

J Mol Model (2014) 20:2277  
DOI 10.1007/s00894-014-2277-6

ORIGINAL PAPER

# Intermediate electrostatic field for the elongation method

Piotr Kuźniarowicz · Kai Liu · Yuriko Aoki ·  
Feng Long Gu · Anna Stachowicz · Jacek Korchowiec

Received: 29 November 2013 / Accepted: 25 April 2014 / Published online: 31 May 2014  
© The Author(s) 2014. This article is published with open access at [Springerlink.com](http://Springerlink.com)

**Abstract** A simple way to improve the accuracy of the fragmentation methods is proposed. The formalism was applied to the elongation (ELG) method at restricted open-shell Hartree-Fock (ROHF) level of theory. The  $\alpha$ -helix conformer of polyglycine was taken as a model system. The modified ELG method includes a simplified electrostatic field resulting from point-charge distribution of the system's environment. In this way the long-distance polarization is approximately taken into account. The field attenuates during the ELG process to eventually disappear when the final structure is reached. The point-charge distributions for each ELG step are obtained from charge sensitivity analysis (CSA) in force-field atoms resolution. The presence of the intermediate field improves the accuracy of ELG calculations. The errors in total energy and its kinetic and potential contributions are reduced by at least one-order of magnitude. In addition the SCF convergence of ROHF scheme is improved.

**Keywords** Charge sensitivity analysis · Electronegativity equalization equations · Elongation cutoff method · Elongation method · Population analyses · Order-N methods

## Introduction

Development of linear scaling methods, or order-N methods, is of primary importance in quantum chemistry. There are two main approaches leading to linear scaling. They can be classified as numerical and chemical approaches. The former linearizes every step of Hartree-Fock (HF) or Kohn-Sham (KS) schemes. Nowadays, the construction of HF and KS matrices is linear [1–8]. To reach this goal and to solve HF/KS equations several techniques are applied, for example, prescreening techniques [9], the continuous fast multipole method [1, 10–14], order- $N$  exchange [15], near-field exchange [16], tree-code approaches [14, 17], density fitting, density matrix minimization techniques [18–21], linear scaling quadrature methods [22, 23], sparse matrix algebra [24, 25], code parallelization [26], and GPU accelerators [27]. The chemical approaches are based on the so called nearsightedness approximation [28]. It assumes that the electronic structure of a given molecular fragment is predominantly determined by its nearest neighborhood and the influence of very distant fragments is negligible. The nearsightedness approximation is strongly related to generalized product function of McWeeny [29]. Several techniques using generalized product function have been proposed. They are called the fragmentation methods. They differ in conditions imposed on mutual interactions among fragments. In the most advanced fragment molecular orbital (FMO) method [30–34] and elongation (ELG) method [35–41] all interactions are rigorously taken into account. The opposite situation appears when fragments are treated completely independently [42, 43]. Intermediate situations are also possible [44–51]. However, one should

This paper belongs to Topical Collection MIB 2013 (Modeling Interactions in Biomolecules VI)

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

K. Liu · Y. Aoki  
Department of Material Sciences, Faculty of Engineering Sciences,  
Kyushu University, Kasuga, Fukuoka 816-8580, Japan

F. L. Gu  
MOE Key Laboratory of Theoretical Chemistry of Environment;  
School of Chemistry and Environment, South China Normal  
University, Guangzhou 510631, China

A. Stachowicz · J. Korchowiec (✉)  
K. Gumiński Department of Theoretical Chemistry, Faculty of  
Chemistry, Jagiellonian University, Ingardena 3, 30-060 Krakow,  
Poland  
e-mail: [korchow@chemia.uj.edu.pl](mailto:korchow@chemia.uj.edu.pl)

remember that by simplifying the interactions and narrowing the variation space a systematic error is introduced.

In this paper we describe a simple way to improve the accuracy of the fragmentation methods. The self consistent field (SCF) calculations for selected molecular fragments are performed in the electrostatic field created by the point-charge distribution of the fragment's complementary part. The formalism is applied to elongation method where it is an intermediate construction. The field diminishes in each step of the elongation procedure and finally disappears after the electronic structure of the whole system is obtained. This intermediate electrostatic field introduces the long-distance polarization into ELG calculations and is a step beyond the nearsightedness approximation. There have been some attempts to include the polarization effects in charged polymers, however, they limited the fragment's complementary part to charged groups only. Here, we propose a uniform treatment of the fragment's complementary part. Point charge distribution was also used by Exner and Mezey [45] to diminish the distance criterion in the adjustable density matrix assembler method.

The paper is organized as follows. First, we briefly describe the ELG methods and charge sensitivity analysis (CSA) used to derive the point-charge distributions. Next, we give information connected with computational details. Then, the performance of the method is shown. The analysis is focused on the accuracy. Finally, the conclusions are given along with the future prospects.

