Theor Chem Acc (2012) 131:1277
DOI 10.1007/s00214-012-1277-9

REGULAR ARTICLE

Geometry optimization for large systems by the elongation method

**Kai Liu · Talgat Inerbaev · Jacek Korchowiec ·
Feng Long Gu · Yuriko Aoki**

Received: 28 July 2012 / Accepted: 6 September 2012 / Published online: 28 September 2012

Abstract Geometry optimization using the elongation method is developed at the Hartree–Fock level of theory. The formalism of elongation energy gradient and its accuracy have been validated by model systems calculations. The linear poly-hydrogen fluoride, polyethylene, planar polyacetylene and extended polyalanine are optimized using different basis sets and compared with conventional results. The results show that the elongation Hartree–Fock geometry optimization (ELG-HF-OPT) can reproduce conventional calculation results with high accuracy for various basis sets. For the poly-hydrogen fluoride calculation at 6-31G(d,p) basis set, moreover, ELG-HF-OPT gives a structure with lower ground state energy than conventional results with the same optimization convergence threshold. This means the potential possibility of ELG-HF-OPT can locate a more stable structure than conventional calculations with the same optimization convergence criteria. Therefore, the ELG-HF-OPT would provide one more choice for performing optimization on complicated large systems.

Keywords Geometry optimization · Elongation method · Cutoff technique · Large system

Electronic supplementary material The online version of this article (doi:10.1007/s00214-012-1277-9) contains supplementary material, which is available to authorized users.

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

T. Inerbaev · F. L. Gu (✉)
MOE Key Laboratory of Theoretical Chemistry of Environment,
School of Chemistry and Environment, South China Normal
University, Guangzhou 510631, China
e-mail: gu@scnu.edu.cn

J. Korchowiec
Department of Theoretical Chemistry, Faculty of Chemistry,
Jagiellonian University, Ingardena 3, Kraków 30-060, Poland

F. L. Gu · Y. Aoki
Japan Science and Technology Agency, CREST,
4-1-8 Hon-chou, Kawaguchi, Saitama 332-0012, Japan

Y. Aoki (✉)
Department of Material Sciences, Faculty of Engineering
Sciences, Kyushu University, 6-1 Kasuga-Park,
Fukuoka 816-8580, Japan
e-mail: aoki.yuriko.397@m.kyushu-u.ac.jp

1 Introduction

An accurate description of the electronic structure of a molecule is very important in molecular modeling by quantum chemistry. However, it is still challenging to calculate large systems, such as proteins and DNA, because of their high computational costs. For instance, Hartree–Fock (HF) [1] and density functional theory (DFT) [2] scale in form as $O(N^4)$, the coupled cluster (CC) method [3–6] including single and double excitations (CCSD) scales as $O(N^6)$, where N is the number of basis functions. To reduce this expense to a reasonable level, much considerable efforts have been made to linearly $O(N)$ calculate the Fock matrix (e.g. prescreening techniques [7], quantum fast multiple method (QFMM) [8–10]), which is one of the most time-consuming part of calculations [11–16]. On the other hand, fragmentation methods based on the near-sightedness principle [17] have been proposed and developed rapidly. To name just a few, divide and conquer [18–21], fragment molecular orbitals [22–29], systematic fragmentation method [30–33] and elongation method

(ELG) [34–36]. These approaches adopt local molecular orbitals (LMOs) of essential fragments or subunits instead of conventional canonical molecular orbitals (CMOs) of the whole system. In other words, these approaches project the Roothaan–Hall equations [37] (or Kohn–Sham equations [38]) into a subspace associated with molecular fragments instead of solving HF/KS for the whole system. In such a way, $O(N^3)$ diagonalization problem of a huge matrix is omitted.

The ELG method, originally proposed by Imamura et al. in the early of 1990s, uses region localized molecular orbitals (RLMOs) to calculate random quasi-one-dimension periodic and aperiodic systems with high accuracy at a reasonable cost level. Currently, this method has been successfully implemented and linked to the GAMESS software package [39]. It has been applied to molecular properties calculation at HF and DFT [40] levels of theory. For example, band structure [41, 42] and nonlinear optical properties of polymers [43–47] have been investigated using ELG-HF or ELG-KS method. These properties depend on the system geometry. It is important to link geometry optimization with ELG method. In this paper, the performance of ELG-HF method armed with geometry optimization (ELG-HF-OPT) is described. Our efforts are here concentrated on the formalism of ELG-HF-OPT method and its accuracy in comparison with conventional HF optimization scheme.

The paper is organized as follows. After a brief introduction of the ELG method, a description of the ELG geometry optimization is presented. Next, the results for three model systems (linear poly-hydrogen fluoride, polyethylene and planar polyacetylene) are given with the special emphasize put on basis set effect [STO-3G, 6-31G and 6-31G(d,p)]. The optimized geometrical structures of model systems are compared to conventional HF method. The analysis is focused on Z-matrix variables, that is, bond lengths, bond angles and total energies. Additional test calculations are performed for a model built from 20 units of extended polyalanine. Finally, conclusions are given with future prospects.

