Chemical Physics Letters 560 (2013) 66–70

Contents lists available at SciVerse ScienceDirect

Chemical Physics Letters

journal homepage: www.elsevier.com/locate/cplett

Three dimensional elongation method for large molecular calculations

Kai Liu [a], Liang Peng [b], Feng Long Gu [b,d], Yuriko Aoki [c,d,*]

[a] Department of Molecular and Material Sciences, Interdisciplinary Graduate School of Engineering Sciences, Kyushu University, 6-1 Kasuga-Park, Fukuoka 816-8580, Japan
[b] MOE Key Laboratory of Theoretical Chemistry of Environment, School of Chemistry and Environment, South China Normal University, Guangzhou 510631, China
[c] Department of Material Sciences, Faculty of Engineering Sciences, Kyushu University, 6-1 Kasuga-Park, Fukuoka 816-8580, Japan
[d] Japan Science and Technology Agency, CREST, 4-1-8 Hon-chou, Kawaguchi, Saitama 332-0012, Japan

ARTICLE INFO

Article history:
Received 8 November 2012
In final form 21 December 2012
Available online 21 January 2013

ABSTRACT

Three dimensional elongation (3D-ELG) method is introduced to perform the calculations on real large systems based on regionally localized molecular orbitals. The settings are firstly justified by a quasi-one dimensional system calculations at HF/STO-3G basis set. Then two model systems are employed and calculated to validate the accuracy of 3D-ELG method at HF/6-31G basis set. The average deviation of the total energy between elongation and conventional calculations is 4.7×10^{-7} au/atom for these two model systems. This small deviation indicates the feasibility of 3D-ELG scheme for the large molecular calculations. Finally, the further development of 3D-ELG method is pointed out.

1. Introduction

The challenge to quantum chemist nowadays is to perform accurate and practical calculations for the large molecules. Although there is a very rapid development in the computer science, it is still not sufficient to carry out the *ab initio* calculations for large molecules. If N denotes the number of basis functions in the calculation, then the computer time to calculate the electronic energy is formally of order N^4 for Hartree–Fock (HF) method [1] and density functional theory (DFT) [2], of order N^5 for MP2 [3] and of order N^6 for CCSD [4–6]. Although technique like prescreening [7] can reduce these scaling factors, it still limits the size of the targets that can be treated in practice.

Recently, a popular approach to applying quantum mechanical calculation to large molecules is the hybrid quantum mechanics/molecular mechanics (QM/MM) approach [8,9] and ONIOM method [10,11]. In these two approaches, one uses force fields for the larger part of the system such as receptors or solvent molecules while employing the more expensive QM calculations only for the smaller subsystem such as ligands or solute molecules. Many applications to investigate the catalytic mechanism and transition state structures of enzyme reactions using QM/MM and ONIOM methods have been reported [12–15].

Another approach, based on the nearest fragment method [16], is to divide the system into fragments (subsystems) and to obtain the total properties by the calculations of these fragments individually. To name just a few, divide and conquer (DC) [17,18], fragment molecular orbital (FMO) [19,20], systematical molecular fragmentation (SMF) [21–23], molecular fractionation with conjugate caps (MFCC) approach [24,25] and the multireference configuration interaction method (MRCI) [26–29]. By the calculations of individual fragments, these approaches can linearly scale with the size of the system.

The elongation (ELG) method [30–32], originally proposed by Imamura et al. for the quasi-one dimensional periodic and aperiodic polymers, is also to divide the system into fragments but to calculate the whole system based on the regional localized molecular orbitals (RLMOs) of fragments. It has been successfully implemented in GAMESS program package [33] and applied to calculate the molecular properties, such as band structure extracted from finite chain [34], local density of state calculations of DNA [35] and nonlinear optical properties of quasi-one dimensional polymers [36–38]. In this Letter, we present the new development of ELG method to calculate the electronic structure of real three dimensional complicated systems. After a brief review of standard ELG method, we explain the basic idea of 3D-ELG method. Then, the performance of 3D-ELG, especially the accuracy, is currently examined by the comparison with the corresponding conventional calculations. Finally, a summary of current work is given and the future work is pointed out.

2. The idea of elongation method

The original elongation method is proposed based on one dimensional polymer systems. A suitable size of starting clusters is calculated by conventional Hartree–Fock self-consistent-field (HF SCF) procedure. Then, the corresponding canonical molecular orbitals (CMOs) of starting cluster are transformed to an orthogonal atomic basis (OAO) by Löwdin's symmetric orthogonalization.

* Corresponding author at: Department of Material Sciences, Faculty of Engineering Sciences, Kyushu University, 6-1 Kasuga-Park, Fukuoka 816-8580, Japan. Fax: +81 925838834.
E-mail address: aoki.yuriko.397@m.kyushu-u.ac.jp (Y. Aoki).

0009-2614/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cplett.2012.12.046

In the following step, the OAO-based density matrix ($\mathbf{D}^{OAO}$) is divided into two regions: frozen region (A), defined as far away from the chain elongation point; and active region (B), the remaining unit(s) of the starting cluster. By separately diagonalizing the density matrix of these two regions, a transformation matrix from OAOs to regional orbitals (ROs) is obtained. Due to the unwanted mixing between occupied and unoccupied orbitals, the full density matrix in ROs ($\mathbf{D}^{RO}$) is block diagonalized by Jacobi procedure with blocks corresponding to occupied and unoccupied subspaces. After this elongation localization procedure finishes, a set of RLMOs for both A and B regions are obtained.

