

Article

An Efficient Local Molecular Dynamics Polymerization Simulation Combined with an *Ab Initio* MO Method

Peng Xie ¹, Yuuichi Orimoto ² and Yuriko Aoki ^{2,3,*}

¹ Department of Molecular and Material Sciences, Interdisciplinary Graduate School of Engineering Sciences, Kyushu University, 6-1 Kasuga-Park, Fukuoka 816-8580, Japan;

E-Mail: ak-x-peng@mms.kyushu-u.ac.jp

² Department of Material Sciences, Faculty of Engineering Sciences, Kyushu University, 6-1 Kasuga-Park, Fukuoka 816-8580, Japan; E-Mail: orimoto.yuuichi.727@m.kyushu-u.ac.jp

³ Japan Science and Technology Agency, CREST, 4-1-8 Hon-chou, Kawaguchi, Saitama 332-0012, Japan

* Author to whom correspondence should be addressed; E-Mail: aoki.yuriko.397@m.kyushu-u.ac.jp; Tel.: +81-92-583-8834; Fax: +81-92-583-8834.

Received: 17 December 2012; in revised form: 22 February 2013 / Accepted: 4 March 2013 /

Published: 6 March 2013

Abstract: A new local *ab initio* molecular dynamics method, namely elongation molecular dynamics (ELG-MD) is proposed for highly efficient simulations of aperiodic polymer systems. ELG-MD combines the elongation method (ELG) with the Gear predictor corrector (GPC) algorithm of molecular dynamics simulation. In this method, the local gradients acting on the atom's nucleus in the active region are calculated by the ELG method while the equations of the nucleus's motion are solved by the GPC algorithm. In this work, the first application of this ELG-MD method is described to investigate the stable conformation of polyglycine with surrounding water molecules. The water effects on the structure of polyglycine are examined. The ELG-MD simulations show that the formation of the polyglycine helix is strongly induced by the hydrogen bonds observed in two types of H-bond rings.

Keywords: elongation method; molecular dynamics; helix forming; hydrogen bond

1. Introduction

Computer simulations on biological macromolecules have been a hot area for numerous theoretical studies describing biological processes at the molecular level. One of the popular treatments for this purpose is the classical molecular dynamics simulation method in which empirical interatomic potential functions are used. However, it is not so accurate because of the treatment of constant atom charges. On the other hand, in *ab initio* molecular dynamics (AIMD), the electronic structure, energy and forces of the system are directly computed based on quantum mechanics theory. However, the number of atoms one can handle in the quantum mechanical treatment is much smaller than the number of atoms in a biological system because of the limitation of computer resources. The elongation (ELG) method has proved to be an efficient method for calculating the electronic structure of large aperiodic polymers with high accuracy at the *ab initio* level [1–6]. In addition, the analytical energy gradients can be calculated by the ELG method and have been successfully used for the geometry optimization of a series of polymers [7]. Based on these developments, we here propose a new efficient AIMD method named Elongation-MD (ELG-MD) by combining the elongation method with the dynamics algorithm. The ELG-MD method makes it possible to efficiently analyze the mechanisms of chemical reactions by considering dynamics even for huge random polymers such as biomaterials. For the first application of the ELG-MD method to confirm its effectiveness, we selected polyglycine as a target system; polyglycine is representative of the simplest peptide and widely investigated both experimentally and theoretically.

Polyglycine (Gly)_n, containing n residues, forms the backbones of amino acids, peptides and proteins without side chains and their functional groups [8]. Therefore, the investigations on (Gly)_n are important to the understanding of a wide range of biological systems. In particular, the three-dimensional structure is one key factor for the biological activities of a peptide, and hydrogen bonds play important roles in stabilizing the secondary structure of peptides [9]. In general, hydrogen bonds (H-bonds) can be generated both intramolecularly and intermolecularly in the system. In the former case, the H-bonds are formed between many different functional groups in the peptide, but in the latter case the H-bonds are formed between polar groups of the peptide and the surrounding solvent molecules such as water molecules. For a β strand, the hydrogen bonds are formed by the N–H groups in the backbone of one β strand with the C=O groups in the backbone of the adjacent β strands. For an α helix, each hydrogen bond is formed by the N–H group of an amino acid with the C=O group of the amino acid four residues earlier. On the conformation of polyglycine, Ohnishi *et al.* demonstrated that polyglycine in solution exhibits a strong preference for an extended conformation when polymerized in short segments as the intermediate in the extension between a β strand and an α helix [10]. Which kind of hydrogen bond will stabilize this extended conformation of polyglycine in aqueous solution? The hydration effects via direct hydrogen bonds between water molecules and the peptide backbone can be expected. Here we applied the ELG-MD method to investigate the intermolecular H-bonds effects on conformation stability when (Gly)_n is polymerized with water molecules. This application deals with the conformation change of (Gly)_n under the explicit consideration of water molecules, and the AIMD simulation in which energies and forces are generated by the ELG method at each time step.

This paper is organized into four sections. In Section 2, we briefly introduce the ELG formalism including energy and gradient calculations, after that we present a description of the ELG-MD method.

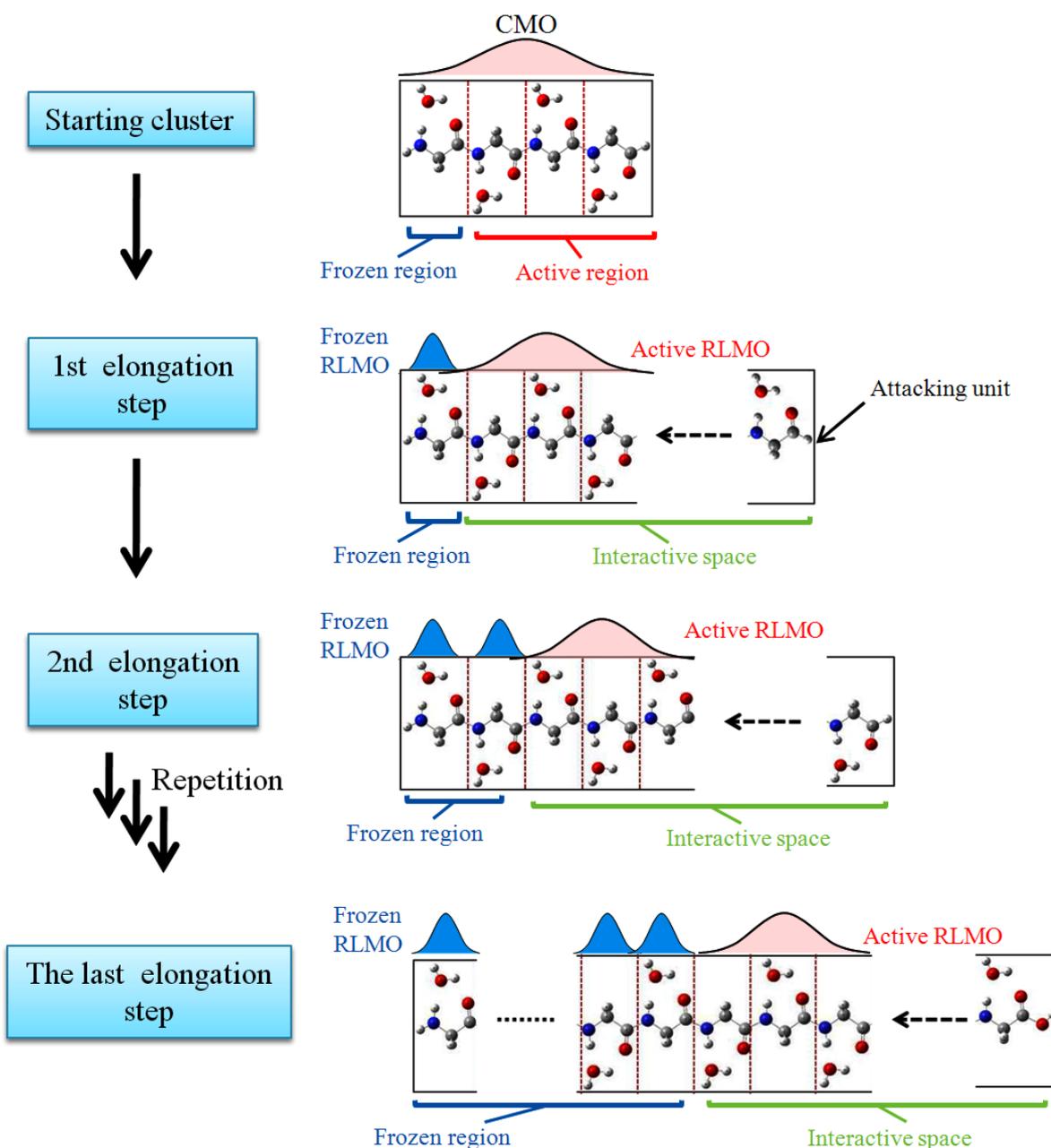
In Section 3, we introduce the structure of model molecules, *i.e.*, polyglycine with water molecules. In the same section we analyze the equilibrium structures of polyglycine and their energetic stability. Final conclusions are collected in Section 4.

