How to determine boundaries for QM/MM calculations: A guideline based on linear response function for glutathione

Shusuke Yamanaka1, Koki Ueda1, Kazuto Nakata2, Mitsutaka Okumura1, Kizashi Yamaguchi3 and Haruki Nakamura4

1. Graduate School of Science, Osaka University, Machikaneyama 1-1, Toyonaka, Osaka, Japan.
2. NEC Corporation, 1-10, Nissincho, Fuchu, 183-8501, Japan.
3. TOYOTA Physical & Chemical Research Institute Nagakute, Aichi, Japan.
4. Protein Institute, Osaka University, 3-2, Yamadaoka Suita, Osaka, Japan.
syama@chem.sci.osaka-u.ac.jp

Abstract. Quantum mechanics/molecular mechanics (QM/MM) methods have grown to be a standard tool for chemical reactions in biological systems. Still, the remaining problem is that the MM point charges induce artificial polarizations in QM regions, spoiling the quality of the QM calculations. Thus, how to determine boundaries between QM and MM regions is an essential issue for QM/MM calculations. Recently, we proposed the use of a linear response function as an indicator to examine the validity of the replacement of QM peripheral ligands with MM point charges. In this study, we examine the glutathione molecule, for which protonation models have been proposed so far. The calculated results are discussed in relation to the QM/MM modeling of this system.

1. Introduction

The ab initio quantum chemical approach has been established for calculations of chemical reactions of various molecules. For most molecules consisting of several dozen of atoms, the chemical accuracy in estimating the activation barrier heights and formation enthalpies has been attained [1]. In contrast, the enzyme reactions are still tough targets because not only quantum fluctuations in the reaction site, but also thermal fluctuations around the site play important roles to determine quantitative values of such chemical properties. In order to tackle this problem, many researchers have developed quantum mechanics/molecular mechanics (QM/MM) methods, and applied them to biomolecular systems [2-12]. In QM/MM methods, the core parts of biomolecular systems, in which quantum events such as chemical reactions and electronic excitations occur, are treated with the QM method, while the peripheral regions such as protein environments, lipids, and waters are treated with the MM method. The QM method, in particular the ab initio QM method, is now a reliable tool to estimate chemical reactions by which we reproduce experimental results within the chemical accuracy (0.1-1 kcal/mol), and the MM method is also well trained to reproduce the experimental results of proteins [13]. However, the problems of QM/MM methods remain in the seam between QM and MM regions, where the MM point charges induce artificial overpolarizations in QM regions near the QM/MM boundaries. More explicitly, the difference between the QM potential and the MM potential, \( \Delta v = v_{QM} - v_{MM} \), which is just the error due to the replacement of the QM electronic structure by the MM point charges, causes the error of calculated density.
\[ \delta \rho (\mathbf{r}) = \int d\mathbf{r}' \frac{\delta \rho (\mathbf{r})}{\delta \nu (\mathbf{r}')} \delta \nu (\mathbf{r}'). \]  

If \( \delta \rho (\mathbf{r}) \) at the QM event site is not negligible, it leads to significant failures of computational results in describing the event. Equation (1) implies that there are two ways to improve the QM/MM methods. The first one is the standard approach that many researchers have taken: a device on the QM/MM boundaries such as link-atom [2-4], generalized hybrid orbital [6] local self-consistent field [7], and effective potential approaches [9]. Although many efforts have been devoted to minimizing \( \delta \nu = \nu_{QM} - \nu_{MM} \) in the right-hand side of (1), the complete solution has not been obtained. The second one consists in referring to \( \delta \rho (\mathbf{r})/\delta \nu (\mathbf{r}') \), which indicates how far the effects of the MM point charges propagate. In striking contrast to many studies for the first one, there are few studies for the second one. In particular, the systematic study to determine QM/MM boundaries has never been implemented so far. Previously, we proposed to use \( \delta \rho (\mathbf{r})/\delta \nu (\mathbf{r}') \) to examine how the effects of QM/MM errors depend on the intermolecular units and chemical bonds [15]. In that study, we examined typical covalent bonds, \( \pi \)-conjugated systems, and dialanine dipeptide systems.