### Elongation and elongation cutoff technique

Elongation method mimics the polymerization/copolymerization reaction mechanism [35, 36]. The electronic structure of the whole molecular system ( $M$ ) is synthesized by enlarging a so called starting cluster ( $M_1$ ). The SCF calculations performed on  $M_1$  initialize the ELG process. This step is complemented with molecular orbital (MO) localization procedure. The canonical  $M$  two fragments:  $A_1$  and  $B_1$  [37]. Next, the size of the system is enlarged by adding a new molecular fragment  $C_1$ . The localized molecular orbitals (LMOs) assigned to fragment  $A_1$ , which is far away from the chain propagation center, are frozen. The LMOs of  $B_1$  and guess MOs of  $C_1$  constitute the variation space  $S_1$ . Then, SCF calculations on  $S_1$  are performed. Such propagation procedure is continued, step by step, until  $M$  is built. Every SCF step of chain propagation is followed by molecular orbital localization. Canonical MOs are localized onto  $A_i$  and  $B_i$  fragments. The LMOs assigned to  $A_i$  are excluded from variation space (they are frozen in the ELG process). The LMOs assigned to  $B_i$  together with MOs of  $C_i$  constitute the active space ( $S_i$ ). The whole ELG procedure can be summarized as follows:

$$M_1 \stackrel{\text{LMO}}{=} (A_1 | B_1) \quad (1a)$$

$$\begin{cases} M_2 = (A_1 | B_1 + C_1) \stackrel{\text{SCF}}{=} (A_1 | S_1) \stackrel{\text{LMO}}{=} (A_1 + A_2 | B_2) \equiv (A^2 | B_2) \\ M_3 = (A^2 | B_2 + C_2) \stackrel{\text{SCF}}{=} (A^2 | S_2) \stackrel{\text{LMO}}{=} (A^2 + A_3 | B_3) \equiv (A^3 | B_3) \\ \vdots \\ M_{n-1} = (A^{n-2} | B_{n-2} + C_{n-2}) \stackrel{\text{SCF}}{=} (A^{n-2} | S_{n-2}) \stackrel{\text{LMO}}{=} (A^{n-2} + A_{n-1} | B_{n-1}) \\ \equiv (A^{n-1} | B_{n-1}) \end{cases} \quad (1b)$$

$$M = M_n = (A^{n-1} | B_{n-1} + C_{n-1}) \stackrel{\text{SCF}}{=} (A^{n-1} | S_{n-1}) \quad (1c)$$

Equations 1a, 1b, and 1c correspond to initialization, chain propagation, and chain termination, respectively. Notice that there is no need to localize MO in the final step (Eq. 1c).

In the ELG scheme the size of variation space is almost constant. One can eventually take advantage of the sparseness in the LMO representation [39, 40, 52, 53]. In the limit of perfect localization, LMOs assigned to the frozen fragment ( $C_{A_i}$ ) have no tails in the active fragment and vice versa, LMOs assigned to the active fragment ( $C_{S_i}$ ) have no tails in the frozen one. Therefore, instead of constructing the Fock or Kohn-Sham matrix of  $M_{i-1}$  ( $\mathbf{F}_{M_{i-1}M_{i-1}}^{\text{AO}}$ ), the  $\mathbf{F}_{S_iS_i}^{\text{AO}}$  block is built. Such elongation cutoff scheme (ELG/C) operates within a low-dimensional subspace,

$$\begin{aligned} C_{S_i}^\dagger \mathbf{F}_{i+1} C_{S_i} &= \begin{pmatrix} \mathbf{0}_{A_iS_i}^\dagger & C_{S_iS_i}^\dagger \end{pmatrix} \begin{pmatrix} \mathbf{F}_{A_iA_i}^{\text{AO}} & \mathbf{F}_{A_iS_i}^{\text{AO}} \\ \mathbf{F}_{S_iA_i}^{\text{AO}} & \mathbf{F}_{S_iS_i}^{\text{AO}} \end{pmatrix} \begin{pmatrix} \mathbf{0}_{A_iS_i} \\ C_{S_iS_i} \end{pmatrix} \\ &\equiv \mathbf{F}_{S_iS_i}^{\text{MO}} = C_{S_iS_i}^\dagger \mathbf{F}_{S_iS_i}^{\text{AO}} C_{S_iS_i} \end{aligned} \quad (2)$$

and avoids the known bottleneck of the SCF calculations, i.e., diagonalization. The lower indices in Eq. (2) are introduced to indicate dimensions and the block  $\mathbf{0}_{A_iS_i}$  is filled with zeros. In addition, the number of two-electron repulsion integrals is substantially reduced as long as the total energy of  $M_i$  is not needed. It was demonstrated that ELG/C scheme is linear in CPU time at HF [52] and KS [53] levels of theory for a linear or quasi-linear polymer. The ELG scheme can be generalized to three-dimensional (3D) systems. In such a generalized ELG scheme the frozen LMOs of a given fragment can reenter the variation space [54–56].

In the ELG (ELG/C) scheme the molecular fragments are not treated equivalently. Namely, the starting cluster does not know its “future” while  $S_{n-1}$  possesses the whole knowledge of  $M$ . Such way of building  $M$  is consistent with the nearsightedness approximation [28] which is a common

assumption in the fragmentation methods. However, it is also a source of errors. To reduce this error, the ELG (ELG/C) process can be performed in an approximate field arising from complementary part of  $M_i$ , for example its point charge distribution. The  $i$ th subsystem's complementary part is denoted by  $\overline{M}_i$ . By enlarging the system size ( $i \rightarrow n$ ) the complementary part becomes smaller and finally ( $i=n$ ) disappears ( $\overline{M}_n$  is a zero set). This electrostatic field is an intermediate construction, therefore, the final energy of  $M$  should be greater than the reference HF (KS) one.

