_{sc_i}^{(trans)}$ , and  $K_{sc_i}^{(int)}$  by eqs 13 – 16.

$$K_{p_i}^{(trans)} = \frac{1}{2} m_p \|v_{p_i}\|^2 = \frac{1}{2} m_p \left\| \sum_{j=0}^{i-1} d\dot{C}_j + \frac{1}{2} d\dot{C}_i \right\|^2 \quad (13)$$

$$K_{p_i}^{(int)} = \frac{1}{2} I_p \|d\dot{C}_i\|^2 \quad (14)$$

$$K_{sc_i}^{(trans)} = \frac{1}{2} m_{sc_i} \|v_{sc_i}\|^2 = \begin{cases} \frac{1}{2} m_{sc_i} \left\| \sum_{j=0}^{i-1} d\dot{C}_j + d\dot{X}_i \right\|^2 & SC_i \neq Gly \\ \frac{1}{2} m_{sc_i} \left\| \sum_{j=0}^{i-1} d\dot{C}_j \right\|^2 & SC_i = Gly \end{cases} \quad (15)$$

$$K_{SC_i}^{(int)} = \begin{cases} \frac{1}{2} I_{SC_i} \|d\dot{X}_i\|^2 & SC_i \neq Gly \\ 0 & SC_i = Gly \end{cases} \quad (16)$$

with, according to eq 11

$$I_p = m_p \int_{-1/2}^{1/2} (z/d_p)^2 \rho(z/d_p) dz/d_p \quad (17)$$

$$I_{SC_i} = m_{SC_i} \int_{-1}^1 (z/d_{SC_i})^2 \rho(z/d_{SC_i}) dz/d_{SC_i} \quad (18)$$

In eqs 17 and 18,  $I_p$  and  $I_{SC_i}$  are the mass coefficients of the moments of inertia of the  $i$ th peptide group and the  $i$ th side chain, respectively, about axes perpendicular to the  $C^{\alpha} \cdots C^{\alpha}$  and  $C^{\alpha} \cdots SC$  bonds, respectively. For  $\rho = \text{constant}$  (a uniform mass distribution),  $I_{p_i} = (1/12)m_p$  and  $I_{SC_i} = (1/3)m_{SC_i}$  (the length of a side chain is twice its virtual-bond length and hence the factor of  $1/3$  instead of  $1/12$ ).

As can be seen from eqs 13–16, the kinetic energy depends only on velocities. The derivatives of the Lagrangian in velocities and time are expressed by eqs 19–24, respectively.

$$\frac{d}{dt} \nabla_{d\dot{C}_j} K_{P_i}^{(trans)} = \begin{cases} m_p \left( \sum_{k=0}^{i-1} d\ddot{C}_k + \frac{1}{2} d\ddot{C}_i \right) & i > j \\ \frac{1}{2} m_p \left( \sum_{k=0}^{i-1} d\ddot{C}_k + \frac{1}{2} d\ddot{C}_i \right) & i = j \\ 0 & i < j \end{cases} \quad (19)$$

and

$$\frac{d}{dt} \nabla_{d\dot{X}_i} K_{P_i}^{(trans)} = 0 \quad (20)$$

$$\frac{d}{dt} \nabla_{d\dot{C}_j} K_{SC_i}^{(trans)} = \begin{cases} m_{SC_i} \left( \sum_{k=0}^{i-1} d\ddot{C}_k + d\ddot{X}_i \right) & i \geq j \text{ and } SC_i \neq Gly \\ m_{SC_i} \sum_{k=0}^{i-1} d\ddot{C}_k & i \geq j \text{ and } SC_i = Gly \\ 0 & i < j \end{cases} \quad (21)$$

$$\frac{d}{dt} \nabla_{d\dot{X}_j} K_{SC_i}^{(trans)} = \begin{cases} m_{SC_i} \left( \sum_{k=1}^i d\ddot{C}_k + d\ddot{X}_i \right) & j = i \\ 0 & j \neq i \end{cases} \quad (22)$$

$$\frac{d}{dt} \nabla_{d\dot{C}_i} K_{P_i}^{(int)} = I_p \mathbf{1} \quad (23)$$

$$\frac{d}{dt} \nabla_{d\dot{X}_i} K_{SC_i}^{(int)} = I_{SC_i} \mathbf{1} \quad (24)$$

where  $\mathbf{1}$  is the  $3 \times 3$  identity matrix.

The derivatives of the Lagrangian in generalized coordinates  $\mathbf{q}$  constitute the gradient of the UNRES energy function, because the kinetic energy does not depend on coordinates (eqs 13–16). Therefore, the left side of the Lagrange equations of motion (eq 9) can be expressed by eq 25.



$$\frac{d}{dt}(\nabla_{\dot{\mathbf{q}}}L) - \nabla_{\mathbf{q}}L = G\ddot{\mathbf{q}} + \nabla_{\mathbf{q}}U \quad (25)$$

with

$$G = \mathbf{A}^T \mathbf{M} \mathbf{A} + \mathbf{H} \quad (26)$$

The term with the matrix  $\mathbf{A}$  corresponds to the translational part of the inertia matrix and that with the matrix  $\mathbf{H}$  to the internal-motion part. The matrix  $\mathbf{A}$  is the matrix of a linear transformation from the space of generalized coordinates and velocities ( $\mathbf{q}$  and  $\dot{\mathbf{q}}$ ) to the space of the Cartesian coordinates and velocities of the interacting sites (eqs 4–7). This matrix is defined by eq 27.

$$\mathbf{A} = \begin{pmatrix} 1 & \frac{1}{2}\mathbf{1} \\ 1 & \mathbf{1} \\ \frac{1}{2}\mathbf{1} & \\ \vdots & \vdots \\ \vdots & \ddots \\ 1 & \mathbf{1} \\ 1 & \dots \\ \frac{1}{2}\mathbf{1} & \\ 1 & \mathbf{0} \\ 0 & \mathbf{0} \\ \dots & \mathbf{0} \\ h_{SC_s} \mathbf{1} & \\ 1 & \mathbf{1} \\ 0 & \mathbf{0} \\ \dots & \mathbf{0} \\ 0 & h_{SC_{s+1}} \mathbf{1} \\ \vdots & \vdots \\ \vdots & \vdots \\ \ddots & \vdots \\ \vdots & \vdots \\ \ddots & \\ 1 & \mathbf{1} \\ 1 & \mathbf{1} \\ \dots & \mathbf{1} \\ 0 & \mathbf{0} \\ \dots & h_{SC_e} \mathbf{1} \end{pmatrix} \quad (27)$$

where  $h_{SC_i}$  is 0 if  $SC_i = \text{Gly}$  and 1 otherwise,  $\mathbf{1}$  is a  $3 \times 3$  identity matrix, and  $\mathbf{0}$  is a  $3 \times 3$  matrix composed of zeros. It should be noted that the columns of the matrix  $\mathbf{A}$  run through all variables (the elements of the vector  $\mathbf{q}$  of eq 3) and the rows run through the Cartesian coordinates of the interacting sites.

The matrix  $\mathbf{M}$  is a diagonal matrix containing the masses of the interacting sites and is defined by eq 28

$$M = \begin{pmatrix} m_p 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & m_p 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & m_{SC_s} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & m_{SC_{s+1}} 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & m_{SC_e} 1 \end{pmatrix} \quad (28)$$

where  $m_p$  is the mass of a peptide group and  $m_{SC_i}$  is the mass of the  $i$ th side chain.

The matrix  $\mathbf{H}$  contains the quantities  $I_p$  and  $I_{SC_i}$  of eqs 17 and 18, respectively. It is defined by eq 29.

$$H = \begin{pmatrix} I_p 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & I_p 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & I_{SC_{i1}} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & I_{SC_{i2}} 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & I_{SC_{im}} 1 \end{pmatrix} \quad (29)$$

The fact that the matrix  $\mathbf{G}$  defined by eq 26 is nondiagonal implies that a system of linear equations must be solved to compute accelerations from forces (see eq 32 in section 2.3). One may ask if, with an appropriate choice of coordinates, the problem could be reduced to a diagonal one, i.e., to classical Newton's equations. Using the Cartesian coordinates of the interacting sites (i.e., the SC and p centers) as variables seems to be a natural choice leading to diagonal equations of motion. These coordinates do not, however, define the geometry of an UNRES chain, because the virtual-bond vectors are also required. The virtual-bond vectors can be defined, if the coordinates of the  $C^\alpha$ 's are specified. Because the peptide groups are located exactly in the middle between the consecutive  $C^\alpha$ 's, a complete set of coordinates could consist of the Cartesian coordinates of the  $C^\alpha$ 's and the attached SCs. However, with such a choice, the system still is nondiagonal because the coordinates of the centers of the peptide groups are linear combinations of those of the  $C^\alpha$ 's.