In this study, we implemented a similar analysis for glutathione, which is known to function as an antioxidant working with glutathione (GSH) peroxidase/reductase [16]. Even without these enzymes, GSH can also react with radical species such as OH-, RO·, NO2·, and carbon-centered radicals, which is why the GSH itself is considered as an antioxidant [16]. The glutathione radical (GS·) is produced by the scavenging reaction,

\[ \text{GSH} + X \rightarrow \text{GS}· + XH, \]  

which is considered to be detoxified in the processes in vivo as follows:

\[ \text{GS}· + \text{GSH} + \text{O}_2 \rightarrow \text{GSSG} + \text{O}_2^- , \]  

\[ 2\text{O}_2^- + 2\text{H}^+ \rightarrow \text{H}_2\text{O}_2 + \text{O}_2 , \]  

\[ 2\text{H}_2\text{O}_2 \rightarrow \text{H}_2\text{O} + \text{O}_2 . \]  

Although the S-H bond has been usually assumed to be dissociated homolytically as in (2), Fiser et al. presented using density functional theory calculations with a polarizable continuum model that the N-H bond in the ammonia group is easier to dissociate than the S-H bond [17]. In their calculations, they partitioned the GSH into several parts by using “hydrogen caps”. The dehydrogenation of GSH in water is a critically complicated problem, because there are various possible modes concerning protonation/deprotonations of carboxyl and amino groups, as well as dehydrogenation modes [18]. Further, the stabilities of the modes must depend on environmental effects, in particular on pH [18]. Thus, instead of studying the various modes, we focus our attention on examination of the validity of the partition methods of glutathione molecule on the basis of the computational results of the linear response function. The results are also discussed in relation to fundamental features of the LRF of molecules, as well as the implications of our results in the context of the usual QM/MM modeling tactics.

2. Methodology

So far, ab initio calculations of linear response functions (LRF) have been implemented by several groups [19, 20]. Morita and Kato calculated \( \delta \rho (\mathbf{r})/\delta \nu (\mathbf{r}') \) on the basis of the coupled perturbed
Hartree-Fock equation [19]. Recently, Geerlings and his coworkers have investigated this property in the context of conceptual density functional theory [20]. They used a perturbation theory to estimate $\frac{\delta \rho(\mathbf{r})}{\delta \nu(\mathbf{r'})}$ and compute the linear response (LR) to the specific site with using Becke’s fuzzy cell scheme [21]. Independently we also calculated LR to the specific site, $\mathbf{X}$, i.e., $\frac{\delta \rho(\mathbf{r})}{\delta \nu_{\mathbf{X}}}$ based on the perturbation theory, but using the linear combination of atomic orbital (LCAO) expansion [14,15].

The LCAO based method is suitable for our purpose: the QM/MM modeling is applied for specific sites, which would be described as quantum electronic structures using atomic orbitals (AOs) if all systems are treated with the QM method. Thus the perturbations and the responses expressed in terms of AOs are efficiently used in the discussion of the validity of the QM/MM boundary setting.

In this study, we computed the linear response for each site of the glutathione, $\frac{\delta \rho(\mathbf{r})}{\delta \nu_{\mathbf{X}}}$ As discussed in the introduction, we focus our attention on the LRF of glutathione molecule. Thus we fixed a standard hydration mode with the geometry that was obtained by Fiser et al. [17], as shown in figure 1.

Figure 1. The structure of glutathione.

The B3LYP functional and the 6-31G** basis set were used for all calculations. Using the B3LYP solutions, we calculated the LRF based on first-order perturbation theory [14]. The code we used is a locally extended version of GAMESS [22], in which our LRF module is implemented. The results will be presented in the next section.

3. Results and discussion

Each of the 36 parts of figure 2 shows the isosurface of the linear response to the perturbation that is applied to the specific site $\mathbf{X}$, $\frac{\delta \rho(\mathbf{r})}{\delta \nu_{\mathbf{X}}}$ of the glutathione molecule. The number described below each figure corresponds to the atom number showed in figure 1. The isosurfaces for $\frac{\delta \rho(\mathbf{r})}{\delta \nu_{\mathbf{X}}} = 0.001$ and $-0.001$ are described as blue and red surfaces, respectively. The plus
value (blue region) of $\frac{\delta \rho(\mathbf{r})}{\delta v_X}$ implies that, if the attractive perturbation, $\delta v_X (< 0)$, is applied for the X site, the density decreases. Contrastingly, the minus value (red region) of $\frac{\delta \rho(\mathbf{r})}{\delta v_X}$ indicates that the density increases for the attractive perturbation.