2. Methodology

2.1. Elongation Method

The elongation method can be considered as a procedure for *theoretical* polymerizations on computers [1–3]. We present a brief overview of the procedures (see also the flowchart in Figure 1).

Figure 1. Flowchart of the elongation method illustrated using polyglycine with water molecules. The canonical molecular orbital (CMO) and region localized molecular orbital (RLMO) indicate canonical and regional localized molecular orbitals, respectively.



First, the polymer chain is divided into several monomers, and then a suitable size of monomers as a starting cluster is selected. A and B regions are defined for the starting cluster, where the A region is the frozen region assumed to be the side furthest away from the chain propagation point, and the B region is the active region near the propagation point. The canonical molecular orbitals (CMOs) of the starting cluster are solved by the conventional Hartree-Fock self-consistent-field (HF-SCF) procedure.

Second, the CMOs are converted to localized MOs (LMOs) over the A or B regions. We define the interacting space as the B region with an attacking molecule, and the A region is furthest away from the interaction spot. The localization procedure is explained in the reference [2]. The obtained LMOs are localized in a special region over several units, and are different from those by Ruedenberg's [11] or Boys's [12] localization method that provide LMOs localized on a bond. Thus, we name the obtained orbitals as region localized molecular orbitals (RLMOs). By using the RLMOs, the density matrix can be expressed by

$$D^{\text{RLMO}} = \{L_{\text{CMO}}^{\text{RLMO}}(\text{A}) + L_{\text{CMO}}^{\text{RLMO}}(\text{B})\} P \{L_{\text{RLMO}}^{\text{CMO}}(\text{A}) + L_{\text{RLMO}}^{\text{CMO}}(\text{B})\} \quad (1)$$

where L indicates the transformation matrix between CMO and RLMO. P is the diagonalized matrix in which the degenerates 2 and 0 are given as the eigenvalues for occupied and vacant spaces, respectively. The D^{RLMO} is actually identical to P because the transformation matrix L is defined under the freedom that comes from the non-uniqueness of the coefficients owing to the degeneracy within all the occupied orbitals and all the vacant orbitals. In other words, the occupied and vacant spaces are not mixed with each other by Equation (1). Therefore, the diagonalized density matrix P is invariant under the unitary transformation. By this feature, we can make preferable RLMOs suitable for local interaction during the polymerization process.

Third, at the elongation step, the attacking monomer (M) is attached to the terminal of the B region. Because the A region is far away from the M region, the interaction between them is negligible. Thus, the ELG HF equation is solved self-consistently only for the localized orbital space including the B and M regions. After the new CMOs of the B and M regions are obtained, the new CMOs are localized into new frozen (A') and active (B') regions. Then a new attacking monomer (M') is attached. The above procedures are repeated step by step until the electronic structure of the target system is obtained.

2.2. Elongation Gradient Calculation

In HF approximation, the total energy of a closed-shell system can be written as

$$E = \sum_{\mu\nu} P_{\nu\mu} H_{\mu\nu}^{\text{core}} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} P_{\nu\mu} P_{\lambda\sigma} (\mu\nu || \sigma\lambda) + V_{\text{NN}} \quad (2)$$

where P , H^{core} and S denote the density matrix, core Hamiltonian and overlap matrix of the system, respectively. The $(\mu\nu || \sigma\lambda)$ and V_{NN} represent the two-electron integral and the nuclear-nuclear repulsion, respectively. We can obtain the first derivative (gradient) of the total energy E by the analytic derivative method with the form of explicit dependence on the nuclear coordinate X_A [13] as

$$\frac{\partial E}{\partial X_A} = \sum_{\mu\nu} P_{\nu\mu} \frac{\partial H_{\mu\nu}^{\text{core}}}{\partial X_A} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} P_{\nu\mu} P_{\lambda\sigma} \frac{\partial (\mu\nu || \lambda\sigma)}{\partial X_A} - \sum_{\mu\nu} Q_{\nu\mu} \frac{\partial S_{\mu\nu}}{\partial X_A} + \frac{\partial V_{\text{NN}}}{\partial X_A} \quad (3)$$

where Q is the energy-weighted density matrix defined as

$$Q_{\nu\mu} = 2 \sum_i^{N/2} \varepsilon_i C_{\mu i} C_{\nu i} \quad (4)$$

A detailed description of how to obtain gradient calculations with the elongation method was presented in a recent paper of our group [7]. We sketch this method out here. Because the elongation method is based on AO basis, the density matrix by the ELG method is almost the same as the density matrix in Equation (3) for the conventional calculation except for the third term $Q_{\nu\mu}$. In each elongation step, the HF equations are solved only for the interacting space including the B and M regions. After the ELG-SCF procedure, the molecular orbitals of the B and M regions are transformed from RLMO to AO basis for calculating the gradient in Equation (3). First, we combine C_A^{RLMO} with $C_{\text{BM}}^{\text{MO}}$ to construct the coefficient matrix for the whole system, C^{MO} , where the C_A^{RLMO} and $C_{\text{BM}}^{\text{MO}}$ denote the coefficients of the frozen region (A) from the previous ELG steps and interacting space (B + M) from the ELG-SCF calculation in the present ELG step, respectively. The ε matrix can be calculated by $C^{\text{MO}} F C^{\text{MO}}$, but the ε is not diagonal any more. Therefore, we can get the AO based energy-weighted density matrix Q for the ELG-HF method by

$$Q_{\nu\mu} = \sum_{i,j}^{N/2} n_i \varepsilon_{ij} C_{\mu i} C_{\nu j} \quad (5)$$

where n is the occupancy number matrix corresponding to the density matrix in the MO representation, in this matrix degenerate 2 is given for the occupied space and degenerate 0 is for the vacant space.