### 2.3. Integrating the Equations of Motion

A variety of integrators have been designed for molecular dynamics algorithms, the most common being the Verlet algorithm,<sup>58</sup> the leapfrog algorithm,<sup>59</sup> the Beeman algorithm,<sup>60</sup> and the velocity Verlet algorithm.<sup>61</sup> We chose the velocity Verlet algorithm, because it yields velocities and positions at a specific time  $t$  rather than at other times.<sup>59</sup> The velocities are, in turn, needed to compute the kinetic energy (which is necessary to couple the system to the thermal bath) and to compute friction forces, as is done in the accompanying paper.<sup>54</sup> Another advantage of the velocity Verlet algorithm is that it is symplectic (allowing, in particular, for an approximate conservation of the total energy in long-time simulations<sup>62–64</sup>) and time-reversible.<sup>65,66</sup> The algorithm consists of two iterating steps. In the first step, the coordinates at time  $t + \delta t$  (where  $\delta t$  is the time step) are calculated from the positions, velocities, and accelerations computed at  $t$  by using eq 30.

$$q(t + \delta t) = q(t) + \dot{q}(t)\delta t + \frac{1}{2}\ddot{q}(t)\delta t^2 \quad (30)$$

In the second step, the velocities at  $t + \delta t$  are calculated from eq 31.

$$\dot{q}(t+\delta t)=\dot{q}(t)+\frac{1}{2}[\ddot{q}(t)+\ddot{q}(t+\delta t)]\delta t \quad (31)$$

The accelerations are computed from eq 32.

$$\ddot{q}(t)=-G^{-1}\nabla_q U[q(t)] \quad (32)$$

From eq 32, it appears that the choice of virtual-bond vectors as variables instead of curvilinear coordinates (such as, e.g., virtual-bond angles and virtual-bond dihedral angles) contributes to reducing the numerical cost of the algorithm. The matrix  $\mathbf{G}$  is a constant matrix, and therefore, its inverse can be computed once and then used to calculate accelerations from eq 32, the cost of this operation being on the order of  $N^2$  where  $N$  is the number of coordinates. Use of curvilinear coordinates would require inverting the matrix  $\mathbf{G}$  at each integration step, the cost of which scales as  $N^3$ .

It is desirable to increase  $\delta t$  as much as possible because that means advancing faster through the trajectory. For all-atom potentials,  $\delta t$  is chosen such that the total time at each step is at least an order of magnitude smaller than the length of time corresponding to the fastest motion in the system. Typically, this motion is the vibration of a hydrogen bond with a period less than 1 fs. Therefore,  $\delta t$  for an all-atom molecular dynamics algorithm is on the order of femtoseconds. The values of  $\delta t$  for some of the molecular dynamic packages are summarized in Table 1. It can be seen that these time steps are orders of magnitude smaller than the time scales of processes occurring in living cells, which are summarized in Table 2. Therefore, the application of all-atom molecular dynamics to study biochemical processes is usually limited to motions about the equilibrium structure of a macromolecule unless specially designed potentials guiding the system to the desired configuration are implemented. If the energy is expressed in kcal/mol, the mass in g/mol, and the distance in Å, then the natural unit of time is 48.9 fs; this unit will hereafter be referred to as the molecular time unit (mtu). In contrast to this, use of the UNRES representation of a polypeptide chain involves averaging over fast motions,<sup>56</sup> and consequently, the effective UNRES unit of time is larger than that resulting from dimensional analysis. We address this question in the accompanying paper.<sup>54</sup>

The effective time step can be increased by using the time-split algorithms<sup>65–69</sup> in which the forces acting in the system are split into the long-range forces ( $\mathbf{F}_L$ ), which are expensive to evaluate but usually vary slowly with conformation, and the short-range forces ( $\mathbf{F}_S$ ), which are not expensive to evaluate and vary quickly. The short-range forces are integrated with a smaller time step ( $\delta t_S$ ), and the long-range forces with a larger time step ( $\delta t_L$ );  $\delta t_S = \delta t_L/n$ , where  $n$  is an integer. The time-split algorithms are derived by splitting the Liouville propagator of the time evolution of the system under study.<sup>65–69</sup> In this study, in part of our calculations (section 3) we adapted the reversible reference system propagator algorithm (RESPA).<sup>65</sup> In the RESPA algorithm, each integration step consists of the following operations:

1. Calculate the velocities at half of the large time step ( $\delta t_L$ ) by using eq 33.

$$\dot{q}_1=\dot{q}\left(t+\frac{1}{2}\delta t_L\right)=\dot{q}(t)+\frac{1}{2}\ddot{q}_L(t)\delta t_L \quad (33)$$

with

$$\ddot{q}_L(t)=G^{-1}\mathbf{F}_L[q(t)]=-G^{-1}\nabla_q U_L[q(t)] \quad (34)$$

where  $U_L$  denotes the sum of the long-range energy components of the UNRES energy function.

2. After the velocities are set at  $\mathbf{q}'_1$  as computed in step 1, carry out  $n$  velocity Verlet steps (eqs 30 and 31) with accelerations computed by using only the short-range forces (eq 35). Let the resulting coordinates and velocities be denoted by  $\mathbf{q}_2$  and  $\mathbf{q}'_2$ , respectively,  $\mathbf{q}_2 = \mathbf{q}(t + \delta t_L)$  being the final coordinates at the end of the RESPA step.

$$\ddot{\mathbf{q}}_s(t) = \mathbf{G}^{-1} \mathbf{F}_s[\mathbf{q}(t)] = -\mathbf{G}^{-1} \nabla_{\mathbf{q}} U_s[\mathbf{q}(t)] \quad (35)$$

where  $U_s$  denotes the sum of the short-range energy components of the UNRES energy function.

3. Through the use of the coordinates and velocities from step 2, compute the velocities at the end of the RESPA step from eq 36.

$$\dot{\mathbf{q}}(t + \delta t_L) = \dot{\mathbf{q}}_2 + \frac{1}{2} \ddot{\mathbf{q}}_L(t + \delta t_L) \delta t_L \quad (36)$$

The accelerations  $\ddot{\mathbf{q}}_L(t + \delta t)$  are computed from eq 34, where  $\mathbf{q}(t + \delta t_L)$  is substituted for  $\mathbf{q}(t)$ .

#### 2.4. Generating Initial Velocities

The initial velocities were generated from a Gaussian distribution corresponding to a given temperature. To derive appropriate formulas, we first note that the kinetic energy can be expressed by eq 37.

$$K = \frac{1}{2} \dot{\mathbf{q}}^T \mathbf{G} \dot{\mathbf{q}} = \frac{1}{2} \sum_{i=1}^n \mu_i \xi_i^2 \quad (37)$$

with

$$\xi = \mathbf{V} \mathbf{q} \quad (38)$$

where the matrix  $\mathbf{G}$  is defined by eq 26,  $\mu_i$ ,  $i = 1, 2, \dots, n$  are the eigenvalues of  $\mathbf{G}$ , and  $\mathbf{V}$  is the matrix of the eigenvectors of  $\mathbf{G}$ . The velocities expressed in normal coordinates ( $\xi_1, \xi_2, \dots, \xi_n$ ) can now be sampled from normal distributions, and the initial velocities  $\mathbf{q}'$  of the virtual-bond vectors can be computed by inverting eq 38, as expressed by eqs 39 and 40, respectively.

$$\xi_i = \sqrt{RT/\mu_i} N(0,1) \quad (39)$$

$$\dot{\mathbf{q}}^{\text{init}} = \mathbf{V}^T \xi \quad (40)$$

where  $N(0,1)$  is the normal distribution with zero mean and unit variance.

### 3. Numerical Tests of the Total Energy Conservation

One of the most desired features of any molecular dynamics algorithm is its symplecticity. A symplectic algorithm ensures approximate conservation of the total energy (without drift) in a long simulation of a microcanonical ensemble, assuming that a sufficiently small time step and a sufficiently smooth potential-energy function are used.<sup>63,64</sup> This is a test that allows one to determine how stable an integration scheme is and the largest integration time step that can be

used. It should be noted, however, that microcanonical simulations do not correspond to realistic models and most simulations are carried out in the canonical mode where energy exchange between the system under study and the environment occurs.

To test the integration scheme, we carried out microcanonical ensemble simulations (i.e., without coupling to a thermal bath) for 100 mtu (4.89 ps) on the unblocked Ala<sub>10</sub> polypeptide with various values of  $\delta t$ . The global minimum for the UNRES potential energy with the F2 set of parameters derived in ref<sup>48</sup> corresponds to a full right-handed  $\alpha$ -helix, as found by the global-optimization CSA method.<sup>50,51</sup> The simulations were started from the right-handed  $\alpha$ -helix, with all virtual-bond angles  $\theta$  set at 90° and all virtual-bond dihedral angles  $\gamma$  set at 45°. The components of the initial velocity vector were sampled from the Gaussian distribution of velocities with zero mean and variance corresponding to a “temperature” of 300 K as described in section 2.4. (We must bear in mind that, since this is a microcanonical simulation, the total energy and not the temperature is supposed to be conserved throughout the run.) The graph of the total energy versus time for different values of  $\delta t$  is shown in Figure 2A. The drift of the total energy from the initial value ( $|\delta E_{\text{Tot}}|$ ) for different values of  $\delta t$  is shown in Table 3.