### Charge sensitivity analysis in force-field atoms resolution

The intermediate point-charge distribution can be easily obtained from charge sensitivity analysis [57]. CSA is based on second-order Taylor expansion of the system's energy  $E_M$  with respect to atomic charges. The CSA formalism in global resolution (without constraints on charge flow) can be summarized in a single matrix equation [58]:

$$\begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & \eta_{11} & \eta_{12} & \cdots & \eta_{1N} \\ 1 & \eta_{21} & \eta_{22} & \cdots & \eta_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \eta_{N1} & \eta_{N2} & \cdots & \eta_{NN} \end{pmatrix} \begin{pmatrix} -\chi \\ q_1 \\ q_2 \\ \vdots \\ q_N \end{pmatrix} = \begin{pmatrix} q \\ -\chi_1^* \\ -\chi_2^* \\ \vdots \\ -\chi_N^* \end{pmatrix} \tag{3}$$

where  $\eta = \{\eta_{ij} = \partial^2 E_M / \partial q_i \partial q_j\}$  is the hardness matrix,  $q$  is the total charge and  $\chi = \partial E_M / \partial q$  is the global electronegativity [59, 60] of a molecular system  $M$  composed of  $N$  atoms. Vectors  $\mathbf{q} = (q_1, q_2, \dots, q_N)^T$  and  $\mathbf{\chi} = (\chi_1^*, \chi_2^*, \dots, \chi_N^*)^T$  group the atomic charges and electronegativities, respectively. The first equation in (3) is a closure relation:

$$q = \sum_{i=1}^N q_i \tag{4}$$

The remaining equations

$$\chi = \chi_i = \chi_i^* + \sum_{j=1}^N \eta_{ij} q_j, \quad j = 1, 2, \dots, N \tag{5}$$

are the electronegativity equalization equations ( $\chi_1 = \chi_2 = \dots = \chi_N = \chi$ ) [61]. The charge distribution inside  $M$  can be obtained by inverting Eq. (3):

$$\begin{pmatrix} -\chi \\ q_1 \\ q_2 \\ \vdots \\ q_N \end{pmatrix} = \begin{pmatrix} -\eta & f_1 & f_2 & \cdots & f_N \\ f_1 & -\beta_{11} & -\beta_{12} & \cdots & -\beta_{1N} \\ f_2 & -\beta_{21} & -\beta_{22} & \cdots & -\beta_{2N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_N & -\beta_{N1} & -\beta_{N2} & \cdots & -\beta_{NN} \end{pmatrix} \begin{pmatrix} q \\ -\chi_1^* \\ -\chi_2^* \\ \vdots \\ -\chi_N^* \end{pmatrix} \tag{6}$$

Here,  $\beta = \{\beta_{ij} = \partial^2 E_M / \partial v_i \partial v_j = -\partial q_i / \partial v_i = -\partial q_j / \partial v_j\}$  is the polarization matrix (linear response matrix). The vector  $\mathbf{v} = (v_1, v_2, \dots, v_N)$  denotes the external potential due to nuclei. The remaining undefined quantities are the global hardness [62, 63]  $\eta = \partial^2 E_M / \partial q^2$  and the Fukui function (FF) [63] vector  $\mathbf{f} = \{f_i = (\partial q_i / \partial q)_v = -(\partial \mu / \partial v_i)_q\}$ . The hardness matrix  $\eta$  and the vector of atomic electronegativities  $\chi$  are the base parameters of CSA. In the force-field atoms resolution they depend on the atomic number, hybridization and local chemical environment of atoms constituting the system. We have recently derived these parameters for different population analyses [64, 65].

### Computational details

All ELG (ELG/C) calculations were performed at restricted open-shell Hartree-Fock (ROHF) level of theory using GAMESS package [66, 67]. Three different basis sets, namely: STO-3G, 6-31G, and 6-31G(d), were applied. The alpha-helix conformer of polyglycine was taken as a model system. The starting cluster was built of 15 amino acid units. In each step of the elongation eight units were frozen and another eight units were added to the system. This ELG propagation scheme is denoted as 15/8. The elongation process was terminated for 55 glycine units. In the case of ELG/C calculations, the cutoff procedure was initialized for a system made up of 31 units (15/8:31). The geometry of alpha-helix was the same as in our earlier paper [52]. We have chosen this conformer since, for a given partitioning scheme, the error in its total energy was bigger than for other conformers (C5, C7, and 3<sub>10</sub>-helix).

The point charge distribution for different population analyses was obtained from CSA calculations. Five different charge distributions were used, namely, Bader (B) [68], Hirshfeld (H) [69], M\"ulliken (M) [70], Natural (N) [71], and Voronoi (V) [72] population analyses. The ELG calculations were carried out for each charge distribution. The errors in total (tot), kinetic (kin), and potential (pot) energies were defined with respect to the conventional (supermolecule) ROHF energies:

$$\Delta E_x = E_x^{\text{ELG}} - E_x^{\text{ROHF}}, \quad x = (\text{tot}, \text{kin}, \text{pot}) \tag{7}$$

By definition  $\Delta E_M = \Delta E_{\text{tot}}$  is greater than zero since ELG (ELG/C) is a variational method. There are no such restrictions on its components, therefore,  $\Delta E_{\text{kin}}$  ( $\Delta E_{\text{pot}}$ ) may be either negative or positive. We have chosen ROHF calculations since the propagation at RHF level of theory would require saturation of "broken" bonds. This means that each intermediate subsystem (15, 23, 31, 39, and 47 units) would have to be saturated by a hydrogen atom. To avoid perturbation created by this artificial hydrogen atom ROHF scheme

