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

YASUOMI KIYOTA,[1] JUN-YA HASEGAWA,[1,2] KAZUHIRO FUJIMOTO,[1] BEN SWERTS,[1] HIROSHI NAKATSUJI[1,2]

[1]*Department of Synthetic Chemistry and Biological Chemistry, Graduate School of Engineering, Kyoto University, Kyoto-Daigaku-Katsura, Nishikyo-ku, Kyoto 615-8510, Japan*

[2]*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.

 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).[1,2] 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[3] and retinal proteins.[4,5]

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.[6–9] 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.[10–17] Electronic structure obtained by a local computational model is very close to that by a large computational model, if appropriately treated.[18] 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 $i, j, \ldots$ and $A, B, \ldots$ 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 α of an atom A in QM subregion I, the gradient is calculated by a ordinary scQM/MM calculation in which only the subregion I is treated by quantum mechanics. The QM subregion J is treated by an effective MM Hamiltonian. The total Hamiltonian of a scQM/MM method is composed of three terms:

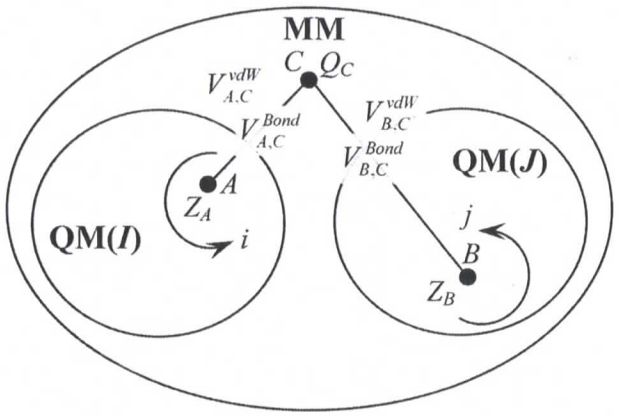

Figure 1. A model system of the multicore QM/MM calculation. For simplicity, a case of two QM subregions surrounded by the MM region is considered.

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

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

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

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

$$\hat{H}^{\mathrm{MM}}(QM:I) = \sum_{(B>C)\notin I}\left(\frac{Q_B Q_C}{|\mathbf{r}_B - \mathbf{r}_C|} + V^{\mathrm{vdW}}_{B.C} + V^{\mathrm{Bond}}_{B.C}\right) \tag{3}$$

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

$$\hat{H}^{\mathrm{QM-MM}}(QM:I) = \sum_{i\in I}\sum_{B\notin I}-\frac{Q_B}{|\mathbf{r}_i - \mathbf{r}_B|} + \sum_{A\in I}\sum_{B\notin I}V^{\mathrm{vdW}}_{A.B} + \sum_{A\in I}\sum_{B\notin I}V^{\mathrm{Bond}}_{A.B}. \tag{4}$$

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

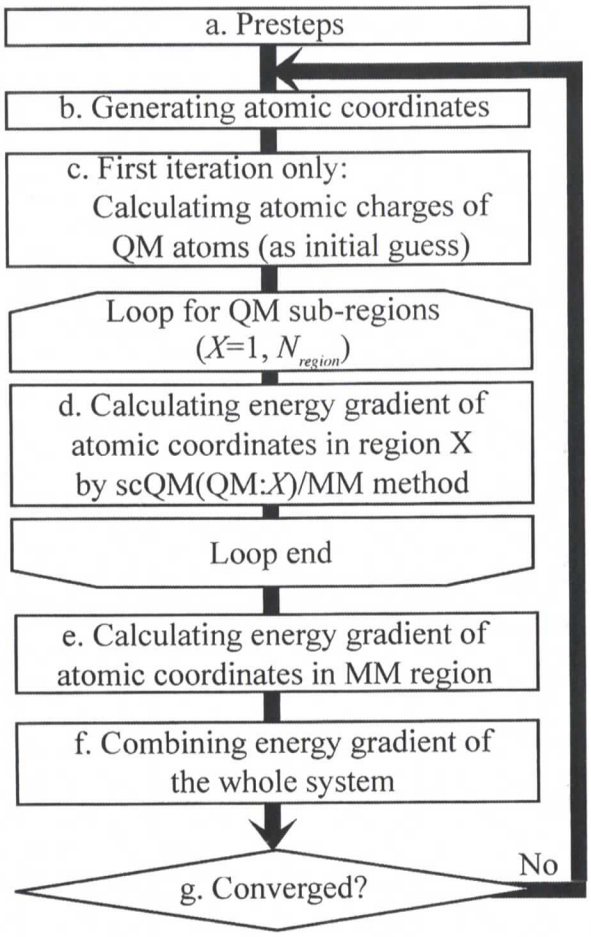

Figure 2. A flow chart of the mcQM/MM calculation for the geometry optimization.

energy gradient of QM subregions. In the final step, the energy derivatives are calculated for atoms in the genuine MM region.