It can be seen from Figure 2A and Table 3 that the largest allowable time step for UNRES is 0.02 mtu, which corresponds to the formal time step of about 1 fs (cf section. 2.3). For larger values of  $\delta t$ , the total energy drifts significantly from the initial value in a stepwise manner. For reasonably small values of  $\delta t$ , the drift of the total energy is negligible compared to the variation of the potential and the kinetic energy, as can be seen from Figure 2B.

The rapid changes of the total energy correlate with sharp spikes in the magnitude of the maximum change of the acceleration at a given time step, as can be seen in Figure 2C for  $\delta t = 0.03$  mtu. Clearly, when switching to the canonical mode, in which the total energy is not conserved and the system is subjected to stochastic forces (considered either explicitly in Langevin dynamics or implicitly when applying the Berendsen or Nose–Hoover thermostat) which lead to quite significant changes of the total energy, the time step can be larger because the stochastic forces overwhelm the error in the total energy inherent in the integration algorithm.

An analysis of the components of the forces revealed that the most rapid changes occur in the gradient of the side-chain-rotamer energy,  $U_{\text{rot}}$  (cf. eq 1). This contribution to the energy consists of logarithms of the sums of the Gaussian terms in the angles  $\alpha$  and  $\beta$  that define the rotameric state of a side chain (Figure 1A). These rapid jumps of the gradient of  $U_{\text{rot}}$  arise in the neighborhood of  $\alpha_{\text{SC}} = 180^\circ$ . The explanation for this behavior is simple: Here, we consider the gradient in the virtual-bond vectors. Suppose that  $\alpha_{\text{SC}_i}$  approaches 180°. Then, a very small change of the corresponding virtual-bond vector  $d\mathbf{X}_i$  will result in a big change of the angle  $\beta_{\text{SC}_i}$  and, consequently, in the pertinent component of  $U_{\text{rot}}$ , which results in high gradients in  $d\mathbf{X}_i$ . It should be noted that, when  $\alpha_i = 180^\circ$ , the angle  $\beta_{\text{SC}_i}$  and, thereby, the gradients in  $d\mathbf{X}_i$  are indefinite. In our previous implementation of the UNRES force field (in which we performed only local energy minimizations),<sup>39,43,70</sup> this problem did not arise because we used the angular coordinates ( $\theta$ ,  $\gamma$ ,  $\alpha_{\text{SC}}$ , and  $\beta_{\text{SC}}$ ) as variables. Such a choice of variables is, however, not appropriate for molecular dynamics because it involves a greater cost for computing accelerations from forces (eq 32). Besides, the present  $U_{\text{rot}}$  is one of the few terms remaining in UNRES that were derived as statistical potentials from the PDB.<sup>36,37</sup> We are now working on replacing these statistical terms (including  $U_{\text{rot}}$ ) with physics-based terms that will not depend on angles explicitly and will, therefore, not result in numerical instabilities in the forces. For this reason, instead of tuning the present integration scheme to handle the gradients of  $U_{\text{rot}}$  perfectly, we worked out a practical temporary solution, which is described below.

To prove that the gradient of  $U_{\text{rot}}$  is the cause of the spikes in the acceleration and, consequently, in the energy drift, we carried out a series of 100 000 mtu (4.89 ns) of microcanonical simulations on the Gly<sub>10</sub> polypeptide where the  $U_{\text{rot}}$  energy component is absent. With the force field used, Gly<sub>10</sub> has a doubly degenerate global minimum, this being either a right- or a left-handed  $\alpha$ -helix. The plots of the total energy versus time for the time steps of 0.05, 0.1, 0.15, and 0.2 mtu, respectively (2.445, 4.89, 7.335, and 9.78 fs, respectively), are presented in Figures 3A–D, while a graph of the potential, kinetic, and total energy for the time step of 0.15 mtu is shown in Figure 3E. It can be seen that oscillations of the total energy take place only up to the time step of  $\delta t = 0.15$  mtu (7.335 fs), and the energy drift starts to appear (after more than 40 000 mtu or about 2 ns) only with the time step  $\delta t = 0.20$  mtu (9.78 fs). For larger time steps, the trajectory is unstable with the total energy rapidly diverging.

To prove that the spikes in accelerations are not caused by other components involving the side-chain virtual-bond vectors  $\mathbf{dX}$ , we carried out a series of 100 000 mtu simulations on the Ala<sub>10</sub> peptide with the UNRES energy function in which  $U_{\text{rot}}$  was replaced with a simple harmonic potential defined by eq 41.

$$U_{\text{rot}}^{\text{simple}}(\alpha_{\text{sc}_i}, \beta_{\text{sc}_i}) = 100[(\alpha_{\text{sc}_i} - 2\pi/3)^2 + (\beta_{\text{sc}_i} + 2\pi/3)^2] \quad (41)$$

where the angles are expressed in radians. The potential defined by eq 41 forces the angles to be centered about ( $\alpha = 120^\circ$ ,  $\beta = -120^\circ$ ), close to the position of the center of the alanine side chain determined from the PDB.<sup>37</sup> The large force constant (100 kcal/(mol rad<sup>2</sup>)) keeps the angles  $\alpha$  away from the instability region close to  $\alpha = 180^\circ$ . The graphs of the total energy for the time steps of 0.05, 0.1, and 0.15 mtu are shown in Figures 4A–C, respectively (the total energy diverges for higher  $\delta t$  values), while the plots of the potential, kinetic, and the total energy are shown in Figure 4D. It can be seen that the total energy only oscillates now, as in the case of the Gly<sub>10</sub> system. It can, therefore, be concluded that the energy drift observed when carrying out molecular dynamics simulations with the full UNRES is caused by numerical instability of the gradient of  $U_{\text{rot}}$ .

Because the total energy is reasonably conserved for most of an UNRES molecular dynamics trajectory run in the microcanonical mode, even for larger time steps for most of a trajectory except for the spikes in acceleration, we conclude that reducing the time step in the regions of the spikes, while retaining a larger time step for most of a run, could increase the time scale of the UNRES dynamics. One algorithm for reducing the time step in these dangerous regions was designed by Gibson and Scheraga.<sup>71</sup> They derived rigorous formulas based on the Lyapunov theory of stability for the length of the time step, depending on the shape of the energy function at a given point. However, these formulas require the computation of the third derivatives of the potential-energy function (this is a third-order tensor), which is very expensive in the case of UNRES because of the presence of the multibody terms. We, therefore, designed a simpler algorithm, which can be summarized as follows:

1. Store the coordinates, velocities, and accelerations from the previous time step.
2. Compute the new coordinates from eq 30 and the current value of the time step.
3. Through the use of the new coordinates, compute the forces and accelerations.
4. Determine the maximum change of acceleration  $\delta a_{\text{max}} = \max |\ddot{\mathbf{q}}_i - \ddot{\mathbf{q}}_i^{(\text{old})}|$ .
5. If  $\delta a_{\text{max}}$  is less than the cutoff value (we use 1 or 2  $\text{\AA}/\text{mtu}^2$ ), then compute the new velocities from eq 31 and go to the next time step. If not, then reduce the time step according to eq 42 and go back to step 2.

$$\delta t_{\text{reduced}} \leftarrow \delta t \times 2^{-\max\{\text{int}[\log_2(\delta a_{\text{max}}/\delta a_{\text{cut}})], 1\}} \quad (42)$$

where  $\text{int}(x)$  denotes the integer closest to  $x$  and smaller or equal to it. Scaling  $\delta t$  by powers of 2 is not important in this work but becomes important when quantities dependent on the value of the time step are needed (such as, e.g., those used in the integration of the Langevin equations of motion; see the accompanying paper<sup>54</sup>).

Figure 5A shows the variation of the total energy for five representative trajectories of the unblocked Ala<sub>10</sub> polypeptide obtained with the constant time step  $\delta t = 0.05$  mtu, while Figure 5B shows the corresponding trajectories (each drawn with the same line style as its counterpart in Figure 5A) obtained with the variable-time-step algorithm. We noted earlier (Figure 2 and Table 3) that this time step is too large for the constant-time-step velocity Verlet integrator to yield stable trajectories. Each pair of corresponding trajectories (one of Figure 5A and one of Figure 5B) was started from the same set of initial velocities (generated as described in section 2.4). It can be seen that the use of the variable time step reduces the energy drift greatly, and the total energy is reasonably conserved.

We also carried out test simulations with the variable time step for more values of  $\delta t$ . The results are summarized in Table 4. It can be seen that, even with  $\delta t = 0.1$  mtu (4.89 fs), the drift of the total energy reaches 8.8 kcal/mol only after 10 ps. This is certainly a large value for a microcanonical simulation; however, for canonical simulations, the fluctuations of the total energy for the Ala<sub>10</sub> polypeptide are on the order of 50 kcal/mol,<sup>54</sup> and therefore, a time step that causes a larger energy drift can be used. It can also be seen by comparing Table 4 with Table 3 that the shortening of the total length of the trajectory due to occasional reduction of the time step (eq 42) appears insignificant.