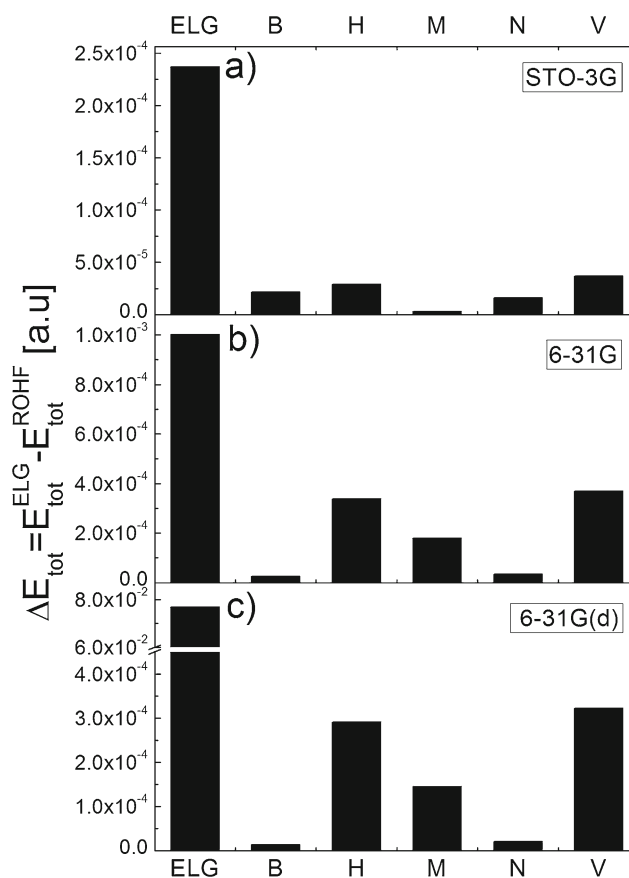
was selected. One can expect that the intermediate electrostatic field should stabilize the system and should improve the SCF convergence. The proposed modification of ELG (ELG/C) scheme has no influence on CPU time since one-electron integrals are computed only once, before the SCF process.

## Results and discussion

The errors introduced by ELG and modified ELG schemes are plotted in Fig. 1. Parts a, b, and c correspond to STO-3G, 6-31G, and 6-31G(d) basis sets, respectively. The error in the total energy is always positive. The ELG and modified ELG methods are variational, therefore the calculated energies are higher than the reference ROHF energies. All energies are computed for the final system built up of 55 units. The intermediate energies cannot be directly compared with the reference ROHF energies since they have quite a different meaning. The first error bar in each figure corresponds to standard ELG calculations with  $q_{M_i} = 0$  (no field). The remaining error bars correspond to different population analyses employed in the modified ELG procedure. The error in ELG calculation depends on the basis set and increases with its size. For 6-31G(d) basis set it is about two-orders of magnitude greater than for the minimal basis set. The presence of the intermediate field arising from charges from every population analysis reduces the error in the total energy. The reproduction of conventional ROHF energies is improved by about one order of magnitude for STO-3G (Fig. 1a) and 6-31G (Fig. 1b) basis sets. The improvement for 6-31G(d) basis is the most pronounced (Fig. 1c). The error is reduced by three orders of magnitude for B, M, and N population analyses. For all basis sets the modified ELG scheme with V and H charges works slightly worse than with B, M, and N charges. The changeability in total energy with the basis set size is the smallest for B and N charges. The errors in total energies are equal to 0.01, 0.02, and 0.02 kcal mol<sup>-1</sup> for STO, 6-31G, and 6-31G(d) basis sets, respectively.

Therefore, by taking the long-distance polarization into account, even in the simplified way, the reproduction of the reference ROHF energy is much better and the error does not exceed 1 kcal mol<sup>-1</sup>. It should also be mentioned that, although Fig. 1 corresponds to ELG and modified ELG schemes, they present the errors in the ELG/C and modified ELG/C schemes. The reason for that is that the cutoff error ( $E_{\text{tot}}^{\text{ELG/C}} - E_{\text{tot}}^{\text{ELG/C}}$ ) is at least two orders of magnitude smaller than the elongation error ( $E_{\text{tot}}^{\text{ELG}} - E_{\text{tot}}^{\text{ROHF}}$ ).

To understand the improvement in reproduction of the system's total energy we have decomposed  $E_{\text{tot}}^{\text{ELG}}$  into its potential ( $E_{\text{pot}}^{\text{ELG}}$ ) and kinetic ( $E_{\text{kin}}^{\text{ELG}}$ ) components. The errors with respect to conventional potential ( $E_{\text{pot}}^{\text{ROHF}}$ ) and kinetic ( $E_{\text{kin}}^{\text{ROHF}}$ ) energies are illustrated in Fig. 2. The black error bars