2.3. Elongation Molecular Dynamics

We can straightforwardly combine the ELG method with the MD algorithm to construct the ELG-MD method in the following manner. Newton's equation of motion is solved by calculating the ELG gradients on the fly using

$$F = m \frac{d^2 r}{dt^2} = -\nabla E^{\text{ELG}} \quad (6)$$

The time evolution of the atomic nuclei is performed according to the fourth order Gear predictor corrector (GPC) algorithm [14,15]. The GPC is one of the most widely-used algorithms in molecular dynamics simulations. In general, any time-dependent property can be estimated from a Taylor series expansion. In the predictor stage of the GPC, therefore, the atomic coordinate vector r^p can be expanded in the Taylor series in the time step Δt :

$$r^p(t + \Delta t) = r(t) + \frac{dr}{dt} \Delta t + \frac{1}{2} \frac{d^2 r}{dt^2} \Delta t^2 + \frac{1}{6} \frac{d^3 r}{dt^3} \Delta t^3 + \frac{1}{24} \frac{d^4 r}{dt^4} \Delta t^4 + \dots \quad (7)$$

The atomic velocity vector v^p and the acceleration vector a^p could be updated by the other appropriate expansions similar to Equation (7), respectively. We here define the first, second, and third derivative of r as $v = \frac{dr}{dt}$, $a = \frac{d^2 r}{dt^2}$ and $b = \frac{d^3 r}{dt^3}$, respectively.

$$v^p(t + \Delta t) = v(t) + \frac{d^2r}{dt^2} \Delta t + \frac{1}{2} \frac{d^3r}{dt^3} \Delta t^2 + \dots \quad (8)$$

$$a^p(t + \Delta t) = a(t) + \frac{d^3r}{dt^3} \Delta t + \dots \quad (9)$$

$$b^p(t + \Delta t) = b(t) + \dots \quad (10)$$

Although we get a set of predicted vectors from these procedures, they are not correct because of their truncation errors. Thus, all the vectors are corrected in the corrector stage of GPC. From the new predicted coordinates, we can actually calculate the accelerations a^c . The difference between a^c and a^p , *i.e.*, Δa , are used to correct the predicted value.

$$\Delta a(t + \Delta t) = a^c(t + \Delta t) - a^p(t + \Delta t) \quad (11)$$

$$r^c(t + \Delta t) = r^p(t) + c_0 \Delta a(t + \Delta t) \quad (12)$$

$$v^c(t + \Delta t) = v^p(t) + c_1 \Delta a(t + \Delta t) \quad (13)$$

For example, suitable values of c_0 and c_1 for the fourth-order GPC algorithm are 19/120 and 3/4, respectively. In a similar way, the other vectors can be also corrected.

As mentioned above, the ELG method is similar to the procedure for the theoretical synthesis of polymers. When the attacking monomer (M region) approaches the B region, the structure of the interacting space (B and M regions) will change because of the interaction between them. In contrast, the frozen region (A region) is far away from the M region and thus we can presume there is no interaction between the A region and the M region. Therefore, the structure of the A region will not change any more when M approaches B. For this reason, in the ELG-MD calculation, we only perform molecular dynamics on the interacting space in the last step of ELG-MD. For the previous ELG steps, only energy calculations are carried out. The flowchart of the ELG-MD method is presented in Figure 2. On the right side of Figure 2, it is shown that the A, B and M regions are renewed for the next elongation step after localization, so A denotes a different region in a different elongation step, as do B and M. All the procedures of the ELG-MD method were implemented with the GAMESS program package [16].

3. Results and discussion

A (Gly)₁₄ peptide in β -strand conformation with 14 water molecules was selected as a model system to test and demonstrate the ELG-MD method (Figure 3). For the first testing calculation for ELG-MD, we did not consider more water molecules around polyglycine. One reason was to save computing time, another reason is that for the central fragment of the glycine molecule, one side including C=O and N-H groups is hydrophilic and these two groups can form a hydrogen bond ring C₇(C=O...H-O...H-N) (shown in Figure 4b) with one water molecule, it is not easy to add more water molecules forming hydrogen bonds with these two groups in this area because of space limitation. Another side including the CH₂ group is hydrophobic. Here, the notation C_n(X...H-O...Y) represents the hydrogen bond ring structure consisting of n atoms enclosed by the H-O group of the adjacent

water molecule. The (Gly)₁₄ is divided into seven units ((Gly)₂ for each unit), and assigned A, B, and M₁~M₃ regions as shown in Figure 3. The simulation in the present article has two stages.

Figure 2. The flowchart of the elongation molecular dynamics (ELG-MD) method. The “Energy and Density (B + M)” means that the energy and density of the B + M region are obtained with the contribution from the A region.

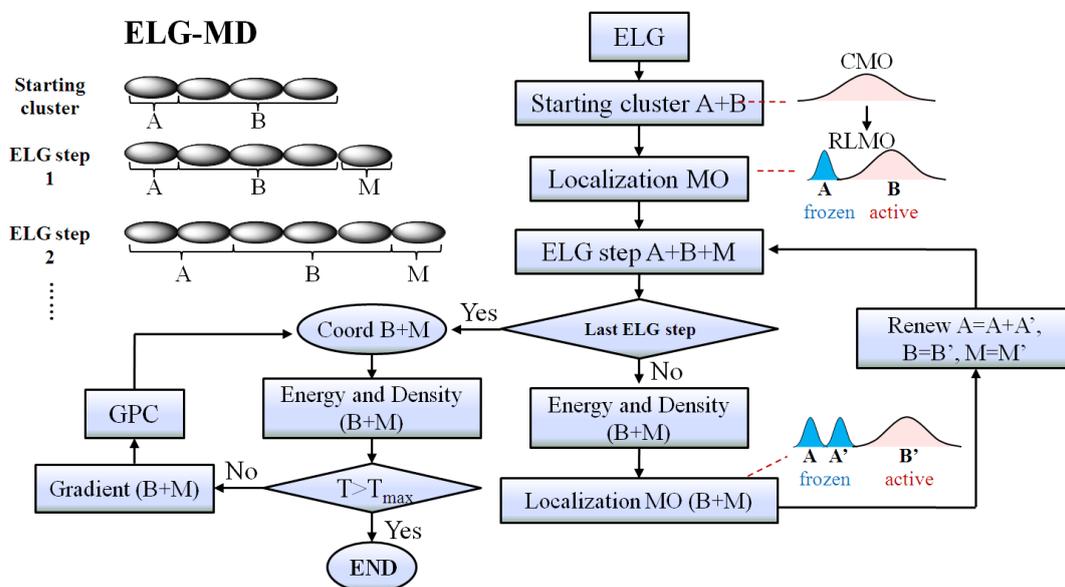
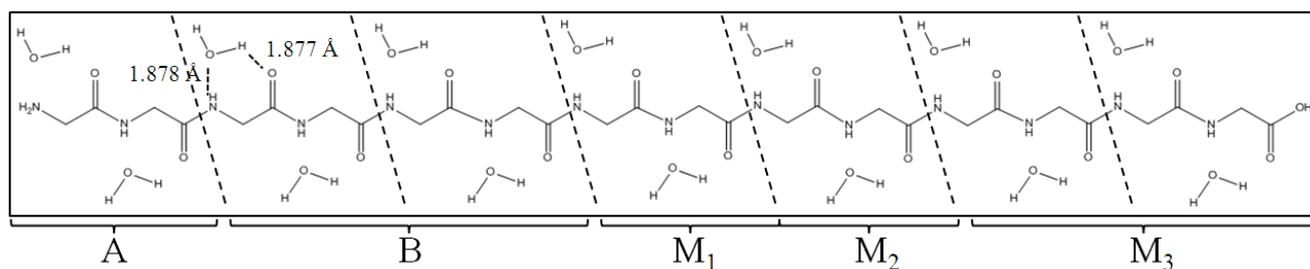