While the variable-time-step algorithm described above seems to be a practical temporary solution, it should be noted that it is not time-reversible and not symplectic. We, therefore, tried to implement the symplectic multiple-time-step RESPA algorithm,<sup>65</sup> as described in section 2.3. We found that, for reasonable stability, the short time step  $\delta t_S$  must be on the order of 0.005 mtu (0.2445 fs). The graphs of the total energy for  $\delta t_L/\delta t_S = 0.02/0.004$ ,  $0.05/0.005$ , and  $0.1/0.005$  mtu are shown in Figure 6; some drift in the total energy is still observed. However, first, it occurs on a much longer time scale (on the order of nanoseconds) than when applying the variable-time-step algorithm and, second, has the form of long-period waves and not energy jumps. Therefore, numerically, the use of RESPA is a much better solution to handle the instability of the gradient of  $U_{\text{rot}}$  than using the variable-time-step algorithm. However, RESPA involves a major increase of computational cost. For the time step of  $\delta t_L = 0.05$  mtu when 10 steps with  $\delta t_S$  are required, the cost increases by 90%, and for  $\delta t_L = 0.10$  mtu (20 steps with  $\delta t_S$ ), the cost increases by 180%. The major contribution to this increase arises from the fact that computation of accelerations from forces involves a matrix multiplication (eq 32). In comparison to this, the use of the variable-time-step algorithm involves only a 5% increase in computational cost. The increase of computational cost when using RESPA is even more significant for larger systems, because the cost of computation of accelerations from forces scales with the square of the size of the system (cf. section 2.3). For example, for the N-terminal portion of the B-domain of staphylococcal protein A (a 46-residue protein) studied in the accompanying paper,<sup>54</sup> 10 and 20 small time steps increase the cost by 140% and 270%, respectively. Because the variable-time-step algorithm proved to provide sufficient energy stability to maintain constant average temperature when carrying out canonical simulations,<sup>54</sup> we decided to use it as a temporary solution of the problem of instability of the gradient of  $U_{\text{rot}}$ .

Figure 7 shows snapshots from a trajectory of the Ala<sub>10</sub> system started from an extended structure (i.e, all of the virtual-bond dihedral angles  $\gamma$  have been set at 180° obtained with  $\delta t = 0.02$  mtu (about 1 fs)). It can be seen that, even though the simulation was carried out in the microcanonical mode, which means that there was no possibility to transfer the high potential

energy of the starting structure to a thermal bath, helix nucleation, propagation, and unwinding can be observed, all these events occurring within about 10 ps. Because helix formation requires at least  $0.6 \mu\text{s}$ ,<sup>5,72</sup> the value of  $\delta t$  taken in these calculations must actually correspond to a value much larger than 1 fs, resulting from unit conversion, as remarked in the Methods section.

## 4. Conclusions

In this paper, we derived the Lagrange equations of motion for the UNRES force field, implemented the velocity Verlet algorithm to integrate them numerically, and analyzed the stability of the integration scheme by carrying out microcanonical molecular dynamics runs. The total energy of the system appears to be sufficiently conserved (i.e., the algorithm is stable) for up to 10 ps (Table 3) with a time step of 0.02 molecular time units, which formally corresponds to 0.978 fs. However, energy drift and not just oscillations is observed even for the smallest time step applied. The energy drift occurs because of rare but dramatic changes in the magnitude of the forces (and, thereby, accelerations); such changes usually result in stepwise changes of the total energy. An analysis of UNRES energy components showed that the present statistics-based side-chain-rotamer potentials ( $U_{\text{rot}}$  in eq 1) are responsible for the drift in the total energy. These  $U_{\text{rot}}$  potentials depend explicitly on the angles  $\alpha_{\text{SC}}$  and  $\beta_{\text{SC}}$  defining the orientation of a side-chain center with respect to the peptide backbone (Figure 1A), and therefore, their gradients in the virtual-bond vectors are numerically unstable in the vicinity of the side-chain angle  $\alpha = 180^\circ$ . When  $U_{\text{rot}}$  is not present or replaced with a numerically stable potential, the total energy does not drift but undergoes oscillations only until the time step of 0.15 mtu (7.335 fs) and for the total length of a trajectory equal to 100 000 mtu (4.89 ns). This time-step value is comparable with that achieved by Rudnicki et al.<sup>32</sup> with their mesoscopic model of nucleic acids. These results demonstrate that the functional form of the present  $U_{\text{rot}}$  must be replaced with one implying stable gradients in virtual-bond vectors. Work in this direction is now being carried out in our laboratory.

To be able to carry out molecular dynamics simulations with the present UNRES, we introduced a variable-time-step algorithm in which the time step is scaled down if a rapid change of acceleration is detected. With this modification, the energy drift is greatly reduced and the time step can be increased to 0.05 mtu (2.445 fs). In the accompanying paper,<sup>54</sup> we demonstrate that, even when the time step is set at 0.1 mtu (4.89 fs), the average temperature is kept constant while running simulations in the canonical mode. Using the variable-time-step algorithm involves only about a 5% increase of computational cost. Thus, the solution worked out in this paper appears to be sufficient for the time being. It should be noted, however, that use of the variable-time-step algorithm breaks the symplecticity and time reversibility of the integration scheme, resulting in possible worsening of the total energy conservation in a longer microcanonical simulation.

We also tried to implement the time-reversible multiple-time-step RESPA algorithm<sup>65</sup> to handle the instability of the gradient of  $U_{\text{rot}}$ . We found that, with the RESPA algorithm, the energy is reasonably stable even for long-time trajectories up to a time step of 0.15 mtu (7.335 fs) and the energy drift is manifested as long-period oscillations (the energy drift cannot be eliminated completely because of the inherent instability of the gradient of  $U_{\text{rot}}$ ). Therefore, RESPA is clearly superior to the variable-time-step algorithm as far as stability is concerned. However, it involves a dramatic increase of computational cost (90% for the Ala<sub>10</sub> sequence and 140% for the 46-residue fragment of the staphylococcal protein A); this cost is associated with the need to perform a matrix multiplication when computing accelerations from forces (eq 32) and will therefore increase as the square of the size of the system. We, therefore, decided to keep the variable-time-step method as a temporary solution for the instability of the gradient of  $U_{\text{rot}}$ .



Even when the problem of instability of the  $U_{\text{rot}}$  gradient is eliminated by revising this energy component, the use of the multiple-time-step algorithm can lead to an increase in the integration time step because rapid oscillations in the short-range fast-varying forces are eliminated (such as, e.g., the gradients of the virtual-bond-stretching energy  $U_{\text{vib}}$ ). (It should be noted that these oscillations do not result from the instability of the respective gradient components, as opposed to the case of  $U_{\text{rot}}$ , but from the fact that the corresponding forces vary faster than the long-range forces.) Preliminary studies suggest that extension of the time step to about 20 fs appears feasible; this would expand the time scale of simulations. However, the problem of the increased cost of computing accelerations from forces must be solved; otherwise the advantage of increasing the time step is largely reduced. The cost of RESPA also depends on the selection of the slow and the fast part of the potential and can be reduced further by implementing a variable time step. These issues are now being addressed in our laboratory.

Even the present variable-time-step algorithm appears sufficient to carry out ab initio simulations of protein folding. Even in the microcanonical test simulation of Ala<sub>10</sub>, we were able to observe formation and unwinding of the  $\alpha$ -helix within picoseconds on the UNRES time scale, while the experimental folding times of model short  $\alpha$ -helical polypeptides are on the order of 0.6  $\mu$ s. Therefore, the UNRES time scale seems to be larger by orders of magnitude than it seems from the analysis of units. We address this issue in more detail in the accompanying paper.<sup>54</sup>