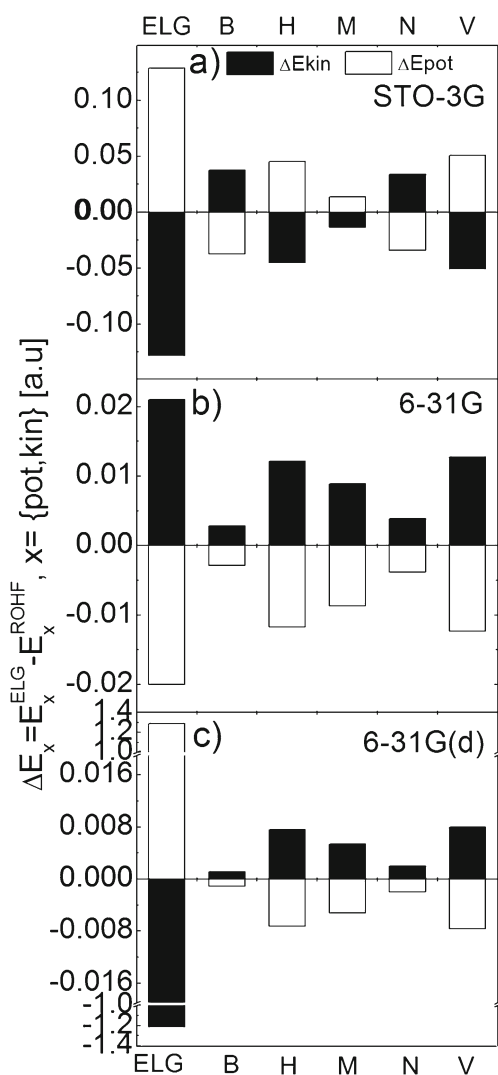


**Fig. 1** The error [a.u.] in the total energies introduced by 15/8 ELG and modified ELG schemes [with Bader (B), Hirshfeld (H), Mülliken (M), Natural (N), and Voronoi (V) charge distributions] with respect to conventional ROHF calculations for polyglycine chain made up of 55 units. Parts (a), (b), and (c) correspond to STO-3G, 6-31G, and 6-31G(d) basis sets

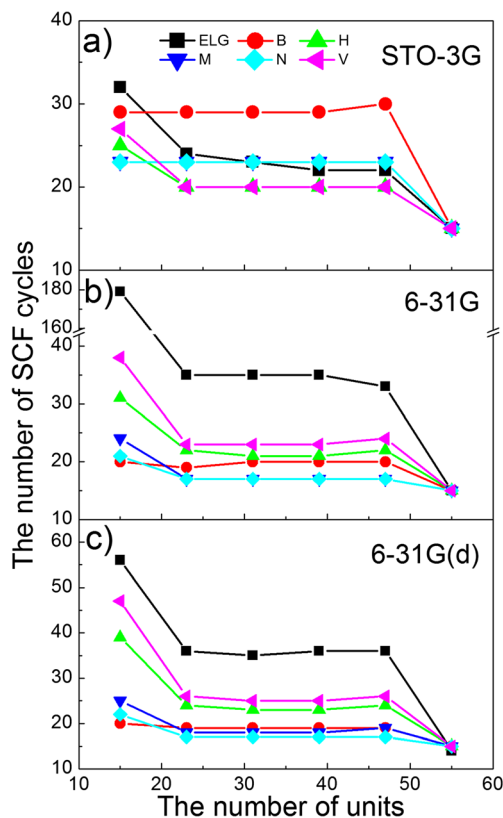
correspond to  $\Delta E_{\text{kin}}$  while white error bars to  $\Delta E_{\text{pot}}$ . Again, the first error bar in each figure corresponds to standard ELG scheme. As it was mentioned in **Computational details**, the errors in potential and kinetic energies may be either positive or negative. It is clearly seen in Fig. 2a. Depending on the population analysis, the error in kinetic energy is positive (B and N) or negative (H, M and V). Regardless of the basis set and the population analysis, the error in potential energy always has the opposite sign to  $\Delta E_{\text{kin}}$ . Its magnitude is almost the same as that of  $\Delta E_{\text{pot}}$ . This error cancellation causes the accuracy in the total energy to be one order of magnitude greater than in its potential and kinetic components. Based on the virial theorem, a different ratio  $\Delta E_{\text{pot}}/\Delta E_{\text{kin}}$  should be expected. However, one should remember that the virial theorem is exact for the true ground-state wave function. The approximate wave function fulfils it only approximately. For more extended basis sets [6-31G and 6-31G(d)],  $\Delta E_{\text{pot}}$  for modified ELG scheme is negative and  $\Delta E_{\text{kin}}$  is positive. In the case of standard ELG scheme,  $\Delta E_{\text{pot}}$  is negative for 6-31G and positive for 6-31G(d) basis sets, respectively.

It is clear from the figures that the intermediate electrostatic field improves the reproduction of kinetic and potential energies. Such improvement is evident for 6-31G(d) basis set. One should expect such behavior since slightly populated polarization functions on heavy atoms should be sensitive to the intermediate field used in the modified ELG calculations. All the results of modified ELG calculations at 6-31G(d) basis set show that the kinetic energy is overestimated due to the influence of point charges.

Finally, let us analyze the SCF convergence during ELG calculations. All intermediate subsystems are radicals (doublets). Only the whole system is closed-shell. The SCF convergence for radicals is worse than for closed-shell singlet states. The SCF convergence is illustrated in Fig. 3. Parts a, b, and c correspond to STO-3G, 6-31G, and 6-31G(d) bases,



**Fig. 2** The error [a.u.] in the potential (white bars) and kinetic (black bars) energies introduced by 15/8 ELG and modified ELG schemes [with Bader (B), Hirshfeld (H), Mülliken (M), Natural (N), and Voronoi (V) charge distribution] with respect to conventional ROHF calculations for polyglycine chain made up of 55 units. Parts (a), (b), and (c) correspond to STO-3G, 6-31G, and 6-31G(d) basis sets



**Fig. 3** The SCF convergence during ELG and the modified ELG [with Bader (B), Hirshfeld (H), Mülliken (M), Natural (N), and Voronoi (V) charge distribution] calculations. Parts (a), (b), and (c) correspond to STO-3G, 6-31G, and 6-31G(d) basis sets

respectively. One can observe that, except for B, M, and N charges at STO-3G basis set calculations, the field stabilizes the radicals. It means that fewer SCF cycles are required to reach the same assumed accuracy in comparison to standard ELG scheme. The differences between ELG and modified ELG schemes are more pronounced for the starting cluster. For larger basis set, the M and N population analyses give the fastest convergence. In general, the natural population analysis (N) gives the best performance of calculations both in accuracy and convergence. At the end of the elongation process the number of iterations is the same for all curves. The system is then closed-shell and the number of iterations in the SCF cycle is small.