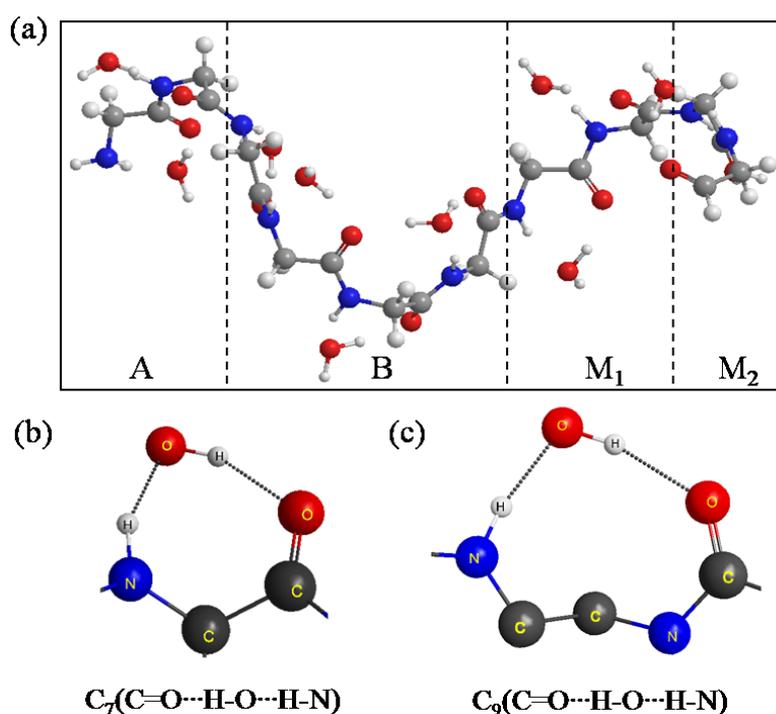
Figure 3. Polyglycine (Gly)₁₄ in β -strand conformation with 14 water molecules. (Gly)₁₄ are divided into A, B, M₁, M₂ and M₃ regions for the ELG-MD procedures. The notations A and B do not correspond to those in Figure 2.



First, we need a stable conformation of A, B, M₁, and M₂ regions to create a more real surrounding for the new attacking monomer (M₃ region) before investigating the structure of (Gly)₁₄ including the A, B, and M₁~M₃ regions. Several research groups have devised experiments for measuring β -sheet stability with various peptides, and they found polyglycine has a low β -sheet propensity [17–20]. Therefore, to find a more reasonable conformation than β -sheet conformation, the conventional AIMD simulation was performed for the A, B, M₁ and M₂ regions (other than the M₃ region). In the AIMD simulation, the energies and forces were obtained at the HF level of theory with the STO-3G basis set. The time step is set as 0.5 fs and we perform in total 5 ps simulation (10,000 steps). The temperature of the system is associated with the classical kinetic energy and the constant temperature (298 K) is achieved by the algorithm proposed by Berendsen *et al.* [21]. Figure 4a shows the equilibrium structure obtained by the simulation. It was found that the conformation of A, B, M₁ and M₂

significantly changed from a β -sheet type to a quasi- α -helix type in which the structure is more outstretched than α -helix. In general, α -helix forms a right-handed helix and every N–H group in the backbone donates a hydrogen bond to the C=O group of the amino acid belonging to four residues earlier ($i + 4 \rightarrow i$ hydrogen bonding) [22]. Our results in Figure 4a show that the right-handed helix is induced by two kind of hydrogen bond rings, that is, $C_7(C=O \cdots H-O \cdots H-N)$ in Figure 4b and $C_9(C=O \cdots H-O \cdots H-N)$ in Figure 4c. $C_9(C=O \cdots H-O \cdots H-N)$ shows that the H–O group of the water donates two hydrogen bonds to the C=O and N–H groups of the backbone for instance.

Figure 4. (a) The structure of the A, B, M_1 and M_2 regions of $(Gly)_{14}$ in a quasi- α -helix conformation as the initial structure for ELG-MD simulation. Panels (b) and (c) show two types of H-bond ring. The dotted lines in the panels denote hydrogen bonds.



Second, for the ELG-MD calculation, there are four calculation steps including starting cluster (A and B regions), the first elongation step (A, B and M_1 regions), the second elongation step (A, B, M_1 and M_2 regions) the third elongation step (A, B, M_1 , M_2 and M_3 regions). Energy calculations are carried out for the first three steps, and MD only performed for the last step. In the last step, we presume that a new monomer M_3 (β form) furthermore attacks the propagating site of the M_2 region of the structure obtained above. Therefore, the initial structure for the ELG-MD method consists of the A + B + M_1 + M_2 equilibrium structure (quasi- α form) and an additional M_3 region (β form). The $M_1 \sim M_3$ region of the initial structure is shown in Figure 5. Because we are interested in the structure variation of M_3 and the associated regions such as the M_1 , M_2 regions, the MD treatment in the ELG-MD procedure is applied only for the interacting space in the last elongation step, that is, the $M_1 \sim M_3$ region while the A and B regions are frozen. For the dynamics part of the ELG-MD calculation, we used the same setting as conventional AIMD simulation.

To know the efficiency of the ELG-MD method, 2.5 ps (5000 steps) simulation was performed by both ELG-MD and conventional AIMD for the initial structure in Figure 5. The total wall clock time is

206,362.3 s (ELG-MD) and 298,391.3 s (conventional AIMD), respectively. Because the ELG-MD is a local MD method, its efficiency is obviously higher (about 31%) than the conventional AIMD. Furthermore, the potential energies (HF energies) of the entire system can be compared between the two treatments (see Figure 6).

Figure 5. The initial structure of the M_1 , M_2 and M_3 regions (atoms from 62 to 143). C60 and O61 belong to the B region.