## Acknowledgements

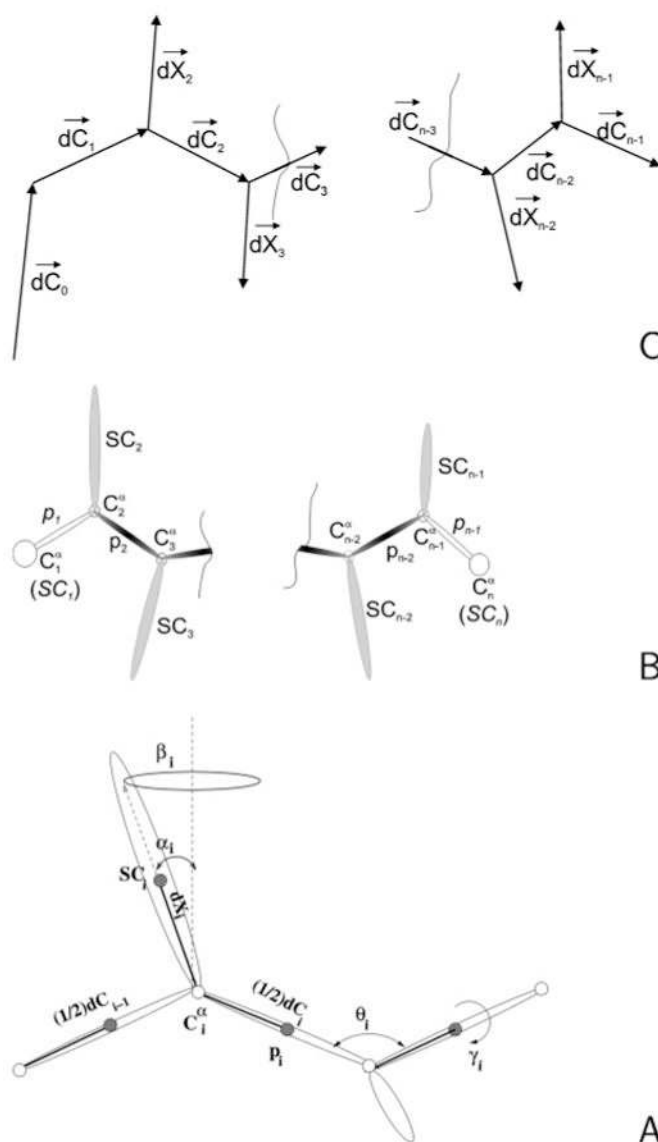
We are indebted to Professor Bogdan Lesyng, University of Warsaw, for valuable suggestions and comments on the manuscript. This work was supported by grants from the National Institutes of Health (NIH; GM-14312), the National Science Foundation (MCB00-03722), the NIH Fogarty International Center (TW1064), and the Polish Ministry of Scientific Research and Information Technology (3 T09A 032 26 and 6 T11 2003 C/06098). Studies of P.G. and F.R. were partially supported by the European Centre of Excellence MAMBA (project QLRI-CT-2002-90383). This research was conducted by using the resources of (a) our 392-processor Beowulf cluster at the Baker Laboratory of Chemistry and Chemical Biology, Cornell University, (b) the National Science Foundation Terascale Computing System at the Pittsburgh Supercomputer Center, (c) our 45-processor Beowulf cluster at the Faculty of Chemistry, University of Gdańsk, (d) the Informatics Center of the Metropolitan Academic Network in Gdańsk, and (e) the Interdisciplinary Centre for Mathematical and Computational Modelling at Warsaw University.

## References and Notes

1. van Gunsteren, WF. Molecular dynamics and stochastic dynamics: A primer. In: van Gunsteren, WF.; Weiner, PK.; Wilkinson, AJ., editors. *Computer Simulation of Biomolecular Systems*. ESCOM; Leiden, The Netherlands: 1993. p. 3-36.
2. Fersht AR, Daggett V. *Cell* 2002;108:573. [PubMed: 11909527]
3. Day R, Daggett V. *Adv Protein Chem* 2003;66:373. [PubMed: 14631823]
4. Shea JE, Brooks CL III. *Annu Rev Phys Chem* 2001;52:499. [PubMed: 11326073]
5. Kubelka J, Hofrichter J, Eaton WA. *Curr Opin Struct Biol* 2004;14:76. [PubMed: 15102453]
6. Elber R, Ghosh A, Cárdenas A. *Acc Chem Res* 2002;35:396. [PubMed: 12069624]
7. Ghosh A, Elber R, Scheraga HA. *Proc Natl Acad Sci U S A* 2002;99:10394. [PubMed: 12140363]
8. Duan Y, Kollman PA. *Science* 1998;282:740. [PubMed: 9784131]
9. Pande VS, Baker I, Chapman J, Elmer SP, Khaliq S, Larson SM, Rhee YM, Shirts MR, Snow CD, Sorin EJ, Zagrovic B. *Biopolymers* 2003;68:91. [PubMed: 12579582]
10. Jang S, Kim E, Shin S, Pak Y. *J Am Chem Soc* 2003;125:14841. [PubMed: 14640661]
11. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE III, DeBolt S, Ferguson D, Seibel G, Kollman P. *Comput Phys Commun* 1995;91:1.
12. Case, DA.; Pearlman, DA.; Caldwell, JW.; Cheatham, TE., III; Wang, J.; Ross, WS.; Simmerling, CL.; Darden, TA.; Merz, KM.; Stanton, RV.; Cheng, AL.; Vincent, JJ.; Crowley, M.; Tsui, V.; Gohlke, H.; Radmer, RJ.; Duan, Y.; Pitera, J.; Massova, I.; Siebel, GL.; Singh, UC.; Weiner, PK.; Kollman, PA. *AMBER 7*. University of California; San Francisco, CA: 2002.

13. Cramer CJ, Truhlar DG. *Chem Rev* 1999;99:2161. [PubMed: 11849023]
14. Bashford D, Case DA. *Annu Rev Phys Chem* 2000;51:129. [PubMed: 11031278]
15. Veitshans T, Klimov D, Thirumalai D. *Folding Des* 1996;2:1.
16. Klimov DK, Thirumalai D. *Phys Rev Lett* 1997;79:317.
17. He S, Scheraga HA. *J Chem Phys* 1998;108:271.
18. He S, Scheraga HA. *J Chem Phys* 1998;108:287.
19. Hoang TX, Cieplak M. *J Chem Phys* 2000;112:6851.
20. Cieplak M, Hoang TX, Robbins MO. *Proteins: Struct Funct Genet* 2002;49:104. [PubMed: 12211020]
21. Sorenson JM, Head-Gordon T. *Proteins: Struct Funct Genet* 2002;46:368. [PubMed: 11835512]
22. Koretke KK, Luthey-Schulten Z, Wolynes PG. *Proc Natl Acad Sci U S A* 1998;95:2932. [PubMed: 9501193]
23. Hardin C, Eastwood MP, Luthey-Schulten Z, Wolynes PG. *Proc Natl Acad Sci U S A* 2000;97:14235. [PubMed: 11114172]
24. Takada S. *Proteins: Struct Funct Genet* 2001;42:85. [PubMed: 11093263]
25. Hardin C, Eastwood MP, Prentiss M, Luthey-Schulten Z, Wolynes PG. *J Comput Chem* 2002;23:138. [PubMed: 11913379]
26. Hardin C, Eastwood MP, Prentiss MC, Luthey-Schulten Z, Wolynes PG. *Proc Natl Acad Sci U S A* 2003;100:1679. [PubMed: 12554830]
27. Fujitsuka Y, Takada S, Luthey-Schulten ZA, Wolynes PG. *Proteins: Struct Funct Genet* 2004;54:88. [PubMed: 14705026]
28. Kolinski A, Godzik A, Skolnick J. *J Chem Phys* 1993;98:7420.
29. Kolinski A, Skolnick J. *Proteins: Struct Funct Genet* 1994;18:338. [PubMed: 8208726]
30. Kolinski A, Skolnick J. *Proteins: Struct Funct Genet* 1994;18:353. [PubMed: 8208727]
31. Kolinski A, Rotkiewicz P, Ilkowski B, Skolnick J. *Prog Theor Phys Suppl* 2000;138:292.
32. Rudnicki WR, Bakalarski G, Lesyng B. *J Biomol Struct Dyn* 2000;17:1097. [PubMed: 10949175]
33. Kilmister, CW. *Lagrangian Dynamics: An Introduction for Students*. Logos Press Limited; London, U K: 1967.
34. Maciejczyk M, Rudnicki WR, Lesyng B. *J Biomol Struct Dyn* 2000;17:1109. [PubMed: 10949176]
35. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. *Protein Sci* 1993;2:1715. [PubMed: 8251944]
36. Liwo A, Oldziej S, Pincus MR, Wawak RJ, Rackovsky S, Scheraga HA. *J Comput Chem* 1997;18:849.
37. Liwo A, Pincus MR, Wawak RJ, Rackovsky S, Oldziej S, Scheraga HA. *J Comput Chem* 1997;18:874.
38. Liwo A, Kaźmierkiewicz R, Czaplewski C, Groth M, Oldziej S, Wawak RJ, Rackovsky S, Pincus MR, Scheraga HA. *J Comput Chem* 1998;19:259.
39. Liwo A, Lee J, Ripoll DR, Pillardy J, Scheraga HA. *Proc Natl Acad Sci U S A* 1999;96:5482. [PubMed: 10318909]
40. Liwo A, Czaplewski C, Pillardy J, Scheraga HA. *J Chem Phys* 2001;115:2323.
41. Lee J, Ripoll DR, Czaplewski C, Pillardy J, Wedemeyer WJ, Scheraga HA. *J Phys Chem B* 2001;105:7291.
42. Pillardy J, Czaplewski C, Liwo A, Wedemeyer WJ, Lee J, Ripoll DR, Arłukowicz P, Oldziej S, Arnautova YA, Scheraga HA. *J Phys Chem B* 2001;105:7299.
43. Pillardy J, Czaplewski C, Liwo A, Lee J, Ripoll DR, Kaźmierkiewicz R, Oldziej S, Wedemeyer WJ, Gibson KD, Arnautova YA, Saunders J, Ye YJ, Scheraga HA. *Proc Natl Acad Sci U S A* 2001;98:2329. [PubMed: 11226239]
44. Liwo A, Arłukowicz P, Czaplewski C, Oldziej S, Pillardy J, Scheraga HA. *Proc Natl Acad Sci U S A* 2002;99:1937. [PubMed: 11854494]
45. Oldziej S, Kozłowska U, Liwo A, Scheraga HA. *J Phys Chem A* 2003;107:8035.
46. Liwo A, Oldziej S, Czaplewski C, Kozłowska U, Scheraga HA. *J Phys Chem B* 2004;108:9421.
47. Liwo A, Arłukowicz P, Oldziej S, Czaplewski C, Makowski M, Scheraga HA. *J Phys Chem B* 2004;108:16918.
48. Oldziej S, Liwo A, Czaplewski C, Pillardy J, Scheraga HA. *J Phys Chem B* 2004;108:16934.