2 Elongation method

In general, ELG procedure is analogous to experimental polymer chain synthesis as shown in Fig. 1. More detailed descriptions are found in [36] and [48]. First, a suitable size of initial monomers (starting cluster) is chosen to initiate the ELG procedure. The energy of this starting cluster is computed by conventional HF self-consistent-field (HF-SCF) procedure; then, the CMOs of starting cluster are transformed to an orthogonal atomic basis (OAO) by Löwdin's symmetric orthogonalization [49]. In the following step, the OAO-based density matrix ($\mathbf{D}^{\mathrm{OAO}}$) is to partition into frozen region (A) and active region (B). The frozen region is assumed to be far away from the chain propagation point, while the active region is consisted of the remaining part of the starting cluster. After that, the transformations to the regional orbitals are obtained by separately diagonalizing the subspaces $\mathbf{D}^{\mathrm{OAO}}$ (A) and $\mathbf{D}^{\mathrm{OAO}}$ (B). Due to the unwanted mixing between occupied and unoccupied orbitals, the full density matrix in regional orbital basis is block-diagonalized by Jacobi procedure with blocks corresponding to occupied and unoccupied subspaces. After this localization procedure finishes, a set of RLMOs for both A and B regions are obtained. This ELG localization scheme has been successfully performed on several model systems including strongly delocalized system like cationic cyanines [36].

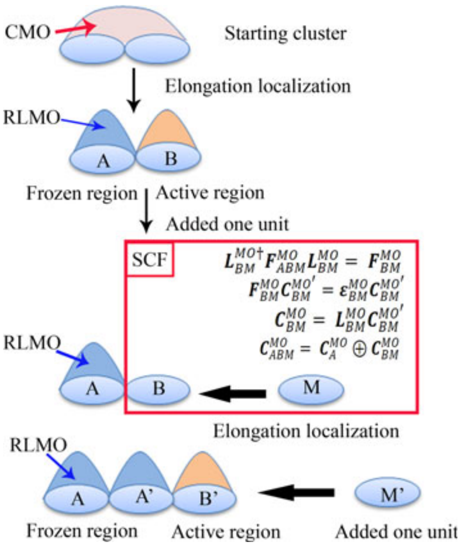

Fig. 1 The flowchart of the elongation method, CMO and RLMO correspond to the canonical molecular orbital and regional localized molecular orbital, respectively. After the elongation localization procedure, CMOs of staring cluster will be localized to RLMO A (frozen part) and RLMO B (active part). The elongation HF equations will be solved at local molecular orbital basis

Then, one attacking unit (M) is added to the chain propagation point for the next ELG step. Because the frozen region (A) is far away from M, the exchange interactions between them are neglected. In the subsequent ELG single point energy calculation, only B and M will be included in the ELG HF-SCF calculations, and the Roothaan–Hall equations are solved only for molecular orbitals (MOs) of B and M regions.

After the ELG HF-SCF converges, the CMOs of B and M regions will be localized again to form a new frozen region (RLMOs A′) and a new active region (RLMOs B′). Then, a new attacking unit (M′) is added to repeat the above procedures until the desired length is reached.

The important feature of the ELG method is that the HF/KS equations are solved only for small subunits instead of the whole system. If the same size of unit (M) is added each time, the corresponding HF equations will be solved with the same dimension. Meanwhile, if the distance (or interaction) between A and M is less than threshold, the calculations of two-electron repulsion integrals (ERIs) will be partly omitted by cutoff technique [48, 50–52]. It is proved by the large model system calculations that the ELG method can evaluate the total electronic energy of system linearly (or sublinearly) while maintaining a high accuracy.

3 Elongation geometry optimization

In the framework of HF calculation, the first derivative (gradient) of total energy (E) with respect to the nuclear coordinate X_{A} at atomic orbitals basis can be written as follows[1]:

$$\frac{\partial E}{\partial X_{\mathrm{A}}} = \sum_{\mu\nu} \mathbf{D}_{\nu\mu} \frac{\partial \mathbf{H}^{\mathrm{core}}_{\mu\nu}}{\partial X_{\mathrm{A}}} + \frac{1}{2} \sum_{\mu\nu\lambda\sigma} \mathbf{D}_{\nu\mu} \mathbf{D}_{\lambda\sigma} \frac{\partial (\mu\nu||\sigma\lambda)}{\partial X_{\mathrm{A}}} + \frac{\partial \mathbf{V}_{\mathrm{NN}}}{\partial X_{\mathrm{A}}} - \sum_{\mu\nu} \mathbf{Q}_{\nu\mu} \frac{\partial S_{\mu\nu}}{\partial X_{\mathrm{A}}} \tag{1}$$

where $(\mu\nu\|\sigma\lambda)$ denotes two-electron integral, $\mathbf{D}$ is the density matrix, $\mathbf{H}^{\mathrm{core}}$ and $\mathbf{S}$ correspond to the core Hamiltonian and overlap matrices of system, respectively. The nuclear–nuclear repulsion is defined by $\mathbf{V}_{\mathrm{nn}}$. The energy-weighted density matrix $\mathbf{Q}$ of conventional method is defined as follows:

$$\mathbf{Q}_{\nu\mu} = \sum_{i}^{N/2} n_i \varepsilon_i C_{\mu i} C_{\nu i} \tag{2}$$

where $\mathbf{n}$ is the matrix of occupancy number (density matrix in MO representation).