Then, one new attacking unit (C) is added to the chain propagation point for the next ELG step. The eigenvalue problem is solved again but only concentrated on the B and C. In other words, during the ELG SCF procedure, the calculation of frozen part is omitted. After the SCF converges, the CMOs of B and C will be localized again to a new frozen region (RLMOs A′) and a new active region (RLMOs B′) by the procedure described above. Another new attacking unit (D) is continuously added to repeat the above calculations until the desired length is reached. The detailed descriptions of the elongation method and the elongation localization are given by Ref. [32] and [38].

As the elongation procedure sufficiently proceeds, the corresponding HF equations are solved within a very low dimension of only active region and attacking unit. Meanwhile, if the distance (or interaction) between A and C is far away, the calculations of two-electron repulsion integrals (ERIs) is partly omitted by cutoff technique [39,40]. The calculations of quasi-one dimensional model systems confirm the quit good performance on both accuracy and efficiency [32,39].

3. 3D-ELG for real three dimensional systems

For the three dimensional systems calculations, because of the uncertain direction of the attacking monomer, these frozen units may be close to the attacking monomer in a given ELG step. Therefore, it should not always keep these units as frozen as current standard one dimensional ELG procedure does. In this sense, a three dimensional elongation method is proposed [41]. Based on the standard ELG method, two additional steps are introduced: activate and re-frozen steps. The activate step is to evaluate the interactions among all the units in frozen region (A), active region (B) and attacking monomer (C). While the re-frozen step is to localize these activated units back to RLMOs for the next ELG step.

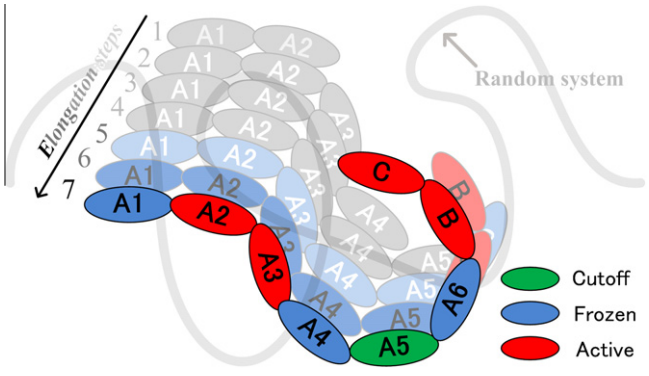

Figure 1. The scheme of three dimensional elongation method for random large system calculations. The whole random system is shown as ribbon. The calculation is approached by continuously adding one unit in each ELG step until the whole structure is reached. Supposing the calculation is performing on B plus C at the 7th ELG step, from A1 to A6 are in frozen region before SCF procedure. The green unit is set as cutoff, meanwhile the A2 and A3 is treated as active region; the rest units (A1, A4 and A6) are frozen.

The basic idea is explained in details with Figure 1. Supposing it is performing the calculation on B plus C regions, at the same time, from unit A1 to A6 are in the frozen region. Then the magnitude of the interaction between the unit in frozen region (A1 to A6) and attacking unit (C) is evaluated separately before current step SCF calculation. If the interaction is stronger than a given threshold, this frozen unit(s) is activated and included in the successive ELG SCF procedure. It means this unit(s) is crucial to the calculation of B plus C. In the example, A2 and A3 are activated and participate in the SCF calculation with B plus C. Because of the far away distance between A5 and B plus C, it is set as cutoff and partly skipped from the ERIs calculations. The rest units (A1, A4 and A6) are considered as frozen region, which is only excluded from the solution of SCF. After the ELG SCF converges, this activated unit(s) is re-frozen again and prepared for the next steps. The re-frozen procedure is similar to the standard elongation localization procedure. Firstly the CMO of A2, A3, B plus C is obtained after the SCF convergence, then A2 is treated as frozen region, the rests (A3, B plus C) is treated as active region which is the same definition of the standard elongation localization procedure. After the separation between A2 and the rests (A3, B plus C) finished, then the program is to set unit A3 as frozen region, and B plus C is treated as active region again. After the localization of A3 is finished, a standard elongation localization procedure is employed for B plus C. After that, one new attacking monomer is added to continue the elongation calculation and repeat the procedure described above until the whole system is reached.

For the HF calculation on large systems, the construction of the Coulomb matrix, exchange matrix and diagonalization of the Fock matrix constitute the three most time-consuming steps. The diagonalization of Fock matrix intrinsically scales as O (N^3). While in the elongation scheme, the diagonalization calculation is only performed on active regions, which makes the SCF solving linearly with the number of units. However, it is important to realize that the elongation algorithm does not help to reduce the scale of the Coulomb and exchange matrix computation, if no cutoff is performed. The linear scaling of calculation of Coulomb matrix and exchange matrix is mainly provided by the fast multipole method (FMM) [42–45] and the linear exchange K approach (LinK) [46–48] respectively.

4. Results and discussion

The aim of this article is to validate the possibility of elongation methodology and its accuracy to perform *ab initio* calculations on real three dimensional systems. Therefore, a strict threshold for the judgment of unit status (active, frozen or cutoff) is currently employed. If the interaction between the unit in frozen region and attacking monomer, evaluated by the maximum value of overlap, is greater than 1.0×10^{-5}, then the status of the unit is set as active. If the maximum absolute value of the eigenvector of the unit in frozen region is less than 5.0×10^{-5}, then this unit is set as cutoff. Finally the statuses of the rest units keep as frozen. The structures of model systems are all downloaded from the RCSB Protein Data Bank [49]. Currently, we focus on the neutral systems, therefore the charges of each residue in all model structures are saturated to neutral by adding/removing hydrogen atoms.

4.1. Accuracy and efficiency comparisons of quasi-one dimensional system

In this section we firstly perform the calculation on a quasi-one dimensional structure (see Figure 2, PDB code: 2KZ9, model 1, total 1117 atoms) to estimate the current settings and assess the performance of ELG approach. According to our previous experience, the