

# A Multicore QM/MM Approach for the Geometry Optimization of Chromophore Aggregate in Protein

YASUOMI KIYOTA,<sup>1</sup> JUN-YA HASEGAWA,<sup>1,2</sup> KAZUHIRO FUJIMOTO,<sup>1</sup> BEN SWERTS,<sup>1</sup> HIROSHI NAKATSUJI<sup>1,2</sup>

<sup>1</sup>Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Kyoto-Daigaku-Katsura, Nishikyo-ku, Kyoto 615-8510, Japan

<sup>2</sup>Quantum Chemistry Research Institute (QCRI), JST-CREST, Kyodai Katsura Venture Plaza, Goryou Oohara 1-36, Nishikyo-ku, Kyoto 615-8245, Japan

Received 23 June 2008; Revised 9 September 2008; Accepted 1 October 2008

DOI 10.1002/jcc.21156

Published online 13 November 2008 in Wiley InterScience (www.interscience.wiley.com).

**Abstract:** In this article, we present the multicore (mc) QM/MM method, a QM/MM method that can optimize the structure of chromophore aggregate in protein. A QM region is composed of the sum of the QM subregions that are small enough to apply practical electronic structure calculations. QM/MM energy gradient calculations are performed for each QM subregion. Several benchmark examinations were carried out to figure out availabilities and limitations. In the interregion distances of more than 3.5–4.0 Å, the mcQM/MM energy gradient is very close to that obtained by the ordinary QM/MM method in which all the QM subregions were treated together as a single QM region. In van der Waals complex, the error exponentially drops with the distance, while the error decreases slowly in a hydrogen bonding complex. On the other hand, the optimized structures were reproduced with reasonable accuracy in both cases. The computational efficiency is the best advantage in the mcQM/MM approach, especially when the QM region is significantly large and the QM method used is computationally demanding. With this approach, we could optimize the structures of a bacterial photosynthetic reaction center protein in the ground and excited states, which consists of more than 14,000 atoms.

© 2008 Wiley Periodicals, Inc. J Comput Chem 30: 1351–1359, 2009

**Key words:** QM/MM method; multiple QM regions; chromophore aggregate; geometry optimization

## Introduction

Targets of computational quantum chemistry are expanding to large molecular systems such as functional proteins, molecular assemblies, and complex surface catalyses. However, computing time and resources required for the electronic structure calculations increase rapidly with the size of the molecule. One of the feasible approaches to treat large molecules is the combined quantum mechanics and molecular mechanics (QM/MM) method (for review articles, see refs. 1 and 2). In this hybrid approach, the target molecule is divided into QM and MM regions. The QM/MM method usually adopts single QM region, which may be called single-core (sc) QM/MM method. The electronic structure calculations are performed within the QM region to describe electronic structures and chemical reactions in the ground and excited states. The rest of the molecule is treated by the classical mechanics such as MM method. The interactions between the QM and MM regions are described by using the MM force field (electrostatic, through-bond, and van der Waals interactions).<sup>1,2</sup> The greatest advantage of the QM/MM approach is the computational efficiency, which realizes the calculations of large-scale

molecules feasible without losing much accuracy. If the size of the QM region is moderate enough to apply highly accurate *ab initio* electronic structure methods, the QM/MM results would be more reliable. In our recent study, we developed a QM/MM program and performed geometry optimization for photobiological proteins such as fluorescent proteins<sup>3</sup> and retinal proteins.<sup>4,5</sup>

However, the QM/MM calculation is still limited by the size of the QM region. Our research interest is, for example, the potential energy profile of the electron transfer in the photosynthetic reaction center proteins which includes more than 14,000 atoms.<sup>6–9</sup> Photo-excitation and electron transfer occur among the six bacteriochlorophylls which involves more than 500 atoms. Such large systems were still outside the scope of the optimizations of the ground- and excited-state structures.

**Correspondence to:** J. Hasegawa or H. Nakatsuji; e-mail: hasegawa@sbchem.kyoto-u.ac.jp or h.nakatsuji@qcri.or.jp

Contract/grant sponsors: JSPS, JST-CREST, ACCMS and IIMC (Kyoto University)

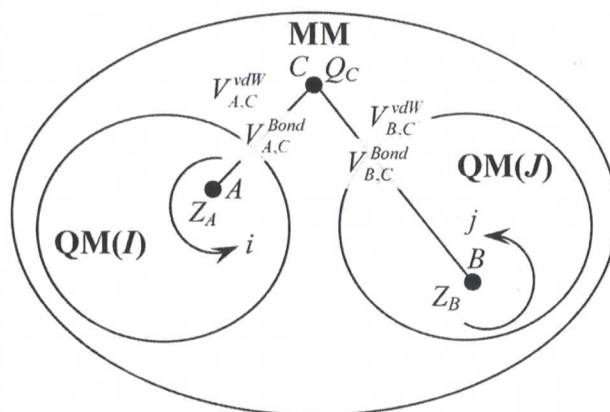
To overcome the difficulty, dividing a large QM region is a reasonable approach.<sup>10–17</sup> Electronic structure obtained by a local computational model is very close to that by a large computational model, if appropriately treated.<sup>18</sup> We sometimes replace a large substituent group into a smaller one. The other reason is that the computational effort of the electronic structure calculation increases rapidly with the size of the system. In other words, the computing time rapidly decreases, when the size of the computational model is reduced. We are interested in optimizing the structures of the ground and excited states of very large molecules, such as photosynthetic reaction center proteins. Since the energy gradient with respect to an atomic coordinate is energy difference against a slight change in the local structural parameter, such a local perturbation would affect only a limited region of a molecule.