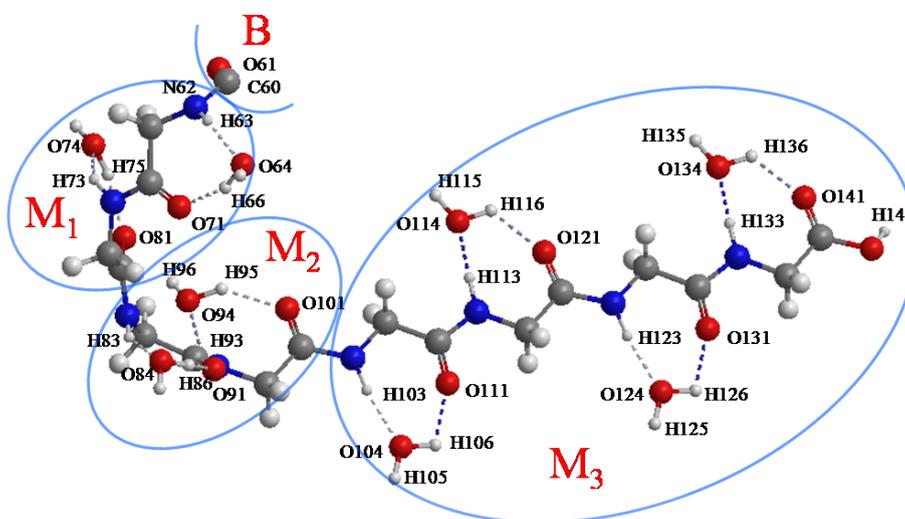
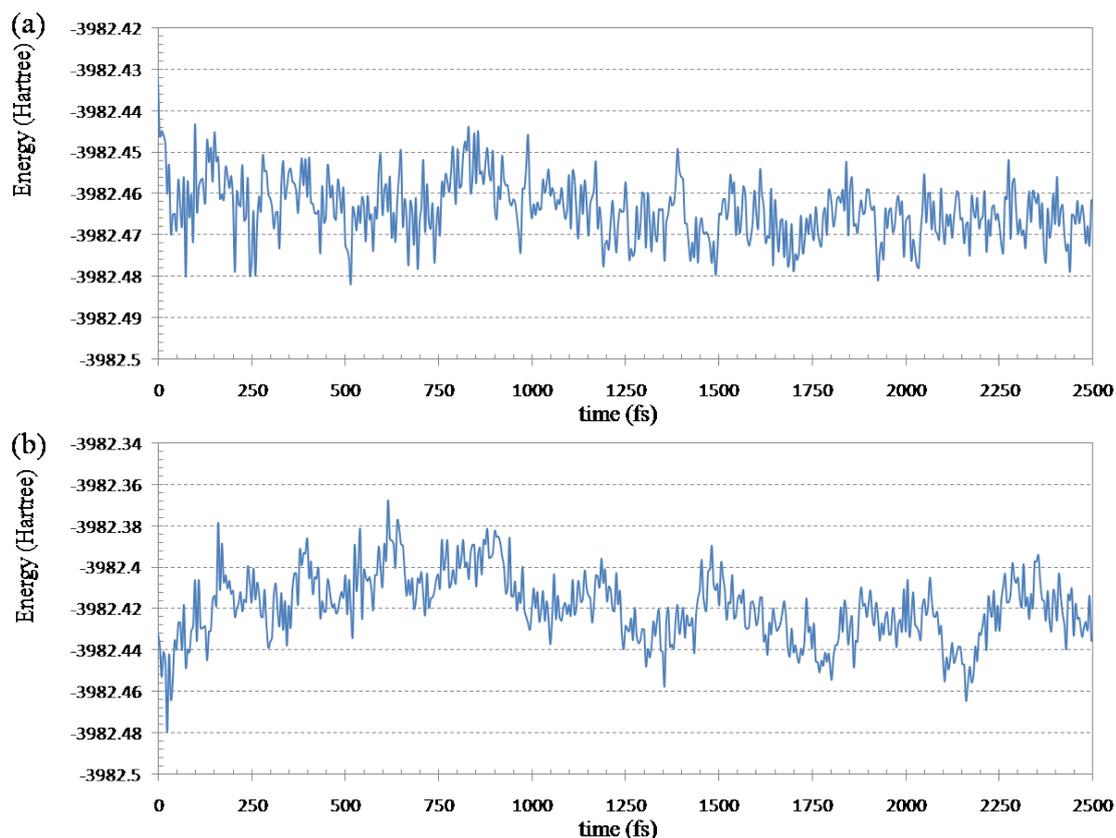


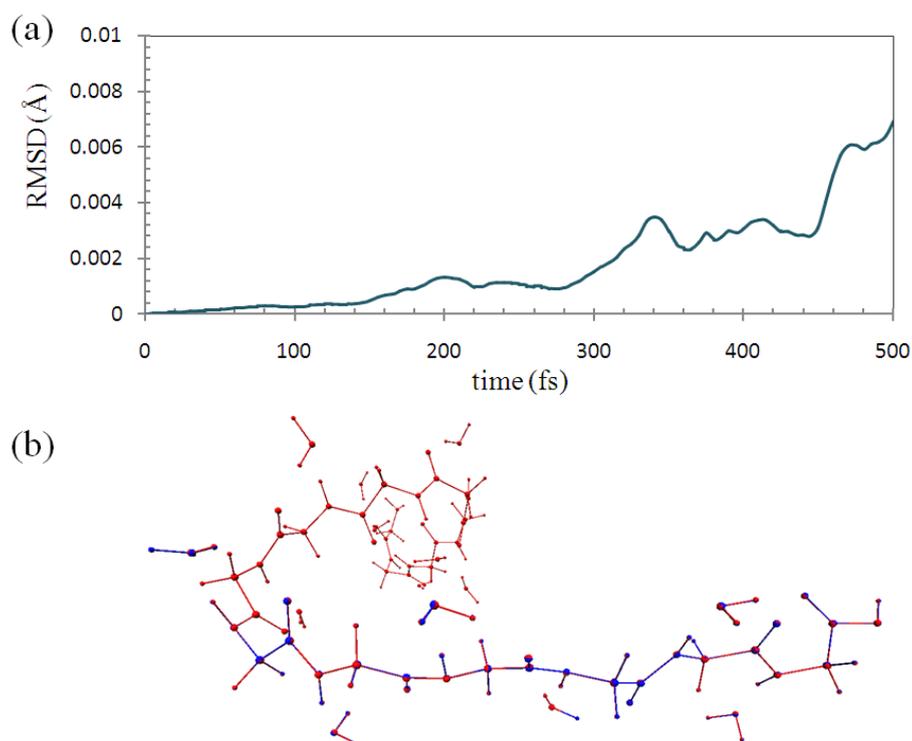
Figure 6. Fluctuations of the Hartree-Fock energy of $(\text{Gly})_{14}$ with 14 water molecules in the simulation at 298.15 K: (a) ELG-MD simulation; (b) conventional *ab initio* MD simulation.



From Figure 6a we can see the HF energy decreases quickly for the ELG-MD method which relaxes only a local region of the polymer. In the ELG-MD calculation, the structure of the A and B regions is automatically frozen when the M_3 region attaches, which is different from the conventional AIMD which treats the whole system at the same time. Because the freedom of the ELG-MD calculation is much less than conventional AIMD, it is much easier for the ELG-MD method to get the energy convergence. Figure 6b shows the fluctuations of the Hartree-Fock energy calculated by conventional AIMD. The energy increases during 900 fs, and then the energy begins to decrease slowly. That is to say, it is hard to get the energy equilibrium in a short period of time in the conventional AIMD. It also should be stressed that the energy of ELG-MD is lower than that of conventional AIMD at 2500 fs. From 2000 to 2500 fs, the average total energy is -3982.4664 Hartree for ELG-MD while -3982.4245 Hartree for AIMD. When the new monomer attaches to the propagating site of the M_2 region, it will mainly affect the structure of the M_1 and M_2 regions because the A and B regions are far away from the M_3 region (the distance between them is about 10 \AA). Therefore, we only focus on the conformation variation of the attacking monomer (M_3) and its vicinity (M_1 and M_2) under the assumption that the A and B regions have no interactions with the M_3 region.