In the ELG method, the HF equations are solved only for active subspaces defined by B and M fragments. After the ELG SCF procedure has converged, all matrices and the molecular orbitals will be transferred back to atomic orbital basis for the gradient calculated by Eq. (1). Comparing with the conventional method, only the last term in Eq. (1) has changed. The eigenvalue matrix $\varepsilon_{\mathrm{ABM}}$ (subscript ABM means the whole system) in Eq. (2) is no longer a diagonal matrix. The non-diagonal ε_{ij}, employed in the ELG-HF-OPT method, is obtained by ${C^{\mathrm{MO}}_{\mu i}}^{\dagger} F^{\mathrm{MO}}_{\mu\nu} C^{\mathrm{MO}}_{\nu j}$ and based on the orthonormal overlap in the mixed basis. The coefficients of the whole system $C^{\mathrm{MO}}_{\mathrm{ABM}}$ consist of $C^{\mathrm{LMO}}_{\mathrm{A}}$ (coefficients of frozen part, localized by the ELG localization procedure) and $C^{\mathrm{MO}}_{\mathrm{BM}}$ (coefficients of active region and attacking unit, transformed from $C^{\mathrm{MO}'}_{\mathrm{BM}}$ after the ELG SCF calculation). Finally, the energy-weighted density matrix $\mathbf{Q}$ of the ELG-HF-OPT method in atomic basis can be rewritten as follows:

$$\mathbf{Q}_{\nu\mu} = \sum_{ij}^{N/2} n_i \varepsilon_{ij} C_{\mu i} C_{\nu j} \tag{3}$$

The ELG-HF-OPT procedure goes as it is described below. The starting cluster is initially optimized using conventional optimization algorithm. After the equilibrium geometry of starting cluster is reached, the electronic orbitals are localized by the ELG localization procedure. Then, an attacking unit is attached. Now, the structure is divided into A, B and M regions. A region is frozen, that is, both electronic orbitals and atomic coordinates are not changed any more. The geometry of remaining BM region is optimized with the ELG-HF-OPT algorithm. Because of the boundary effects between A and BM arisen by the tails after the ELG localization procedure, the coordinates of the one unit of the BM region, which is the closest to the frozen region, are fixed in the gradient calculation to reduce these effects [53–55].

4 Computational details

Both elongation and conventional geometry optimizations were performed by the default geometry optimizer implemented in the GAMESS program package. The optimization convergence threshold is set to be 1×10^{-4} Hartree/Bohr, which means the maximum gradient with respect to coordinate should be less than 1×10^{-4} Hartree/Bohr, and the root mean square gradient (RMSD) should be smaller than $1/3 \times 10^{-4}$ Hartree/Bohr. To illustrate potential uses for elongation geometry optimization, three different types of model systems are chosen. Linear poly-hydrogen fluoride is a model system for non-bonded systems, polyethylene is employed as example of bonding systems, and planar polyacetylene is treated as a system with strongly delocalized π-electrons. STO-3G, 6-31G and 6-31G(d,p) basis sets are used. For the purpose of this paper which is to valid the formalism of ELG-HF-OPT method and its accuracy, the cutoff threshold is restricted to a very low value (the value of overlap between A and M should be less than 1.0×10^{-5}, and the coupling between frozen (A) and active units (B), $X_n = \sum_{\mu\nu \in B} \sum_{ij \in A} \left| C_{\mu i} S_{ij} C_{j\nu} \right|$, should be less than 1.0×10^{-9}) [50]. Finally, a model with 20 units of extended polyalanine is optimized as an example of the application of developed approach to a complicated bio-system using 6-31G basis set. The initial cluster size is chosen systematically. In the ELG-HF-OPT, one unit, the closest to the frozen region, will be fixed in the gradient calculation because of the coupling between frozen region and active region. Then, the size of the starting cluster should be greater than 3 units. In general, larger starting clusters give higher accuracy in elongation calculations. If the size of starting unit is chosen to be small, the program can automatically increase the initial cluster size until the interaction between frozen units and the terminal is less than a given value. The model system of polyethylene and polyacetylene starts from 4 units: one unit is frozen region and the other 3 units are active region. Consider the potential hydrogen bond between units, poly-(HF)s and the polyalanine system starts with 5 units. These sizes of initial cluster should guarantee the frozen region far away from the terminal. The unit size is chosen by trial and error. Smaller unit size will make the SCF solved in a lower dimension but with more elongation steps. Too many elongation steps may offset the time savings in the reduced size of the SCF calculation. In this work, the unit size is not optimized by trial and error; we just simply choose the number of atoms in each unit is about 20 based on previous experience.