![Figure 2](image)

**Figure 2.** The linear response to the perturbation that is applied to the specific site X, $\frac{\delta \rho(\mathbf{r})}{\delta v_X}$ of the glutathione molecule. The numbers below the figures are the sites to which the perturbations are applied.

Because we assume that the number of electrons conserves, the integral of $\frac{\delta \rho(\mathbf{r})}{\delta v_X}$ is always zero:

$$\int d\mathbf{r} \frac{\delta \rho(\mathbf{r})}{\delta v_X} = 0.$$  \hspace{1cm} (4)

Thus blue and red regions appear for all perturbations, $\delta v_X$. A noteworthy point is that, not a monotonically increase or decrease, but oscillations of $\frac{\delta \rho(\mathbf{r})}{\delta v_X}$ are observed: for instance see $\frac{\delta \rho(\mathbf{r})}{\delta v_{22}}$ (22-th) in figure 2. If this phenomenon were dominated by classical mechanics, the
behavior of $\delta \rho(\mathbf{r})$ should be monotonic. Then, it is obvious that the quantum characteristics, for instance phases of orbitals, of this system lead to oscillations of $\delta \rho(\mathbf{r})/\delta \mathbf{v}_X$. A reader might notice that this is not a size effect: a similar oscillation phenomena is known as “Friedel oscillation” in solid state physics since more than 60 years ago [23, 24], and is recently observed as STM images for condensed matter systems [25]. Previous researchers, who compute the LRF of molecular systems, also presented that there are oscillation behaviors of the LRF [19, 20]. This type of oscillations must be universal for molecules, and macromolecules, and condensed phases.

For our primary purpose, it is important that the linear responses do not propagate over the $C_\alpha$ carbons (atom numbers are 5, 10, 13, 20, 22, 31): if the perturbation is applied to the right side of a specific $C_\alpha$, the region defined by $|\delta \rho(\mathbf{r})/\delta \mathbf{v}_X| \leq 0.001$ is mostly localized within the right side of the $C_\alpha$ carbon and is not delocalized the left side of the $C_\alpha$ carbon, and vice versa. However, if the perturbation is applied to an atom next to the $C_\alpha$ carbon, the response might pierce the $C_\alpha$ carbon, because the sp3 bonding orbital that connects the atom and the $C_\alpha$ carbon is delocalized over another side of the $C_\alpha$ carbon (for instance, see 18 in figure 2). In addition, the perturbation on the $C_\alpha$ carbon itself affects to the four directions of sp3 orbitals (see 5, 10, 13, 20, 22, 31 in figure 2), possibly spreading over two amide planes. Thus when a $C_\alpha$ carbon directly contributes to the QM event, the atoms in these two amide planes should not be replaced with the MM charges.

We should note the reason why we choose the threshold $|\delta \rho(\mathbf{r})/\delta \mathbf{v}_X| = 0.001$ to illustrate the linear response surfaces. This surface provides boundaries for errors of order 0.001 on the density if the QM/MM potential yields the errors, $\delta \mathbf{v}_X$, of the order 1.0 Hartree. Then the regions illustrated in figure 2 are actually useful to judge where we should set the QM/MM boundary. Of course, this estimation is not accurate enough to predict the accuracy of the specific QM/MM calculation plans. The LRF is calculated for the whole system of the glutathione molecule using a perturbation scheme [14,15]. In actual QM/MM calculations, the electronic structures of the model systems are relaxed with the self-consistent manner, in which the QM/MM errors might be enlarged (or rarely reduced). Higher order effects are completely neglected in the present calculations. How strongly the higher-order effects affect the results depends on how large the errors due to the QM/MM model, $\delta \mathbf{v}_X$, are. Thus we need to specify the concrete QM/MM scheme in order to refine this type of error analyses for the QM/MM theories.