### Conclusions

A simple modification of the elongation (elongation cutoff) method was proposed. Namely, the intermediate electrostatic field was introduced. The field is exerted by distributed monopoles located in the positions of atoms of the system's complementary part and it disappears in the final stage of the elongation calculations. Therefore, it does not violate the variational character of the ELG method. The modified ELG

scheme was tested for alpha-helix of polyglycine chain (55 glycine units). Several population analyses were applied. Charges were computed using CSA scheme, independently of the quantum chemical calculations. It was shown that the long-distance polarization, introduced by the field, improved the performance of the ELG method. The errors in the total, kinetic, and potential energies were reduced by at least one order of magnitude. The natural, Bader, and Mülliken population analyses gave the best agreement with the reference, conventional ROHF energies for the largest basis set. The proposed method can be easily adopted to other fragmentation techniques.

The modified ELG method improved the convergence of the SCF process at ROHF level of theory. The temporary electrostatic field stabilized the intermediate radical, therefore fewer cycles were involved during the SCF step. We plan to adopt the formalism at RHF and UHF levels of theory.

**Acknowledgments** The works on CSA carried out by JK and AS (Polish group) were supported by Polish Ministry of Science and Higher Education (Project No. UMO-2011/01/B/ST4/00636). CSA calculations were performed on computer cluster purchased with the financial support from the European Regional Development Fund in the framework of the Polish Innovation Economy Operational Program (Contract No. POIG.02.01.00-12-023/08). JK acknowledges the travel and accommodation grant from the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan (short-term stay in Japan in Feb. 2013).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Schwegler E, Challacombe M (1999) Linear scaling computation of the Fock matrix. IV. Multipole accelerated formation of the exchange matrix. *J Chem Phys* 111(14):6223–6229. doi:10.1063/1.479926
- Challacombe M (2000) Linear scaling computation of the Fock matrix. V. Hierarchical cubature for numerical integration of the exchange-correlation matrix. *J Chem Phys* 113(22):10037–10043. doi:10.1063/1.1316012
- Challacombe M, Schwegler E (1997) Linear scaling computation of the Fock matrix. *J Chem Phys* 106(13):5526–5536. doi:10.1063/1.473575
- Gan CK, Challacombe M (2003) Linear scaling computation of the Fock matrix. VI. Data parallel computation of the exchange-correlation matrix. *J Chem Phys* 118(20):9128–9135. doi:10.1063/1.1568734
- Ochsenfeld C, White CA, Head-Gordon M (1998) Linear and sublinear scaling formation of Hartree-Fock-type exchange matrices. *J Chem Phys* 109(5):1663–1669. doi:10.1063/1.476741
- Rayson MJ, Bridson PR (2009) Highly efficient method for Kohn-Sham density functional calculations of 500–10 000 atom systems. *Phys Rev B* 80(20):205104–20510410
- Rudberg E, Rubensson EH, Salek P (2008) Hartree-Fock calculations with linearly scaling memory usage. *J Chem Phys* 128(18):184106. doi:10.1063/1.2918357
- Rudberg E, Rubensson EH, Salek P (2011) Kohn-Sham density functional theory electronic structure calculations with linearly scaling computational time and memory usage. *J Chem Theory Comput* 7(2):340–350. doi:10.1021/ct100611z
- Maurer SA, Lambrecht DS, Flaig D, Ochsenfeld C (2012) Distance-dependent Schwarz-based integral estimates for two-electron integrals: reliable tightness vs. rigorous upper bounds. *J Chem Phys* 136(14):144107–14410710
- White CA, HeadGordon M (1996) Rotating around the quartic angular momentum barrier in fast multipole method calculations. *J Chem Phys* 105(12):5061–5067. doi:10.1063/1.472369
- White CA, Headgordon M (1994) Derivation and efficient implementation of the fast multipole method. *J Chem Phys* 101(8):6593–6605. doi:10.1063/1.468354
- White CA, Johnson BG, Gill PMW, HeadGordon M (1996) Linear scaling density functional calculations via the continuous fast multipole method. *Chem Phys Lett* 253(3–4):268–278. doi:10.1016/0009-2614(96)00175-3
- Rudberg E, Salek P (2006) Efficient implementation of the fast multipole method. *J Chem Phys* 125(8):084106–08410610
- Giese TJ, York DM (2008) Extension of adaptive tree code and fast multipole methods to high angular momentum particle charge densities. *J Comput Chem* 29(12):1895–1904. doi:10.1002/jcc.20946
- Burant JC, Scuseria GE, Frisch MJ (1996) A linear scaling method for Hartree-Fock exchange calculations of large molecules. *J Chem Phys* 105(19):8969–8972. doi:10.1063/1.472627
- Schwegler E, Challacombe M (1996) Linear scaling computation of the Hartree-Fock exchange matrix. *J Chem Phys* 105(7):2726–2734. doi:10.1063/1.472135
- Challacombe M, Schwegler E, Almlöf J (1996) Fast assembly of the Coulomb matrix: a quantum chemical tree code. *J Chem Phys* 104(12):4685–4698. doi:10.1063/1.471163
- Challacombe M (1999) A simplified density matrix minimization for linear scaling self-consistent field theory. *J Chem Phys* 110(5):2332–2342. doi:10.1063/1.477969
- Larsen H, Olsen J, Jorgensen P, Helgaker T (2001) Direct optimization of the atomic-orbital density matrix using the conjugate-gradient method with a multilevel preconditioner. *J Chem Phys* 115(21):9685–9697. doi:10.1063/1.1415083
- Li XS, Millam JM, Scuseria GE, Frisch MJ, Schlegel HB (2003) Density matrix search using direct inversion in the iterative subspace as a linear scaling alternative to diagonalization in electronic structure calculations. *J Chem Phys* 119(15):7651–7658. doi:10.1063/1.1607961
- Nemeth K, Scuseria GE (2000) Linear scaling density matrix search based on sign matrices. *J Chem Phys* 113(15):6035–6041
- Stratmann RE, Scuseria GE, Frisch MJ (1996) Achieving linear scaling in exchange-correlation density functional quadratures. *Chem Phys Lett* 257(3–4):213–223. doi:10.1016/0009-2614(96)00600-8
- Stratmann RE, Burant JC, Scuseria GE, Frisch MJ (1997) Improving harmonic vibrational frequencies calculations in density functional theory. *J Chem Phys* 106(24):10175–10183. doi:10.1063/1.474047
- Birgin EG, Martinez JM, Martinez L, Rocha GB (2013) Sparse projected-gradient method as a linear-scaling low-memory alternative to diagonalization in self-consistent field electronic structure calculations. *J Chem Theory Comput* 9(2):1043–1051. doi:10.1021/ct3009683
- Bock N, Challacombe M (2013) An optimized sparse approximate matrix multiply for matrices with decay. *Siam J Scient Comput* 35(1):C72–C98. doi:10.1137/120870761
- 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(14):6608–6614. doi:10.1063/1.1790891
- Titov AV, Ufimtsev IS, Luehr N, Martinez TJ (2013) Generating efficient quantum chemistry codes for novel architectures. *J Chem Theory Comput* 9(1):213–221. doi:10.1021/ct300321a