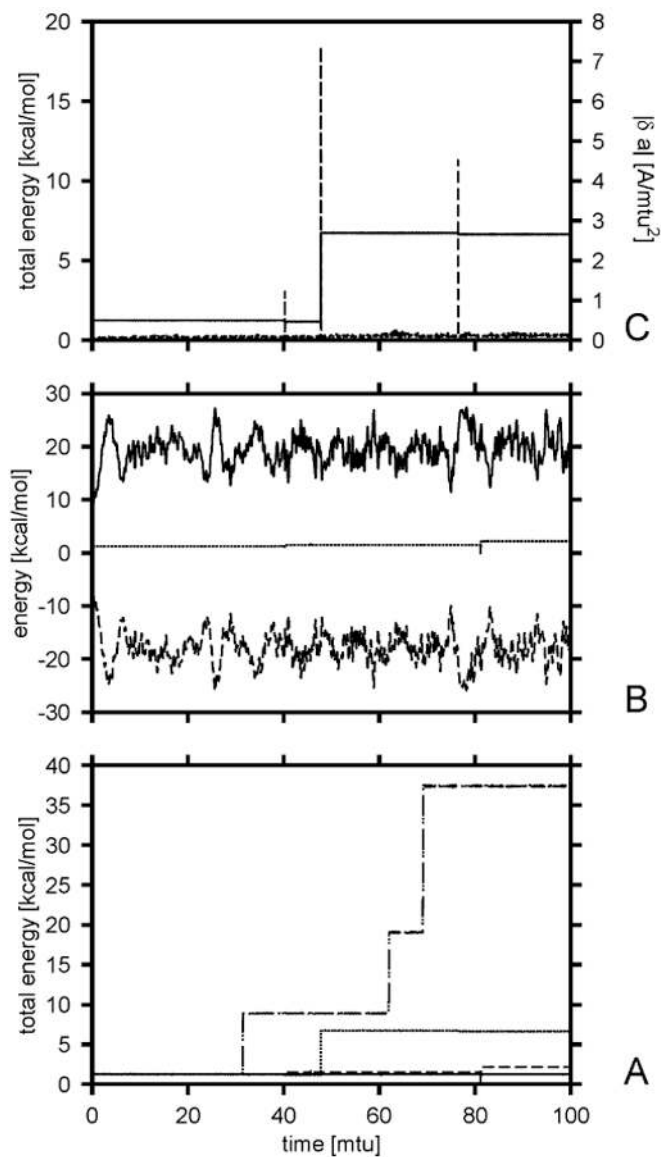
49. Ȯldziej S, Łggiewka J, Liwo A, Czaplewski C, Chinchio M, Nancias M, Scheraga HA. *J Phys Chem B* 2004;108:16950.
50. Lee J, Scheraga HA, Rackovsky S. *J Comput Chem* 1997;18:1222.
51. Lee J, Liwo A, Scheraga HA. *Proc Natl Acad Sci U S A* 1999;96:2025. [PubMed: 10051588]
52. Lee J, Liwo A, Ripoll DR, Pillardy J, Scheraga HA. *Proteins Struct Funct Genet* 1999;204. [PubMed: 10526370]
53. Ȯldziej S, Czaplewski C, Liwo A, Chinchio M, Nancias M, Vila JA, Khalili M, Arnautova YA, Jagielska A, Makowski M, Schafroth HD, Kaźmierkiewicz R, Ripoll DR, Pillardy J, Saunders JA, Kang YK, Gibson KD, Scheraga HA. *Proc Natl Acad Sci U S A* 2005;102:7547. [PubMed: 15894609]
54. Khalili M, Liwo A, Jagielska A, Scheraga HA. *J Phys Chem B* 2005;109:13798. [PubMed: 16852728]
55. Liwo A, Khalili M, Scheraga HA. *Proc Natl Acad Sci U S A* 2005;102:2362. [PubMed: 15677316]
56. van Kampen CW. *Phys Rep* 1985;124:69.
57. Sprik M, Ciccotti G. *J Chem Phys* 1998;109:7737.
58. Verlet L. *Phys ReV* 1967;159:98.
59. Hockney, RW.; Eastwood, JW. *Computer Simulations Using Particles*. McGraw-Hill; New York: 1981.
60. Beeman D. *J Comput Phys* 1976;20:130.
61. Swope WC, Andersen HC, Berens PH, Wilson KR. *J Chem Phys* 1982;76:637.
62. Ruth R. *IEEE Trans Nucl Sci* 1983;30:2669.
63. Calvo, MP.; Sanz-Serna, JM. *Numerical Hamiltonian Problems*. Chapman & Hall; London, U K: 1994.
64. Benettin G, Giorgilli A. *J Stat Phys* 1994;74:1117.
65. Tuckerman M, Berne BJ, Martyna GJ. *J Chem Phys* 1992;97:1990.
66. Ciccotti G, Kalibaeva G. *Philos Trans R Soc London* 2004;362:1583.A
67. Martyna GJ, Tuckerman ME, Tobias DJ, Klein ML. *Mol Phys* 1996;87:1117.
68. Stocker U, Juchli D, van Gunsteren WF. *Mol Simul* 2003;29:123.
69. Litniewski M. *Mol Simul* 2003;29:223.
70. Czaplewski C, Liwo A, Pillardy J, Ȯldziej S, Scheraga HA. *Polymer* 2004;45:677.
71. Gibson KD, Scheraga HA. *J Comput Chem* 1990;11:468.
72. Thompson PA, Muȃoz V, Jas GS, Henri ER, Eaton WA, Hofrichter J. *J Phys Chem B* 2000;104:378.



**Figure 1.**

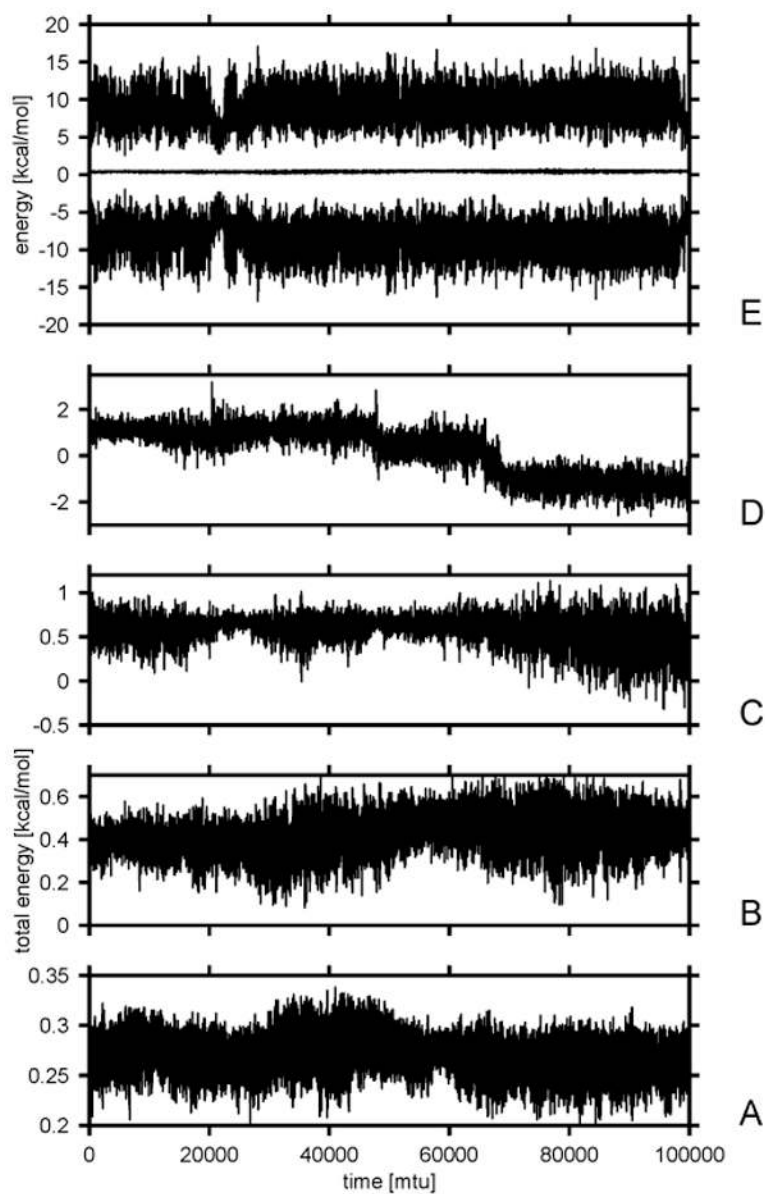
(A) UNRES model of the polypeptide chain. Dark circles represent united peptide groups ( $p$ ), and open circles represent the  $C^\alpha$  atoms, which serve as geometric points. Ellipsoids represent side chains, with their centers of mass at the SCs. The  $p$ 's are located halfway between two consecutive  $C^\alpha$  vectors or  $dC$ s. The SCs are located at the ends of the  $C^\alpha \cdots SC$  vectors or the  $dX$ s. The variables to change the conformation of the polypeptide chain are the virtual-bond angles  $\theta$ , the virtual-bond dihedral angles  $\gamma$ , and the angles  $\alpha_{SC}$  and  $\beta_{SC}$  that define the location of a side chain with respect to the backbone. (B) Ends of a united-residue chain. The peptide groups  $p_1$  and  $p_{n-1}$  are empty ellipsoids and are marked by symbols in italics, because they can be either real peptide groups that interact with the other peptide groups and side chains if the first (last) residue is glycine or a terminal blocking group that contains a backbone amide bond or empty groups. If the first (last) residue is not glycine and the chain is unblocked, then  $p_1$  and  $p_{n-1}$  are dummy objects introduced only to define the first (last) virtual-bond angle ( $\theta$ ) and virtual-bond dihedral angle ( $\gamma$ ), which are, in turn, necessary to define the geometry of the first (last) side chain. Likewise, the first (last) SC center (marked with empty circles and symbols in italics) is a methylene or methyl group (both identified with the glycine SC center)