Nevertheless, the rough estimate described above could certainly provide valuable information for QM/MM applications. In fact, modeling of $C_\alpha$ carbons as hydrogen atoms is widely used. We expect that the discussion above becomes a guideline to determine the QM region in QM/MM calculations. We are now applying our method to small peptides having secondary structures, such as $\alpha$–helix or extended $\beta$–conformations. A larger set of comprehensive guidelines will be presented in the future.

Acknowledgments
S. Yamanaka acknowledges the financial support by the Ministry of Education, Culture, Sports, Science and Technology (MEXT) (Grant-in-Aid for Scientific Research (C) No. 23550016 and (B) No. 23350064) and Research and Development of the Next-Generation Integrated Simulation of Living Matter, as a part of the Development and Use of the Next-Generation Supercomputer Project.

This work was partially supported by MEXT program "Elements Strategy Initiative to Form Core Research Center", MEXT, Ministry of Education Culture, Sports, Science and Technology, Japan.

References
[1] Frank J 2007 Introduction to Computational Chemistry (Chichester England, John Wiley and
Sons)
[2] Warshel A and Levitt M J 1976 J. Mol. Biol. 103 227-249
[3] Senn H M and Thiel W 2009 Angew. Chem. Int. Ed. 48 1198-1229
[4] Lin H and Truhlar D G 2007 Theor. Chem. Acc. 117 185-199
[5] Hu L, Soderhjelm P and Ryde U 2011 J. Chem. Theor. Comput. 7 761-777
[6] Dilabio G A, Hurley M H and Christiansen P A 2002 J. Chem. Phys. 116 9578 –9584
[7] Murphy R. B. D. M. Philipp and R. A. Friesner 2000 Chem. Phys. Lett. 321 113-120
[8] Gao J, Amara P, Alhambra C and Field M J 1998 J. Phys. Chem. A 102 4714-4721
[9] Reuter N, Dejaegere A, Maigret B and Karplus M 2000 J. Phys. Chem. A 104 1720-1735
[10] Dittrich M, Hayashi S and K. Schulten K 2004 Biophys. J. 87 2954-2967
[11] Yasuda K and Yamaki D 2004 J. Chem. Phys. 121 3964.3972
[12] Takahashi H, Matubayasi N, Nakahara M and Nitta T 2004 J. Chem. Phys. 121 3989-3999
[13] Mackerell AD 2004 Empirical force fields for biological macromolecules: overview and issues J. Comput. Chem. 25 1584–604
[14] Ueda K, Yamanaka S, Nakata K, Ebara M, Okumura M, Yamaguchi K and Nakamura H 2013 Int. J. Quantum Chem. 113 336-341
[15] Yamanaka S, Yonezawa Y, Nakata K, Nishihara S, Okumura M, Takada T, Yamaguchi K and Nakamura H 2013 AIP Conf Proc 1504 916-919
[16] Halliwell B and Gutteridge J C 2007 Free radicals in biology and medicine fourth Edition Oxford University Press
[17] Fiser B, Szori M, Jojart B, Csizmadia I. G and Viskolcz B 2011 J. Phys. Chem. B 115 11269-11277
[18] Lampela O, Juffer A H and Rauk A 2003 J. Phys. Chem. A 107 9208-9220
[19] Morita A and Kato S 1997 J. Am. Chem. Soc. 119 4021-4032
[20] For instance, Sablon N, Proft F D and Gerlings P 2010 J. Phys. Chem. Lett. 1 1228-1234
[21] Becke A 1988 J. Chem. Phys. 88 2547-2553
[22] Schmidt MW, Baldrige K K, Boatz J A, Elbert S T, Gordon M S, Jensen J H, Koseki S, Matsunaga N, Nguyen K A, Su S, T L, Windus, Dupuis M and Montgomery Jr A 1993 J. Comp. Chem. 14 1347-1363
[23] Friedel J 1952 Philos. Mag. 43 153-189
[24] Kohn W and Vosko S H 1960 Phys. Rev. 119 912-918
[25] Hasegawa Y, Ono M, Nishigata Y, Nishino T and Eguchi T 2006 Phys. Rev. Lett. 96 016801-4