28. Kohn W (1996) Density functional and density matrix method scaling linearly with the number of atoms. *Phys Rev Letters* 76(17):3168–3171. doi:10.1103/PhysRevLett.76.3168
29. McWeeny R (1959) The density matrix in many-electron quantum mechanics. I. generalized product functions - factorization and physical interpretation of the density matrices. *Proc Roy Soc London A* 253(1273):242–259. doi:10.1098/rspa.1959.0191
30. Fedorov DG, Ishida T, Kitaura K (2005) Multilayer formulation of the fragment molecular orbital method (FMO). *J Phys Chem A* 109(11):2638–2646. doi:10.1021/jp047186z
31. Fedorov DG, Nagata T, Kitaura K (2012) Exploring chemistry with the fragment molecular orbital method. *PCCP* 14(21):7562–7577. doi:10.1039/c2cp23784a
32. Nakata H, Fedorov DG, Nagata T, Yokojima S, Ogata K, Kitaura K, Nakamura S (2012) Unrestricted Hartree-Fock based on the fragment molecular orbital method: energy and its analytic gradient. *J Chem Phys* 137(4):044110–04411010
33. Komeiji Y, Nakano T, Fukuzawa K, Ueno Y, Inadomi Y, Nemoto T, Uebaysai M, Fedorov DG, Kitaura K (2003) Fragment molecular orbital method: application to molecular dynamics simulation, ‘ab initio FMO-MD’. *Chem Phys Lett* 372(3–4):342–347. doi:10.1016/s0009-2614(03)00430-5
34. 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
35. Imamura A, Aoki Y, Nishimoto K, Kurihara Y, Nagao A (1994) Calculations of the electronic-structure of various aperiodic polymers by an elongation method. *Int J Quantum Chem* 52(2):309–319. doi:10.1002/qua.560520206
36. Aoki Y, Suhai S, Imamura A (1994) A density-functional elongation method for the theoretical synthesis of aperiodic polymers. *Int J Quantum Chem* 52(2):267–280. doi:10.1002/qua.560520202
37. Gu FL, Aoki Y, Korchowiec J, Imamura A, Kirtman B (2004) A new localization scheme for the elongation method. *J Chem Phys* 121(21):10385–10391. doi:10.1063/1.1812736
38. Korchowiec J, Gu FL, Aoki Y (2005) Elongation method at restricted open-shell Hartree-Fock level of theory. *Int J Quantum Chem* 105(6):875–882. doi:10.1002/qua.20684
39. 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(5):785–794. doi:10.1002/qua.20448
40. 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(13):1603–1619. doi:10.1002/jcc.20466
41. 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(7):5419–5431. doi:10.1063/1.461658
42. Collins MA, Deev VA (2006) Accuracy and efficiency of electronic energies from systematic molecular fragmentation. *J Chem Phys* 125(10):104104. doi:10.1063/1.2347710
43. Deev V, Collins MA (2005) Approximate ab initio energies by systematic molecular fragmentation. *J Chem Phys* 122(15):154102–15410210
44. Exner TE, Mezey PG (2003) Ab initio quality properties for macromolecules using the ADMA approach. *J Comput Chem* 24(16):1980–1986. doi:10.1002/jcc.10340
45. Exner TE, Mezey PG (2004) The field-adapted ADMA approach: Introducing point charges. *J Phys Chem A* 108(19):4301–4309. doi:10.1021/jp037447p
46. Exner TE, Mezey PG (2005) Evaluation of the field-adapted ADMA approach: absolute and relative energies of crambin and derivatives. *PCCP* 7(24):4061–4069. doi:10.1039/b509557c
47. Li W, Li SH, Jiang YS (2007) Generalized energy-based fragmentation approach for computing the ground-state energies and properties of large molecules. *J Phys Chem A* 111(11):2193–2199. doi:10.1021/jp067721q
48. Li H, Li W, Li SH, Ma J (2008) Fragmentation-based QM/MM simulations: Length dependence of chain dynamics and hydrogen bonding of polyethylene oxide and polyethylene in aqueous solutions. *J Phys Chem B* 112(23):7061–7070. doi:10.1021/jp800777e
49. Gadre SR, Shirsat RN, Limaye AC (1994) Molecular tailoring approach for simulation of electrostatic properties. *J Phys Chem* 98(37):9165–9169. doi:10.1021/j100088a013
50. Gadre SR, Ganesh V (2006) Molecular tailoring approach: towards PC-based ab initio treatment of large molecules. *J Theoret Comput Chem* 5(4):835–855. doi:10.1142/s021963360600260x
51. Ganesh V, Kavathekar R, Rahalkar A, Gadre SR (2008) Software news and update - WebProp: Web interface for ab initio calculation of molecular one-electron properties. *J Comput Chem* 29(3):488–495. doi:10.1002/jcc.20802
52. 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(15):2515–2525. doi:10.1002/jcc.21252
53. 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(12):2130–2139. doi:10.1002/qua.22562
54. Aoki Y, Gu FL (2012) An elongation method for large systems toward bio-systems (vol 14, pg, 2012). *PCCP* 14(48):16781–16781. doi:10.1039/c2cp24033e
55. Liu K, Peng L, Gu FL, Aoki Y (2013) Three dimensional elongation method for large molecular calculations. *Chem Phys Lett* 560:66–70. doi:10.1016/j.cplett.2012.12.046
56. Liu K, Yan Y, Gu FL, Aoki Y (2013) A modified localization scheme for the three-dimensional elongation method applied to large systems. *Chem Phys Lett* 565:143–147. doi:10.1016/j.cplett.2013.02.039
57. Nalewajski RF, Korchowiec J (1998) Charge sensitivity approach to electronic structure and chemical reactivity. World Scientific, Singapore
58. Korchowiec J, Gerwens H, Jug K (1994) Relaxed Fukui function indexes and their application to chemical-reactivity problems. *Chem Phys Lett* 222(1–2):58–64. doi:10.1016/0009-2614(94)00319-x
59. Iczkowski R, Margrave JL (1961) Electronegativity. *J Am Chem Soc* 83(17):3547. doi:10.1021/ja01478a001
60. Parr RG, Donnelly RA, Levy M, Palke WE (1978) Electronegativity - density functional viewpoint. *J Chem Phys* 68(8):3801–3807. doi:10.1063/1.436185
61. Sanderson RT (1951) An interpretation of bond lengths and a classification of bonds. *Science* 114(2973):670–672. doi:10.1126/science.114.2973.670
62. Pearson RG (1963) Hard and soft acids and bases. *J Am Chem Soc* 85(22):3533–3537. doi:10.1021/ja00905a001
63. Yang WT, Parr RG (1985) Hardness, softness, and the Fukui function in the electronic theory of metals and catalysis. *Proc Natl Acad Sci U S A* 82(20):6723–6726. doi:10.1073/pnas.82.20.6723
64. Stachowicz A, Styrcz A, Korchowiec J (2011) Charge sensitivity analysis in force-field-atom resolution. *J Mol Model* 17(9):2217–2226. doi:10.1007/s00894-011-1006-7
65. Stachowicz A, Korchowiec J (2012) Generalized charge sensitivity analysis. *Struct Chem* 23(5):1449–1458. doi:10.1007/s11224-012-0054-7
66. Schmidt MW, Baldrige 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(11):1347–1363. doi:10.1002/jcc.540141112