To check the accuracy of the ELG-MD, we performed 0.5 ps (1000 steps) conventional AIMD calculation for the polymer with fixing of the atoms in the A and B regions in the same manner as with the ELG-MD calculation. We compared the structures of them in each simulation step. Figure 7a shows the root-mean-square deviation (RMSD) of the backbone part ($M_1 \sim M_3$) of the structure between the ELG-MD and conventional AIMD with fixing the atoms. The maximum value of RMSD with 0.00695 \AA is considerably small. This shows that the two structures agree well in each simulation step.

Figure 7. (a) Root-mean-square deviation (RMSD) of the molecular structure between ELG-MD and conventional *ab initio* MD with fixing of the same atoms; (b) Superimposed structures at 500 fs. (Red: conventional AIMD. Blue: ELG-MD).



As shown in Figure 7b, the two structures at 500 fs (1000 steps) calculated by the two methods are almost the same as each other. Hence, it can be concluded that the trajectories of the two simulations have essentially the same quality, and the structure obtained by the ELG-MD simulation is credible.

Now, we performed 5 ps (10,000 steps) ELG-MD simulation for the (Gly)₁₄ peptide starting from the initial structure in Figure 5 to investigate the conformation variation from the viewpoint of the H-bond effects between peptide and water molecules. It should be noted that the frozen region (A and B) is not shown in Figure 5 because the coordinates of the frozen regions are fixed. As mentioned above, in the initial structure the M₃ region has β -strand confirmation and the other region A + B + M₁ + M₂ has quasi- α -helix conformation. All of the H-bonds in the initial structure are of the C₇(C=O \cdots H–O \cdots H–N) type. We measured all of the H-bond lengths during the simulation to examine the configuration change. We assigned these H-bonds into eight groups according to their locations. The changes in the atomic distances of these H-bonds are shown in Figure 8.

Figure 8a shows that the H-bonds (H63 \cdots O64 and H66 \cdots O71) near the curve region in the helix are relatively stable throughout the simulation. Figure 8c–g show that the number of H-bonds is always maintained at two in these areas. In addition, the variation of H-bonds only occurs between O of the backbone's C=O group and the two H atoms of the water molecule in the vicinity induced by the rotation of the water molecule. Figure 8b shows that the H-bond of H75 \cdots O81 changed to that of H76 \cdots O81 at 3330 fs (Figure 9a), and a new H-bond O61 \cdots H75 in C₉(C=O \cdots H–O \cdots H–N) type is observed at 4300 fs (Figure 9b). Then, it was found that the H-bonds H76 \cdots O81 and O61 \cdots H75 became stable after 4300 fs. The formation of new C₉(C=O \cdots H–O \cdots H–N) type H-bonds implies that the helix structure of this area becomes more stable and more compact than the initial structure. The variations of H-bonds in the tail region of the peptide are more active than the other regions as shown in Figure 8h. After 4300 fs, the H135 \cdots O141, H135 \cdots O121 and H136 \cdots O121 distances have a downward trend, but the H136 \cdots O141 distance increases in contrast. This implies that the H-bond of H136 \cdots O141 gradually changes to H135 \cdots O141, and it has a tendency to form a new H-bond H136 \cdots O121 in the C₉(C=O \cdots H–O \cdots H–N) type as shown in the snapshot at 5000 fs (Figure 10c). The snapshot structures of interacting space (M₁, M₂ and M₃) and the entire peptide at 5000 fs are shown in Figure 10a,b, respectively. All the C₇(C=O \cdots H–O \cdots H–N) type H-bonds in the peptide interacting space are stable during the simulation. On the other hand, a new H-bond ring in the C₉(C=O \cdots H–O \cdots H–N) type is formed in front of the interacting space, and the structure of the tail region has the tendency to form another H-bond ring of the C₉(C=O \cdots H–O \cdots H–N) type. The H-bonds in the C₉(C=O \cdots H–O \cdots H–N) type can cause the distortion of the structure leading to the forming of a helix conformation. It can be concluded that the β -strand conformation included in the interacting space (M₃ region) is transformed into a helix conformation due to the newly formed H-bonds.

Figure 8. Fluctuations of the H-bond distances between peptide and water molecules in the interacting space (5 ps ELG-MD simulation). The numbering of the atoms is shown in Figure 5. M_1 , M_2 and M_3 denote the M_1 region, M_2 region and M_3 region shown in Figure 5 respectively. More details of the H-bond distances in Panel b and h are shown in Figures 9 and 10, respectively.