5 Result and discussion

5.1 Non-bonding model system: $(HF)_{n=48}$

The linear poly-hydrogen fluoride (poly-HF) molecules are optimized by both elongation and conventional restricted Hartree–Fock method (RHF) with different basis sets. The initial geometry is arranged from head to tail (H–F…H–F…H–F…H–F…), and the initial structure parameters are set as a kind of random structure, $R_{\mathrm{H-F}} = 0.90 \pm 0.20$ Å, $R_{\mathrm{H\ldots F}} = 1.50 \pm 0.20$ Å. Details of the initial bond lengths for the first 4 molecules are shown in Fig. 2a. For the elongation geometry optimization, the starting cluster contains 30 HF molecules with 6 (one unit) in frozen region (A) and the other 24 (4 units) in active region (B). The large B region will guarantee that the interaction between frozen region(s) and attacking monomer is negligible in the next elongation steps. After a total of 4 steps of optimization, the equilibrium structure by the elongation geometry optimization is achieved. The bond lengths of intra-molecules ($R_{\mathrm{H-F}}$) and inter-molecules ($R_{\mathrm{F\ldots H}}$) for STO-3G, 6-31G, and 6-31G(d,p) basis sets are listed in Supplementary Material Table 1, and the differences of the bond lengths ($\Delta R_{\mathrm{H-F}}$ and $\Delta R_{\mathrm{H\ldots F}}$, $\Delta R = R^{\mathrm{elongation}} - R^{\mathrm{conventional}}$) are shown in Supplementary Material Fig. 2(b, c). It shows that from STO-3G to 6-31G(d,p) basis set, the optimized bond lengths ($R_{\mathrm{H-F}}$) of intra-molecules decrease, while those of inter-molecules increase. The maximum differences of intra-molecular bond lengths are 7.4×10^{-4} Å for STO-3G, 1.2×10^{-4} Å for 6-31G and -1.4×10^{-3} Å for 6-31G(d,p), while the maximum differences of inter-molecular bond lengths are -6.4×10^{-3} Å, 1.5×10^{-3} Å and 6.8×10^{-2} Å for STO-3G and 6-31G and 6-31G(d,p) basis sets, respectively. It also shows that the 6-31G basis set gives the best agreement to the corresponding conventional result. However, The energy differences ΔE ($\Delta E = E^{\mathrm{elongation}} - E^{\mathrm{conventional}}$) of STO-3G, 6-31G and 6-31G(d,p) basis sets are -4.49×10^{-7}, 9.09×10^{-9} and -3.26×10^{-5} Hartree/atom, respectively. As we see that the total energy difference of 6-31G(d,p) is lower than corresponding conventional one. This means the ELG-OPT locates an even lower ground state than conventional results when the same convergence criteria are employed. Based on the structure optimized by the ELG-HF-OPT at 6-31G(d,p) basis set, a conventional single point energy calculation is performed. This total energy difference between ELG-HF-OPT energy and conventional SCF single point calculated ($E^{\mathrm{elongation}} - E^{\mathrm{conventional}}$) is 5.41×10^{-11} Hartree/atom. For a flat energy potential surface of linear poly-HF molecules, ELG-HF-OPT may produce a more promising candidate for the most stable geometry.

The most time-consuming step of optimization is multiple times SCF calculations. As described above, the ELG will solve the SCF equation efficiently with small dimension. If the cutoff is activated, meanwhile, it means if three or four indexes of ERIs belong to cutoff unit(s), the corresponding ERIs will automatically be skipped to calculate. Therefore, the time to form the whole Fock matrix will be faster. Despite the very low cutoff threshold is chosen, after the accumulation of elongation steps, the efficiency of ELG-OPT (sum of all elongation steps about 120.2 min) is little faster (about 8 %) than convention (a one-time calculation for whole molecule about 131.7 min) at the 6-31G(d,p) basis set for the whole system calculation.

By analyzing differences in the total energies and bond lengths, we see that the elongation geometry optimization method produces an optimum structure similar to the structure obtained from conventional Hartree–Fock calculations. It should be mentioned that the maximum errors of bond distances of a sets of model molecules of Hartree–Fock method by cc-pVQZ relative to experiment are about 8.5×10^{-2} Å [56]. Comparing to this error, the difference

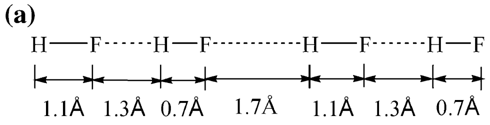

Fig. 2 **a** The structure parameters of the first four hydrogen fluoride molecules. $R_{\mathrm{H-F}} = 0.90 \pm 0.20$ Å, $R_{\mathrm{H\ldots F}} = 1.50 \pm 0.20$ Å

between ELG-HF-OPT and conventional Hartree–Fock method is negligible. For the result of 6-31G(d,p) basis set, ELG-HF-OPT may search over more possibilities than conventional method with the same optimization convergence threshold.