if the first (last) residue is blocked or if it is a blocking group with an amide bond or a dummy center, if the chain is unblocked. The other peptide groups and side chains are black and filled with a dashed pattern, respectively, and marked with regular-text symbols. (C) Illustration of the definition of the virtual-bond vectors  $\mathbf{dC}$  (for  $C^{\alpha}\cdots C^{\alpha}$  virtual bonds) and  $\mathbf{dX}$  (for  $C^{\alpha}\cdots SC$  virtual bonds) and the calculation of the Cartesian coordinates of the interacting sites given by eqs 4 and 6. For clarity, the vectors are marked with arrows over the respective symbols. The vector  $\mathbf{dC}_0$  defines the position of the first atom of the chain in the reference system.

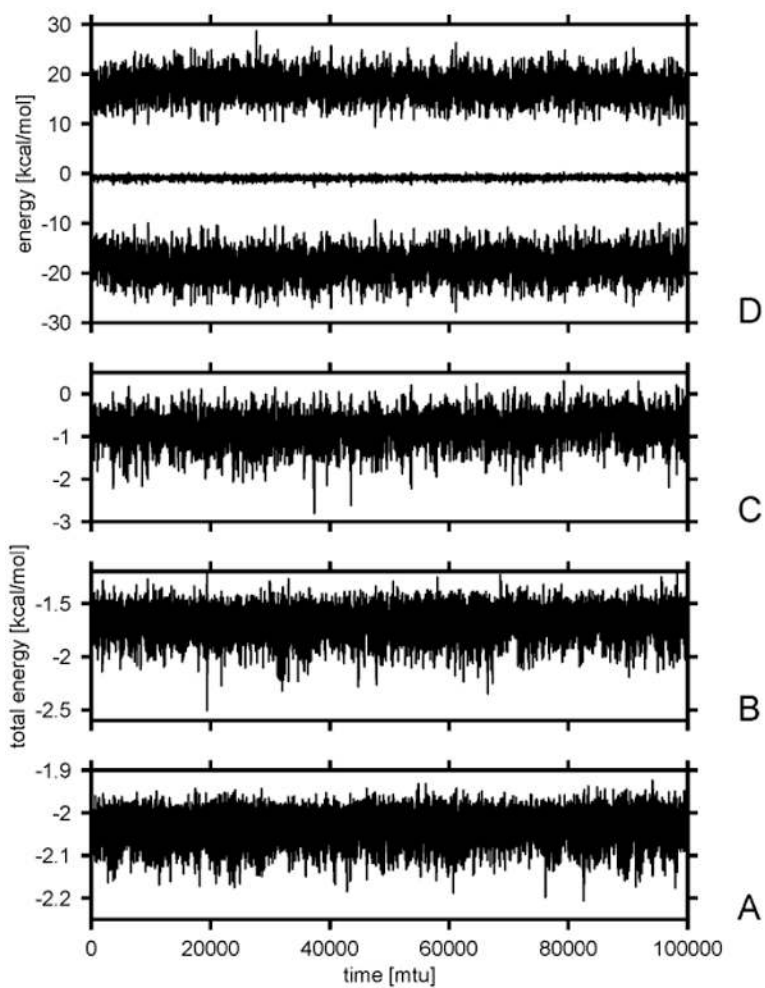


**Figure 2.**

(A) Total energy of the Ala<sub>10</sub> system as a function of time for the time step  $\delta t = 0.01$  and  $0.02$  mtu (solid overlapping lines for both time steps),  $0.03$  mtu (dotted line), and  $0.05$  mtu (dot-dashed line). It can be seen that, for too large values of  $\delta t$ , the total energy increases in a stepwise manner; this usually corresponds to jumps in the acceleration. (B) Variation of the kinetic (solid line), potential (dashed line), and the total (dotted line) energy for the Ala<sub>10</sub> system with the time step  $\delta t = 0.02$  mtu ( $0.978$  fs). (C) Plot of the total energy of panel A for  $\delta t = 0.03$  mtu (solid line) and the maximum difference in acceleration at the beginning and at the end of a given time step ( $|\delta a|$ , dashed line). It can be seen that the spikes in  $|\delta a|$  correlate with the jumps in the total energy.

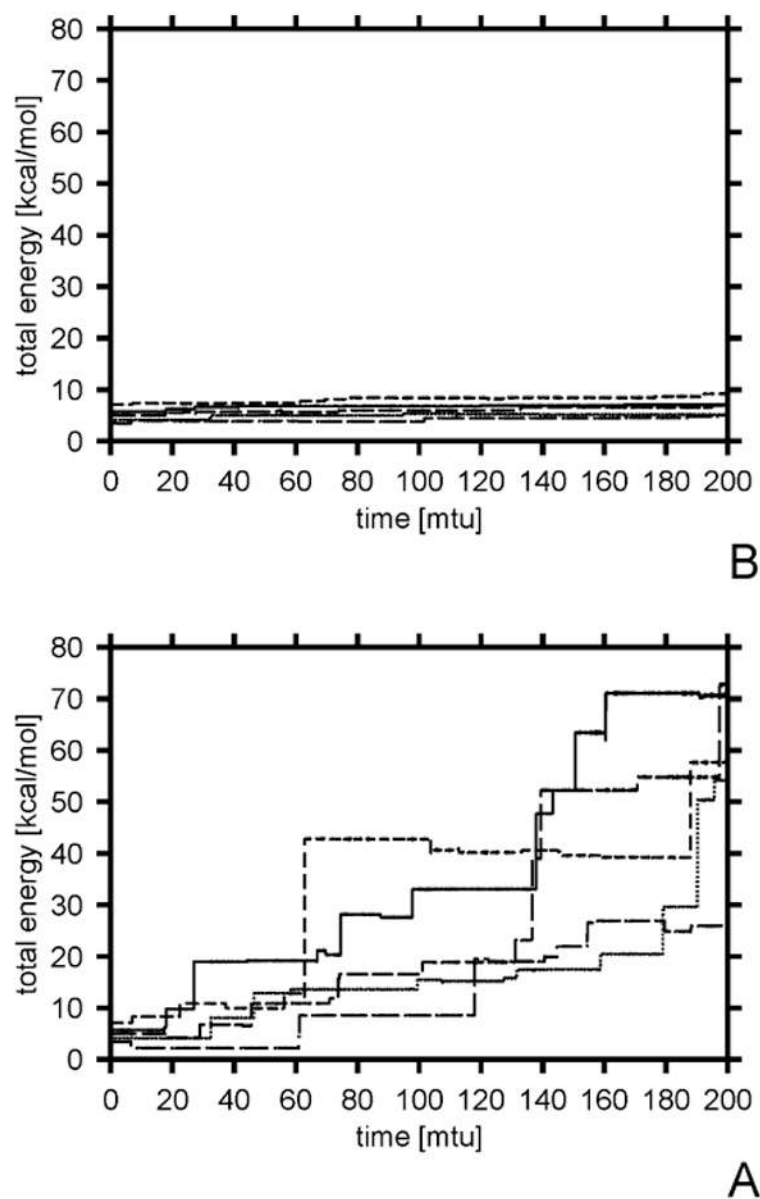


**Figure 3.** Variation of the total energy of the Gly<sub>10</sub> system with time, for the time step  $\delta t = 0.05$  (A), 0.10 (B), 0.15 (C), and 0.20 mtu (D) and the variation of the potential (lower curve), kinetic (upper curve), and total (middle curve) energy with time for the time step  $\delta t = 0.15$  mtu (E).