In this study, we take a simplified approach to calculate the energy gradient of very large systems such as chromophore aggregate in protein. Within the QM/MM framework of methodology, the total QM region was defined as the sum of the QM subregions. To calculate energy gradient regarding atomic coordinates in a QM subregion, the rest of the subregions were treated as a part of the MM region. For each QM subregion, a QM/MM energy gradient calculation was performed. The QM/MM calculations are repeated until the energy gradients of the whole QM regions are obtained. The proposed method, which we hereafter call as multicore QM/MM (mcQM/MM) method, is a feasible approach to obtain the energy gradient of very large system. With the mcQM/MM approach, we could optimize the structures of the photosynthetic reaction center in the ground and excited states.

## Method

In the present approach, the QM region is defined as the sum of the QM subregions. To calculate the energy first derivative with respect to a nuclear coordinate of an atom A, only single QM subregion to which atom A belongs is treated by quantum mechanics. The interactions with the other QM subregions are treated likewise as the other MM region. This approach is based on the following idea. Energy gradient with respect to a nuclear coordinate is energy difference which arises from a slight change in a structural parameter. Such a small local perturbation should affect only limited region of a molecule. To calculate the structural energy derivative of a large QM system, its structure may be divided into small subregions. The derivative calculation is expected to be very quick and, if appropriately treated, to reproduce those obtained with the large parent QM regions.

Figure 1 shows a model system. For simplicity, the QM region is divided into two QM subregions *I* and *J*. The symbols *i*, *j*, ... and *A*, *B*, ... denote electrons and nucleus, respectively. The  $Z_A$ ,  $Z_B$ , ... and  $Q_A$ ,  $Q_B$ , ... are nuclear charges and atomic charges of atoms *A* and *B*, respectively. In the case of energy derivative with respect to a nuclear coordinate  $\alpha$  of an atom *A* in QM subregion *I*, the gradient is calculated by an ordinary scQM/MM calculation in which only the subregion *I* is treated by quantum mechanics. The QM subregion *J* is treated by an effective



**Figure 1.** A model system of the multicore QM/MM calculation. For simplicity, a case of two QM subregions surrounded by the MM region is considered.

MM Hamiltonian. The total Hamiltonian of a scQM/MM method is composed of three terms:

$$\hat{H}^{101}(\text{QM} : I) = \hat{H}^{\text{QM}}(\text{QM} : I) + \hat{H}^{\text{QM-MM}}(\text{QM} : I) + \hat{H}^{\text{MM}}(\text{QM} : I) \quad (1)$$

The first one  $\hat{H}^{\text{QM}}$  describes QM interactions within region *I*.

$$\hat{H}^{\text{QM}}(\text{QM} : I) = \frac{1}{2} \sum_{i \in I} \nabla_i^2 - \sum_{i \in I} \sum_{A \in I} \frac{Z_A}{|\mathbf{r}_i - \mathbf{r}_A|} + \sum_{(i>j) \in I} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{(A>B) \in I} \frac{Z_A Z_B}{|\mathbf{r}_A - \mathbf{r}_B|} \quad (2)$$

The atoms in the rest of the QM subregions and the MM region are treated by a MM Hamiltonian.

$$\hat{H}^{\text{MM}}(\text{QM} : I) = \sum_{(B>C) \notin I} \left( \frac{Q_B Q_C}{|\mathbf{r}_B - \mathbf{r}_C|} + V_{B,C}^{\text{vdW}} + V_{B,C}^{\text{Bond}} \right) \quad (3)$$

For the atomic charges of the QM subregions, there are several definitions such as electrostatic potential fitted (ESP) charges,<sup>19,20</sup> Mulliken charges,<sup>21,22</sup> and natural population analysis charges.<sup>23,24</sup> The  $V_{B,C}^{\text{vdW}}$  and  $V_{B,C}^{\text{Bond}}$  represent van der Waals and through-bond interaction energies, respectively, which are taken from MM force field as AMBER.<sup>25</sup> The interactions between the QM region *I* and the MM region are treated as

$$\hat{H}^{\text{QM-MM}}(\text{QM} : I) = \sum_{i \in I} \sum_{B \notin I} - \frac{Q_B}{|\mathbf{r}_i - \mathbf{r}_B|} + \sum_{A \in I} \sum_{B \notin I} V_{A,B}^{\text{vdW}} + \sum_{A \in I} \sum_{B \notin I} V_{A,B}^{\text{Bond}} \quad (4)$$

Here, the QM subregions other than the region *I* are included in the MM region. These treatments are identical to ordinary QM/MM method. In the same way, sequential scQM/MM calculations are performed for all of the QM subregions to obtain the