5.2 Bonding system: polyethylene ($C_{66}H_{134}$)

The geometry for polyethylene chain was optimized by ELG using a starting cluster containing 4 units with one in frozen part, and the remaining three units in active part, where $–[CH_2]_8–$ group is one unit and added for each elongation step, except $–[CH_2]_{10}–H$ is added for the final step. The initial structural information is shown in Fig. 3a. For the ELG-HF-OPT calculation, there was no symmetry restrictions used during the optimization. In each ELG-HF-OPT step, except final one, a hydrogen atom was used to cap the end of the chain to make a closed-shell system. After the equilibrium structure was reached, the terminal atom (i.e. hydrogen atom) was removed for the next ELG-HF-OPT step. A total of 5 ELG-HF-OPT steps were performed to reach the equilibrium geometry of the whole polyethylene chain. The differences between bond lengths and bond angles calculated with the ELG-HF-OPT and conventional methods are shown in Supplementary Material Fig. 3(b–g). The maximum differences of bond length and bond angle are 5.2×10^{-5} Å and 0.02° for STO-3G basis set, 5.8×10^{-5} Å and 0.03° for 6-31G basis set, and 6.4×10^{-5} Å and 0.02° for 6-31G(d,p) basis set, respectively. These small differences in structure predict the little errors in the total energy. Compared to conventional results, the differences of total energy (ΔE) are 2.99×10^{-9} Hartree/atom for STO-3G basis set, 2.53×10^{-9} Hartree/atom for 6-31G basis set and 1.78×10^{-9} Hartree/atom for 6-31G(d,p) basis set. Differences in bond length, bond angle and the total energy indicate that the reproduction of conventional HF results by ELG-HF-OPT procedure is really very good.

Here, we are focusing on the formalism of ELG-HF-OPT method and its accuracy; therefore, a strict threshold is chosen for the cutoff procedure. In other words, for the bonding system calculation, only a few unit(s) will be cutoff in current calculations to keep the high accuracy. For instance, only one unit is cutoff at the final step with 6-31G(d,p) basis set. Considering the accumulation of elongation steps, the current optimization efficiency of bonding system is not impressed. However, the differences of bond lengths, bond angles and the total energies indicate that the ELG-OPT can reproduce well the conventional results.

5.3 Delocalized system: planar polyacetylene ($C_{64}H_{66}$)

Planar polyacetylene is also optimized by the ELG-HF-OPT to test of a system with delocalized π-elections. The initial bond lengths of double bond and single bond between carbon and carbon are set to 1.30 and 1.45 Å, respectively. More details about the initial geometry parameters are listed in Fig. 4a. For the first elongation geometry optimization step, 4 units ($–[CH{=}CH]_4–$ group is one unit) are optimized; then, each time, one unit is added for each elongation step. In a total of 5 steps, the equilibrant structure is located. The differences of all bond length and bond angle between the elongation results and conventional calculations are shown in Supplementary Material Fig. 4(b–g). The maximum differences in bond length and bond angle between ELG-HF-OPT and conventional Hartree–Fock method are 4.9×10^{-5} Å and 0.02° for STO-3G basis set, 2.6×10^{-5} Å and 0.02° for 6-31G basis, and 3.2×10^{-5} Å and 0.01° for 6-31G(d,p) basis set. An important thing is that the differences in bond length and bond angle are independent with basis sets. As the basis set enlarged from STO-3G to 6-31G(d,p) basis set, the carbon single bond lengths decrease while double bond lengths increase, details are listed in Supplementary Material Table 2. These bond length changes attribute to the effect of delocalized π-orbitals. The same tendency in carbon bond length changes between ELG-OPT and conventional results indicates that although the existence of the elongation localization procedure, the ELG-OPT can well represent the delocalized π-orbitals.

Compared to conventional calculations, the efficiency of current ELG-OPT of polyacetylene is not attractive. This

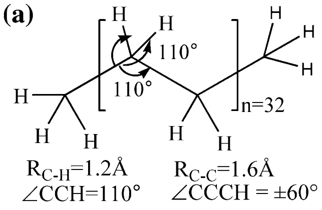

Fig. 3 **a** Initial structural parameters of polyethylene, initial bond lengths of C–H and C–C are set to 1.2 and 1.6 Å, respectively. The bond angle of C–C–H is set to 110°, and the dihedral of C–C–C–H is ±60° of the carbon atomic plane

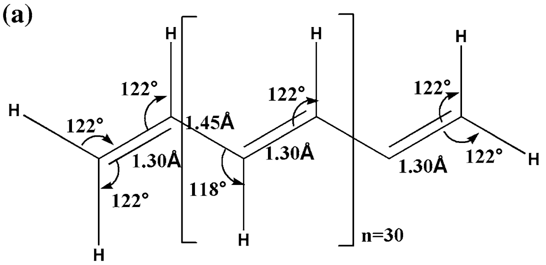

Fig. 4 **a** The geometrical parameters of the initial structure polyacetylene

should attribute to two facts: the symmetry during optimization and the delocalized orbitals. It is difficult to fully localize those π-orbitals into regional localized units because of the "tails" generated during the orthogonal treatment. Due to these few but weighted component elements in tails, no cutoff is activated at current threshold during the whole ELG-OPT. Fortunately, an "orbital shifted" treatment has recently been proposed and developed to deal with these tails during the elongation localization especially for delocalized system [47]. But this treatment for delocalized π-orbitals is not employed in this paper.