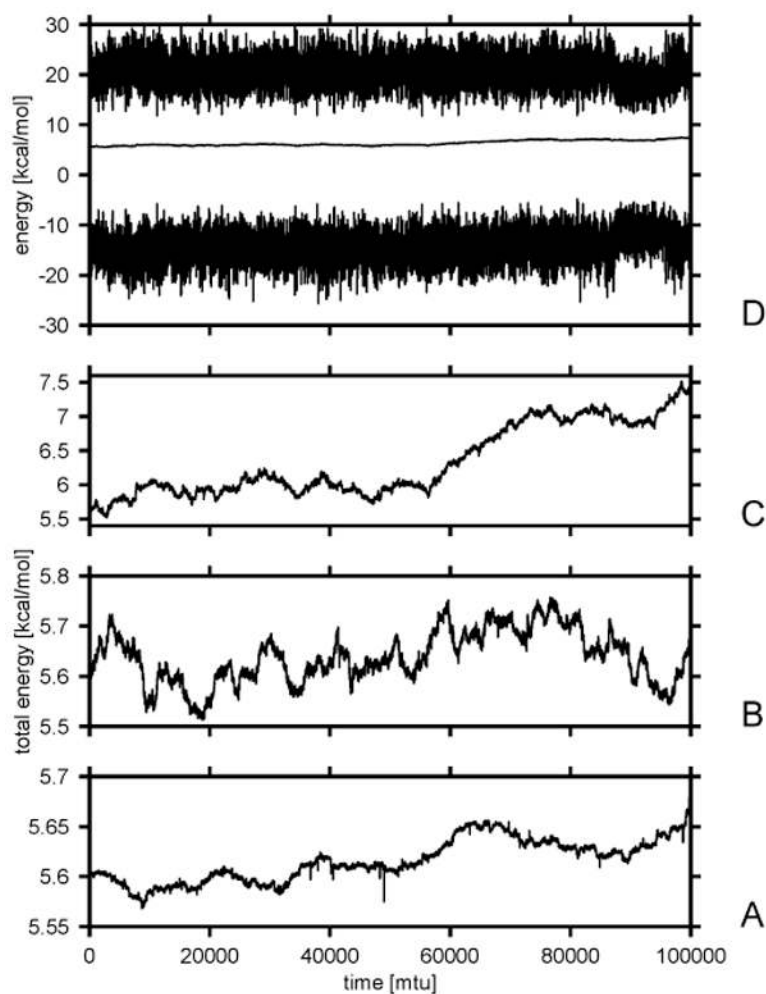


**Figure 4.** Variation of the total energy of the Ala<sub>10</sub> system, with the  $U_{\text{rot}}$  component of the UNRES energy function replaced by a simple harmonic term defined by eq 41, with time for the time step  $\delta t = 0.05$  (A), 0.10 (B), and 0.15 mtu (C) and the variation of the potential (lower curve), kinetic (upper curve), and total (middle curve) energy with time for the time step  $\delta t = 0.15$  mtu (D).

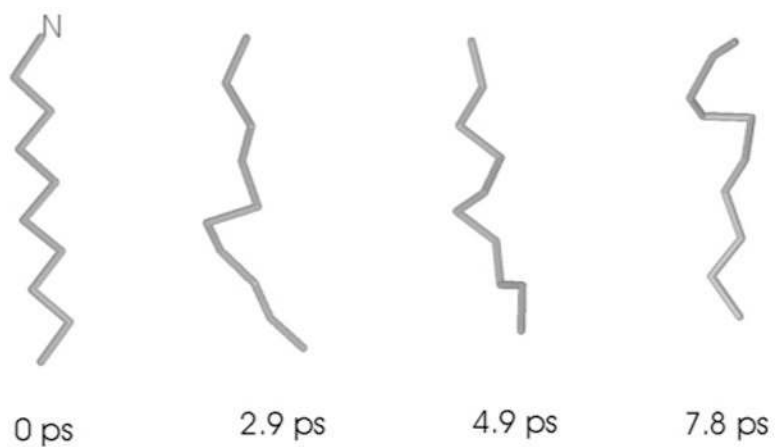




**Figure 5.** Comparison of the variation of the total energy of the Ala<sub>10</sub> system with molecular dynamics step for the fixed (A) and variable-time-step integration algorithm (B) for five pairs of trajectories with  $\delta t = 0.05$  mtu. Trajectories are distinguished by line style; the trajectories corresponding to the same line styles in panels A and B were started from the same initial conditions.



**Figure 6.** Variation of the total energy of the Ala<sub>10</sub> system with time after implementing the reversible RESPA algorithm<sup>65</sup> for the time steps  $\delta t_L/\delta t_S = 0.02$  mtu/0.004 mtu (A), 0.05 mtu/0.005 mtu (B), and 0.10 mtu/0.005 mtu (C) and the variation of the kinetic (upper curve), potential (lower curve), and total (middle curve) energy for the time steps 0.10 mtu/0.005 mtu.



**Figure 7.** Snapshots of the microcanonical-mode trajectory of the Ala<sub>10</sub> polypeptide, starting with a fully extended structure obtained with  $\delta t = 0.03$  mtu. It can be seen that a partially  $\alpha$ -helical structure forms after 4.9 ps of simulation and starts to unfold after 7.8 ps of the simulation.

**Table 1**  
**Time Step  $\delta t$  for Some of the Conventional MD Algorithms**

force field	explicit solvent	implicit solvent
AMBER <sup>a</sup>	1 fs	2 fs
CHARMM <sup>b</sup>	3 fs	4–5 fs
TINKER <sup>c</sup>	1 fs	2 fs

<sup>a</sup><http://amber.scripps.edu/>.

<sup>b</sup><http://www.charmm.org/>.

<sup>c</sup><http://dasher.wustl.edu/tinker/>.

**Table 2**  
**Time Scales Associated with Some of the Biological Activities of the Cell<sup>a</sup>**

local motions	rigid-body motion	large-scale motion
0.01 to 5 Å, 10 <sup>-15</sup> to 10 <sup>-1</sup> s	1 to 10 Å, 10 <sup>-9</sup> to 1 s	>5 Å, 10 <sup>-7</sup> to 10 <sup>4</sup> s
atomic fluctuation side-chain motion loop motion	helix motion domain motion subunit motion	helix-coil transition dissociation/association folding/unfolding

<sup>a</sup> Adapted from Brooks, C. L., III; Karplus, M.; Pettitt, B. M. *A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*; Wiley: 1988; pp 19–20, Table 1.

Table 3

Drift of the Total Energy from the Initial Value ( $|\delta E(t)| = |E_0 - E(t)|$ ), Where  $E_0$  Is the Initial Total Energy and  $E(t)$  Is the Energy at Time  $t$ ) Computed from the Trajectory of the Unblocked Ala10 Run in the Microcanonical Mode with a Fixed Time Step for Various Values of the Time Step  $\delta t$  and Duration of a Run<sup>a</sup>

mtu	$\delta t$ (fs) <sup>a</sup>	$ \delta E $ (kcal/mol)			
		0.1 ps	1.0 ps	5.0 ps	10.0 ps
0.01	(0.489)	0.004	0.02	0.09	0.13
0.02	(0.978)	0.018	0.17	0.57	0.85
0.03	(1.476)	0.040	1.61	6.3	20.
0.04	(1.956)	0.015	2.73	7.8	16.
0.05	(2.445)	0.11	2.48	19.	46.

<sup>a</sup>The values are averages over 10 trajectories.

<sup>b</sup>Values in parentheses are in femtoseconds.

Table 4

Drift of the Total Energy from the Initial Value ( $|\delta E|$ ) and the Average Reduction of the Initial Time Step ( $t/(n\delta t)$ ,  $t$  Being the Total Length of the Trajectory and  $n$  Being the Number of MD Steps) Computed from the Trajectories of the Unblocked Ala10 Run in the Microcanonical Mode with a Variable Time Step for Various Values of the Time Step  $\delta t$  and Duration of a Run<sup>a</sup>

mtu	$\delta t$ (fs) <sup>b</sup>	$ \delta E $ (kcal/mol)					$t/(n\delta t)$
		0.1 ps	1.0 ps	5.0 ps	10.0 ps		
0.02	(0.978)	0.018	0.06	0.27	0.50	0.995	
0.05	(2.445)	0.11	0.52	1.4	2.2	0.977	
0.10	(4.89)	0.47	0.70	7.1	8.8	0.917	
0.20	(9.87)	4.2	32.	154.	291.	0.752	

<sup>a</sup>The values are averages over 10 trajectories and  $|\delta \alpha_{\max}| = 1 \text{ kcal}/(\text{mol} \text{ \AA} \text{ mtu}^2)$ .

<sup>b</sup>Values in parentheses are in femtoseconds.