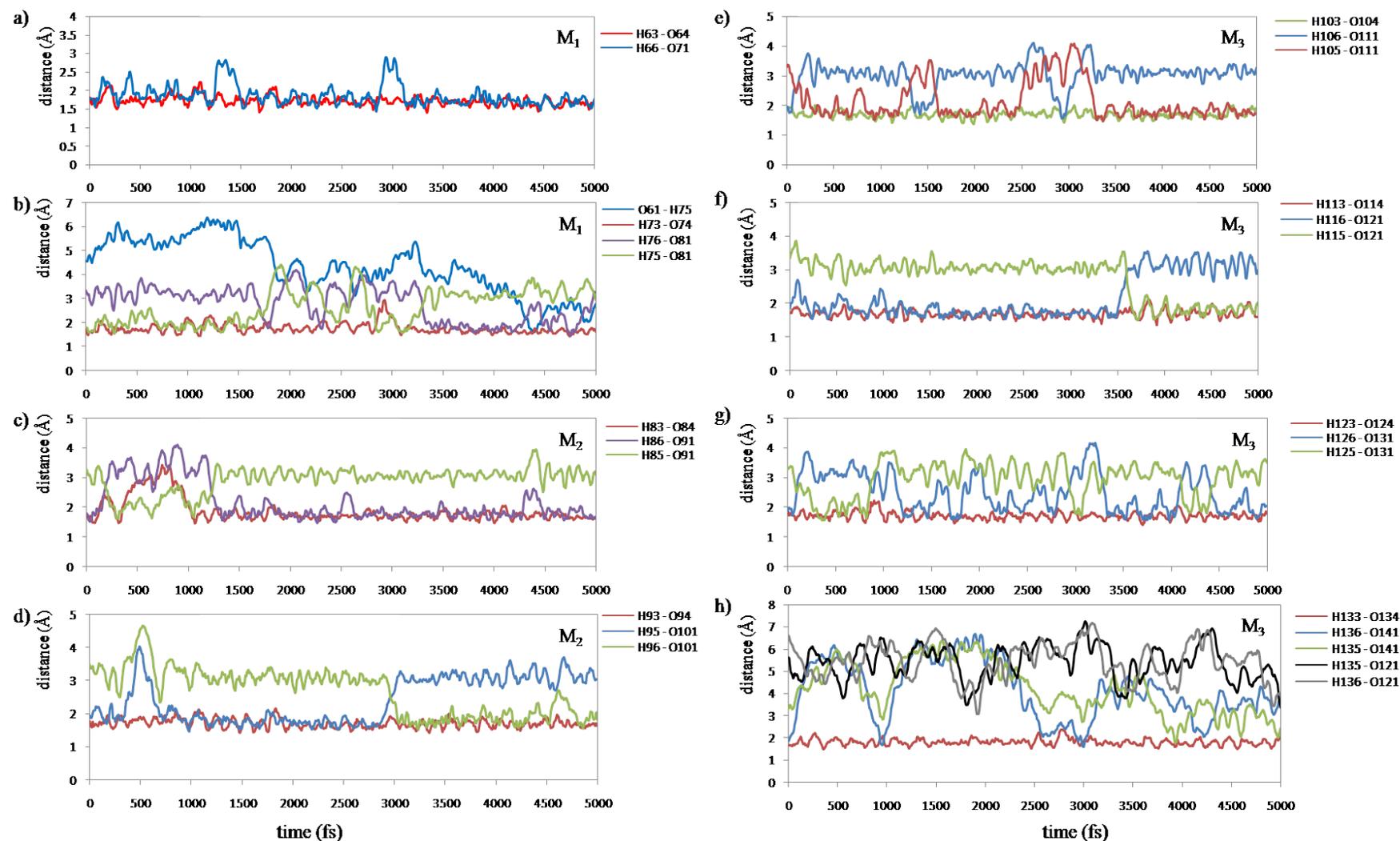


Figure 9. Snapshot of H-bond rings in the M₁ region at (a) 3330 fs and (b) 4300 fs.

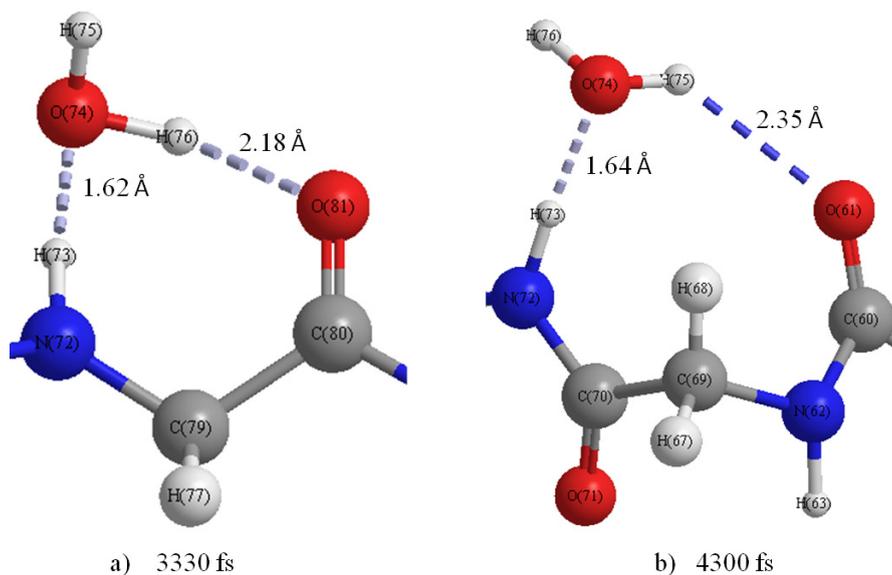
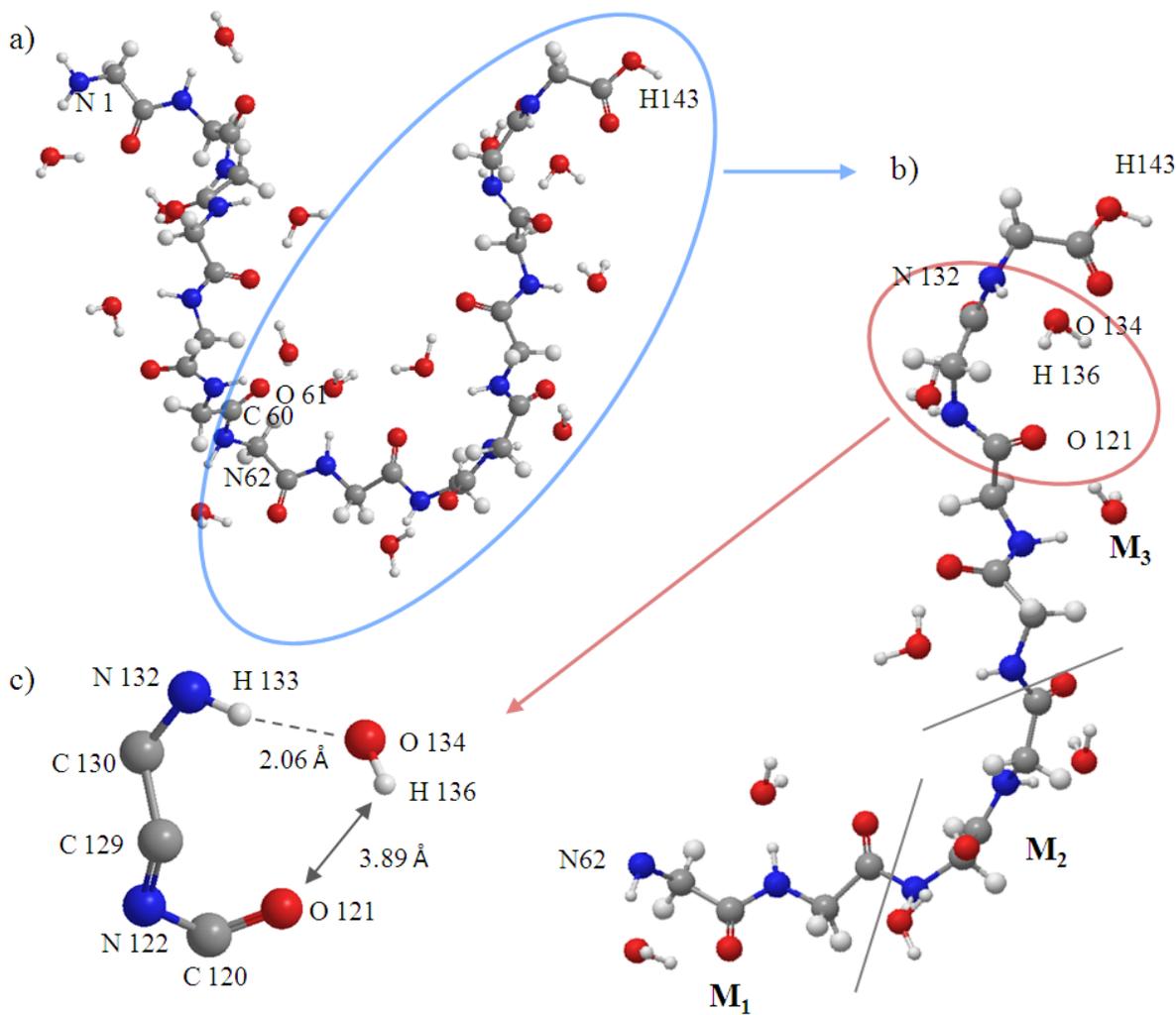


Figure 10. Snapshot at 5000fs of (a) entire peptide (Gly)₁₄. Atoms from 1 to 61 belong to the frozen region; atoms from 62 to 143 belong to the active region; (b) Interacting space (M₁, M₂ and M₃); and (c) H-bond.