The total energy difference between the elongation and conventional calculations of planar polyacetylene of STO-3G, 6-31G and 6-31G(d,p) is 1.14×10^{-9}, 1.90×10^{-9} and 1.06×10^{-9} Hartree/atom, respectively. The difference of total energies and structures indicate that the ELG-OPT will represent the conventional results.

5.4 Polyalanine

Finally, a model of 20 units of extended alanines is optimized by both the ELG-HF-OPT and conventional method using 6-31G basis set. Two alanines are one unit. The initial geometry of the extended polyalanine is generated by Spartan [57] with Psi (155.0°) and Phi (−157.0°). The starting cluster consists of 5 units with one unit in the frozen region, 4 units in active region. Due to the large molecular size, the QFMM is employed by both conventional and elongation method. The optimized geometry difference between the ELG-HF-OPT and conventional SCF-OPT is compared. The maximum differences in bond length and bond angle are 3.4×10^{-4} Å and 0.18°, respectively. The RMSD of the whole backbone (excluding hydrogen) shown in Supplementary Material Fig. 5 is 0.18, calculated by VMD program package [58]. The difference of total energy between ELG-OPT and conventional result is 2.37×10^{-7} Hartree/atom. These small differences in structure and total energy indicate the good agreement between ELG-OPT and conventional calculations.

6 Conclusions and prospects

The geometry of the four model systems (linear poly-hydrogen fluoride, polyethylene, planar polyacetylene and polyalanine) is optimized using the ELG-HF-OPT procedure. Compared to conventional calculations, it is obvious that the overall agreement is very good: in terms of the total energy differences, at worst about 10^{-7} Hartree/atom order of magnitude, for bond length 10^{-3} Å and for bond angles 0.2°.

The result of poly-hydrogen fluoride at 6-31G(d,p) basis set is very promising. Despite of searching on the flat energy potential surface, the total energy of the equilibrium structure located by the ELG-HF-OPT is lower than the corresponding conventional one with the same optimization convergence threshold. The ELG-OPT method focuses not only on the whole structure, but also on each elongation geometry optimization step. It may search over more possibilities for local minimum, which may produce a more promising candidate for the most stable geometry. Actually, we also applied to some structures of DNA and proteins and found such phenomenon that ELG-OPT method could search more stable structure than those from conventional direct optimization for entire system with the same convergence criteria. It must be useful to search more optimized structures toward global minimum in complicated biosystems. We would report those data in the next step after systematical and careful investigations are accomplished.

In this paper, we have examined the formalism of ELG-HF-OPT method and its accuracy by model systems calculations. At the same time, these calculations also predict our future works. It is not only to set a rough threshold to achieve an acceptable result, but also to introduce more approximations to speed up current calculations. For example, the possibility of employing point charges approximation [59]. The elongation method, on the other hand, is being developed to deal with more complicate three-dimension real system. For details of the elongation method and its applications, please refer to the perspective report of [60]. Our final goal is to optimize the real and complicate large system by the elongation geometry optimization method.

Acknowledgments K. L. was supported by the JST-RA system by the group CREST, Japan Science and Technology Agency (JST) and the Kuma Toshimi international scholarship of Interdisciplinary Graduate School of Engineering Sciences at Kyushu University. K. L. thanks D. R. Price for assistance in revising in the manuscript. The calculations were performed on the Linux PC cluster in our laboratory and the high performance computing systems of the Research Institute for Information Technology at Kyushu University. T. I. was financially supported by China Postdoctoral Science Foundation.

References

1. Szabo A, Ostlund NS (1996) Modern quantum chemistry—introduction to advanced electronic structure theory. Dover Publications lnc, New York
2. Parr RG, Yang WT (1995) Density-functional theory of the electronic-structure of molecules. Annu Rev Phys Chem 46:701–728
3. Bartlett RJ (1991) Coupled-cluster theory in atomic physics and quantum-chemistry—introduction and overview. Theor Chim Acta 80:71–79
4. Bartlett RJ (2000) Perspective on "On the correlation problem in atomic and molecular systems. Calculation of wavefunction components in Ursell-type expansion using quantum-field theoretical methods"—Cizek J (1966) J Chem Phys 45: 4256. Theor Chem Acc 103:273–275