67. Gordon MS, Schmidt MW (2005) In: Dykstra CE, Frenking G, Kim KS, Scuseria GE (eds) *Advances in electronic structure theory: GAMESS a decade later*. Elsevier, Amsterdam, pp 1167–1189
68. Bader RFW (1991) A quantum-theory of molecular-structure and its applications. *Chem Rev* 91(5):893–928. doi:[10.1021/cr00005a013](https://doi.org/10.1021/cr00005a013)
69. Hirshfeld FL (1977) Bonded-atom fragments for describing molecular charge-densities. *Theoret Chim Acta* 44(2):129–138. doi:[10.1007/bf00549096](https://doi.org/10.1007/bf00549096)
70. Mülliken RS (1955) Electronic population analysis on LCAO-MO molecular wave functions. I. *J Chem Phys* 23(10):1833–1840. doi:[10.1063/1.1740588](https://doi.org/10.1063/1.1740588)
71. Reed AE, Weinstock RB, Weinhold F (1985) Natural-population analysis. *J Chem Phys* 83(2):735–746. doi:[10.1063/1.449486](https://doi.org/10.1063/1.449486)
72. Bickelhaupt FM, Hommes N, Guerra CF, Baerends EJ (1996) The carbon-lithium electron pair bond in  $(\text{CH}_3\text{Li})_n$  ( $n=1, 2, 4$ ). *Organometallics* 15(13):2923–2931. doi:[10.1021/om950966x](https://doi.org/10.1021/om950966x)