It is a big challenge to understand the interactions in biomolecules at the atomic level with the surrounding water molecules because of their transient character and the inhomogeneity of H-bonding in liquid water [23]. Here, we have proposed a new kind of theoretical investigation on the interactions between peptide and explicit water molecules, and have been able to show the H-bond effects on the conformation transformation of the peptide as a first AIMD application.

4. Conclusions and Prospects

In this study, we presented the theoretical framework of the ELG-MD method and applied it to polyglycine (Gly)₁₄ with 14 water molecules. Firstly, we tested ELG-MD over a short period of time and compared the results with those of conventional AIMD. We found from the comparison that the efficiency of ELG-MD is higher than the conventional AIMD method (about 31%). In addition, ELG-MD can easily be used to obtain the energy convergence of the system by treating only the local interacting space. The two methods produced similar trajectories with fixing of the same atoms. The advantage on efficiency and accuracy demonstrates the effectiveness of our ELG-MD method. Secondly, we performed 5ps (10,000 steps) ELG-MD simulation for this system. It was found from the simulation that the β -strand conformation in the (Gly)₁₄ can be transformed into a random helix conformation by forming new H-bonds with the surrounding water molecules.

Although ELG-MD is not a full molecular dynamics, it can perform simulation in the active region which we are interested in, at the end of a polymer. In chain growth polymerization, the new monomer molecules add onto the active site on a growing polymer chain. The ELG-MD method is suitable for this kind of simulation. Many chemical reactions in biomolecular systems are similar to this process such as peptide chain elongation, the binding between inhibitor and enzyme *etc.* Although ELG-MD can only handle very simple quasi-one-dimensional systems at this stage, three-dimensional ELG-MD for more complicated real biomolecular systems is in progress [24].

Acknowledgments

P. Xie would like to thank the financial support of the JST-RA system by the group CREST, Japan Science and Technology Agency (JST). The calculations were carried on the Linux PC cluster in our laboratory and the high performance computing systems of the Research Institute for Information Technology at Kyushu University.

References

1. Imamura, A.; Aoki, Y.; Maekawa, K. A theoretical synthesis of polymers by using uniform localization of molecular-orbitals-proposal of an elongation method. *J. Chem. Phys.* **1991**, *95*, 5419–5431.
2. Gu, F.L.; Aoki, Y.; Korchowiec, J.; Imamura, A.; Kirtman, B. A new localization scheme for the elongation method. *J. Chem. Phys.* **2004**, *121*, 10385–10391.
3. Korchowiec, J.; Gu, F.L.; Imamura, A.; Kirtman, B.; Aoki, Y. Elongation method with cutoff technique for linear SCF scaling. *Int. J. Quantum Chem.* **2005**, *102*, 785–794.

4. Orimoto, Y.; Gu, F.L.; Imamura, A.; Aoki, Y. Efficient and accurate calculations on the electronic structure of B-type poly(dG)·poly(dC) DNA by elongation method: First step toward the understanding of the biological properties of aperiodic DNA. *J. Chem. Phys.* **2007**, *126*, 215104:1–215104:7.
5. Xie, P.; Liu, K.; Gu, F.L.; Aoki, Y. Counter-Ion Effects of A- and B-type Poly(dG)·Poly(dC)· and Poly(dA)·Poly(dT) DNA by Elongation Method. *Int. J. Quantum Chem.* **2012**, *112*, 230–239.
6. Aoki, Y.; Gu, F.L. An elongation method for large systems toward bio-systems. *Phys. Chem. Chem. Phys.* **2012**, *14*, 7640–7668.
7. Liu, K.; Inerbaev, T.; Korchowicz, J.; Gu, F.; Aoki, Y. Geometry optimization for large systems by the elongation method. *Theor. Chem. Acc.* **2012**, *131*, 1–8.
8. Branden, C.; Tooze, J. *Introduction to Protein Structure*, 2nd ed.; Garland Publishing: New York, NY, USA, 1999.
9. Dean, J.C.; Buchanan, E.G.; Zwier, T.S. Mixed 14/16 Helices in the Gas Phase: Conformation-Specific Spectroscopy of Z-(Gly)_n, n = 1, 3, 5. *J. Am. Chem. Soc.* **2012**, *134*, 17186–17201.
10. Ohnishi, S.; Kamikubo, H.; Onitsuka, M.; Kataoka, M.; Shortle, D. Conformational preference of polyglycine in solution to elongated structure. *J. Am. Chem. Soc.* **2006**, *128*, 16338–16344.
11. Edmiston, C.; Ruedenberg, K. Localized Atomic and Molecular Orbitals. *Rev. Mod. Phys.* **1963**, *35*, 457–464.
12. Boys, S.F. Construction of Some Molecular Orbitals to Be Approximately Invariant for Changes from One Molecule to Another. *Rev. Mod. Phys.* **1960**, *32*, 296–299.
13. Szabo, A.; Ostlund, N.S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications Inc.: New York, NY, USA, 1996.
14. Allen, M.P.; Tildesley, D.J. *Computer Simulation of Liquids*; Oxford University Press: Oxford, UK, 1987.
15. Sadus, R.J. *Molecular Simulation of Fluids: Theory, Algorithms and Object-Oriented*; Elsevier Science Inc.: New York, NY, USA, 1999.
16. Schmidt, M.W.; Baldrige, K.K.; Boatz, J.A.; Elbert, S.T.; Gordon, M.S.; Jensen, J.H.; Koseki, S.; Matsunaga, N.; Nguyen, K.A.; Su, S.J.; Windus, T.L.; Dupuis, M.; Montgomery, J.A. General Atomic and Molecular Electronic Structure System. *J. Comput. Chem.* **1993**, *14*, 1347–1363.
17. Minor, D.L.; Kim, P.S. Measurement of the β -sheet-forming propensities of amino acids. *Nature* **1994**, *367*, 660–663.
18. Smith, C.K.; Withka, J.M.; Regan, L. A Thermodynamic Scale for the β -Sheet Forming Tendencies of the Amino Acids. *Biochemistry* **1994**, *33*, 5510–5517.
19. Chou, P.Y.; Fasman, G.D. Conformational parameters for amino-acids in helical, β -sheet, and random coil regions calculated from proteins. *Biochemistry* **1974**, *13*, 211–222.
20. Kim, C.A.; Berg, J.M. Thermodynamic β -sheet propensities measured using a zinc-finger host peptide. *Nature* **1993**, *362*, 267–270.
21. Berendsen, H.J.C.; Postma, J.P.M.; Vangunsteren, W.F.; Dinola, A.; Haak, J.R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

22. Rules, T. Abbreviations and Symbols for the Description of the Conformation of Polypeptide Chains. *Eur. J. Biochem.* **1970**, *17*, 193–201.
23. Nagornova, N.S.; Rizzo, T.R.; Boyarkin, O.V. Interplay of Intra- and Intermolecular H-Bonding in a Progressively Solvated Macrocyclic Peptide. *Science* **2012**, *336*, 320–323.
24. Liu, K.; Peng, L.; Gu, F.L.; Aoki, Y. Three dimensional elongation method for large molecular calculations. *Chem. Phys. Lett.* **2013**, *560*, 66–70.

© 2013 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).