5. Bartlett RJ, Musial M (2007) Coupled-cluster theory in quantum chemistry. Rev Mod Phys 79:291–352
6. Cizek J (1966) On the correlation problem in atomic and molecular systems. Calculation of wavefunction components in Ursell-type expansion using quantum-field theoretical methods. J Chem Phys 45:4256–4266
7. Haser M, Ahlrichs R, Baron HP, Weis P, Horn H (1992) Direct computation of 2nd-order SCF properties of large molecules on workstation computers with an application to large carbon clusters. Theor Chim Acta 83:455–470
8. Choi CH, Ivanic J, Gordon MS, Ruedenberg K (1999) Rapid and stable determination of rotation matrices between spherical harmonics by direct recursion. J Chem Phys 111:8825–8831
9. Choi CH, Ruedenberg K, Gordon MS (2001) New parallel optimal-parameter fast multipole method (OPFMM). J Comput Chem 22:1484–1501
10. Choi CH (2004) Direct determination of multipole moments of Cartesian Gaussian functions in spherical polar coordinates. J Chem Phys 120:3535–3543
11. Burant JC, Scuseria GE, Frisch MJ (1996) A linear scaling method for Hartree-Fock exchange calculations of large molecules. J Chem Phys 105:8969–8972
12. Schwegler E, Challacombe M (1996) Linear scaling computation of the Hartree-Fock exchange matrix. J Chem Phys 105:2726–2734
13. Challacombe M, Schwegler E (1997) Linear scaling computation of the Fock matrix. J Chem Phys 106:5526–5536
14. Ochsenfeld C, White CA, Head-Gordon M (1998) Linear and sublinear scaling formation of Hartree-Fock-type exchange matrices. J Chem Phys 109:1663–1669
15. Schwegler E, Challacombe M (1999) Linear scaling computation of the Fock matrix. IV. Multipole accelerated formation of the exchange matrix. J Chem Phys 111:6223–6229
16. Gan CK, Tymczak CJ, Challacombe M (2004) Linear scaling computation of the Fock matrix. VII. Parallel computation of the Coulomb matrix. J Chem Phys 121:6608–6614
17. Kohn W (1996) Density functional and density matrix method scaling linearly with the number of atoms. Phys Rev Lett 76:3168–3171
18. Yang W (1991) Direct calculation of electron density in density-functional theory. Phys Rev Lett 66:1438–1441
19. Yang WT, Lee TS (1995) A density-matrix divide-and-conquer approach for electronic-structure calculations of large molecules. J Chem Phys 103:5674–5678
20. Akama T, Kobayashi M, Nakai H (2007) Implementation of divide-and-conquer method including Hartree-Fock exchange interaction. J Comput Chem 28:2003–2012
21. Kobayashi M, Imamura Y, Nakai H (2007) Alternative linear-scaling methodology for the second-order Moller-Plesset perturbation calculation based on the divide-and-conquer method. J Chem Phys 127:074103
22. Kitaura K, Ikeo E, Asada T, Nakano T, Uebayasi M (1999) Fragment molecular orbital method: an approximate computational method for large molecules. Chem Phys Lett 313:701–706
23. Fedorov DG, Kitaura K (2004) The importance of three-body terms in the fragment molecular orbital method. J Chem Phys 120:6832–6840
24. Fedorov DG, Kitaura K (2004) On the accuracy of the 3-body fragment molecular orbital method (FMO) applied to density functional theory. Chem Phys Lett 389:129–134
25. Fedorov DG, Kitaura K (2007) Extending the power of quantum chemistry to large systems with the fragment molecular orbital method. J Phys Chem A 111:6904–6914
26. Mochizuki Y, Koikegami S, Nakano T, Amari S, Kitaura K (2004) Large scale MP2 calculations with fragment molecular orbital scheme. Chem Phys Lett 396:473–479
27. Mochizuki Y, Fukuzawa K, Kato A, Tanaka S, Kitaura K, Nakano T (2005) 1A configuration analysis for fragment interaction. Chem Phys Lett 410:247–253
28. Mochizuki Y, Ishikawa T, Tanaka K, Tokiwa H, Nakano T, Tanaka S (2006) Dynamic polarizability calculation with fragment molecular orbital scheme. Chem Phys Lett 418:418–422
29. Mochizuki Y, Yamashita K, Murase T, Nakano T, Fukuzawa K, Takematsu K, Watanabe H, Tanaka S (2008) Large scale FMO-MP2 calculations on a massively parallel-vector computer. Chem Phys Lett 457:396–403
30. Deev V, Collins MA (2005) Approximate ab initio energies by systematic molecular fragmentation. J Chem Phys 122:154102
31. Collins MA, Deev VA (2006) Accuracy and efficiency of electronic energies from systematic molecular fragmentation. J Chem Phys 125:104104
32. Addicoat MA, Collins MA (2009) Accurate treatment of nonbonded interactions within systematic molecular fragmentation. J Chem Phys 131:104103
33. Mullin JM, Roskop LB, Pruitt SR, Collins MA, Gordon MS (2009) Systematic fragmentation method and the effective fragment potential: an efficient method for capturing molecular energies. J Phys Chem A 113:10040–10049
34. Imamura A, Aoki Y, Maekawa K (1991) A theoretical synthesis of polymers by using uniform localization of molecular-orbitals—proposal of an elongation method. J Chem Phys 95: 5419–5431
35. Aoki Y, Imamura A (1992) Local density of states of aperiodic polymers using the localized orbitals from an abinitio elongation method. J Chem Phys 97:8432–8440
36. Gu FL, Aoki Y, Korchowiec J, Imamura A, Kirtman B (2004) A new localization scheme for the elongation method. J Chem Phys 121:10385–10391
37. Roothaan CCJ (1951) New developments in molecular orbital theory. Rev Mod Phys 23:69–89
38. Kohn W, Sham LJ (1965) Self-consistent equations including exchange and correlation effects. Phys Rev 140:1133–1138
39. Schmidt MW, Baldridge KK, Boatz JA, Elbert ST, Gordon MS, Jensen JH, Koseki S, Matsunaga N, Nguyen KA, Su SJ, Windus TL, Dupuis M, Montgomery JA (1993) General atomic and molecular electronic-structure system. J Comput Chem 14:1347–1363
40. Aoki Y, Suhai S, Imamura A (1994) An efficient cluster elongation method in density-functional theory and its application to poly-hydrogen-bonding molecules. J Chem Phys 101:10808–10823
41. Pomogaeva A, Kirtman B, Gu FL, Aoki Y (2008) Band structure built from oligomer calculations. J Chem Phys 128:074109
42. Pomogaeva A, Springborg M, Kirtman B, Gu FL, Aoki Y (2009) Band structures built by the elongation method. J Chem Phys 130:194106
43. Ohnishi S, Orimoto Y, Gu FL, Aoki Y (2007) Nonlinear optical properties of polydiacetylene with donor-acceptor substitution block. J Chem Phys 127:084702
44. Chen W, Yu GT, Gu FL, Aoki Y (2009) Investigation on the electronic structures and nonlinear optical properties of pristine boron nitride and boron nitride-carbon heterostructured single-wall nanotubes by the elongation method. J Phys Chem C 113:8447–8454
45. Yan LK, Pomogaeva A, Gu FL, Aoki Y (2010) Theoretical study on nonlinear optical properties of metalloporphyrin using elongation method. Theor Chem Acc 125:511–520
46. Pomogaeva A, Gu FL, Imamura A, Aoki Y (2010) Electronic structures and nonlinear optical properties of supramolecular associations of benzo-2,1,3-chalcogendiazoles by the elongation method. Theor Chem Acc 125:453–460
47. Aoki Y, Loboda O, Liu K, Makowski M, Gu F (2011) Highly accurate O(N) method for delocalized systems. Theor Chem Acc 130:595–608

48. Makowski M, Korchowiec J, Gu FL, Aoki Y (2006) Efficiency and accuracy of the elongation method as applied to the electronic structures of large systems. J Comput Chem 27:1603–1619
49. Lowdin P-O (1950) On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals. J Chem Phys 18:365–375
50. Korchowiec J, Gu FL, Imamura A, Kirtman B, Aoki Y (2005) Elongation method with cutoff technique for linear SCF scaling. Int J Quantum Chem 102:785–794
51. Korchowiec J, Lewandowski J, Makowski M, Gu FL, Aoki Y (2009) Elongation cutoff technique armed with quantum fast multipole method for linear scaling. J Comput Chem 30:2515–2525
52. Korchowiec J, De Silva P, Makowski M, Gu FL, Aoki Y (2010) Elongation cutoff technique at Kohn–Sham level of theory. Int J Quantum Chem 110:2130–2139
53. Vreven T, Morokuma K, Farkas O, Schlegel HB, Frisch MJ (2003) Geometry optimization with QM/MM, ONIOM, and other combined methods. I. Microiterations and constraints. J Comput Chem 24:760–769
54. Wang B, Truhlar DG (2010) Combined quantum mechanical and molecular mechanical methods for calculating potential energy surfaces: tuned and balanced redistributed-charge algorithm. J Chem Theory Comput 6:359–369
55. Wang B, Truhlar DG (2011) Geometry optimization using tuned and balanced redistributed charge schemes for combined quantum mechanical and molecular mechanical calculations. Phys Chem Chem Phys 13:10556–10564
56. Helgaker T, Jorgensen P, Olsen J (2000) Molecular electronic-structure theory. John Wiley & Sons Ltd., West Sussex
57. Spartan’10 Wavefunction. Inc. Irvine, CA
58. Humphrey W, Dalke A, Schulten K (1996) VMD: visual molecular dynamics. J Mol Graph 14:33–38
59. Nagata T, Fedorov DG, Kitaura K (2009) Derivatives of the approximated electrostatic potentials in the fragment molecular orbital method. Chem Phys Lett 475:124–131
60. Aoki Y, Gu FL (2012) An elongation method for large systems toward bio-systems. Phys Chem Chem Phys 14:7640–7668