Computational procedures of the geometry optimization are briefly illustrated in a flow chart (see Fig. 2). In the first iteration, atomic charges of all QM atoms are calculated. They are used for the initial guess to be used in the next step (step c). Loop for the QM subregions runs until scQM/MM energy gradient calculations are performed for total N_{region} QM subregions (step d). In each X, subregion X is treated by quantum mechanics, and the rest of the QM subregions are treated as a part of the MM region (scQM(QM:X)/MM calculation). We did not perform micro-iteration among the QM subregions, and therefore, self-consistency of the electron densities was not achieved within the step d. This was intended not to increase number of the QM calculations. The atomic charges computed in this step were used in the gradient calculations with the next trial geometry. In step e, the energy gradients with respect to the MM atom coordinates are calculated. Now, we have the energy gradients for all of the atomic coordinates. They are combined together in step f, and the convergence of the optimization is evaluated in step g. In the present program, maximum force, root mean square (rms) force, maximum displacement, and rms displacement are used for the convergence criteria. If not satisfied, a new set of atomic coordinates is generated in step b, and the energy gradients are calculated in the same way until convergence. At the optimized structure, the calculated electron density of the QM subregions also converges and becomes self-consistent. The present program was developed within the Gaussian 03 package.[26]

Results and Discussion

In the present approach, it is desirable for the mcQM/MM method to reproduce the energy gradient obtained by the scQM/MM method. Dividing the QM region, however, introduces approximated treatment for the interactions between two QM subregions. It is necessary to evaluate how these approximations affect numerical results. As benchmark calculations, we chose two typical systems. The first one is a hydrogen-bonding system, water dimer (Fig. 3a). The second one is a van der Waals (vdW) complex, benzene dimer (Fig. 3b). The calculations were performed with MP2 method with D95(d)[27] basis sets. The structures at each reaction coordinate were optimized using the single QM region. Therefore, the error represents how the present approximation affects the intermolecular and intramolecular geometries.

Figure 4 shows distance dependence in maximum and rms errors in the gradient. The horizontal lines labeled as "max" and "rms" indicate the convergence criteria adopted in Gaussian 03 program[26] as "Loose" options (0.00250 and 0.00167 hartree/bohr for the max and rms, respectively). Errors within these criteria are expected not to cause significant differences in the optimized structures. The result for the water dimer is shown in Figure 4a. As the O—H distance becomes short, the error monotonically increases. If we adopt the Gaussian "Loose" criteria, a hydrogen-bonding complex having the O—H distance longer than 3.5 Å can be divided into two QM subregions. In the case of benzene dimer interacting with the vdW forces, the plane-to-plane distance of 3.5 Å is also a reasonable boundary. In the case of water, the error in the energy gradient significantly increases at 2.5 Å. In the case of benzene dimer, the rapid increase occurs at 3.0 Å. Since similar behaviors are observed in the overlap integrals between the monomers, the error is related to the orbital overlaps between the two wave functions. In benzene dimer, the overlap between carbon $2p_\pi$ orbitals suddenly increases around 2.5–3.0 Å. The H_{1s}-O_{2p} overlap in water dimer increases around 2.0–2.5 Å, which is around 0.5 Å shorter in the intermolecular distance. We also note that the error in the water case decreases very slowly with the distance. In a hydrogen-bonding system such as water dimer, the Coulomb interaction itself is much larger than those in the vdW complex such as benzene dimer.

The accuracy of the mcQM/MM energy gradient was also examined in two cases: acetone molecules in water cluster (Figs. 3c and 3d) and amino acid residues in crambin protein (Fig. 3e). They represent two of the most popular applications of the QM/MM method. B3LYP[29,30] and CIS calculations were performed for the ground and excited states, respectively. The 6-31G(d,p)[31,32] basis sets were employed. AMBER and TIP3P[28]

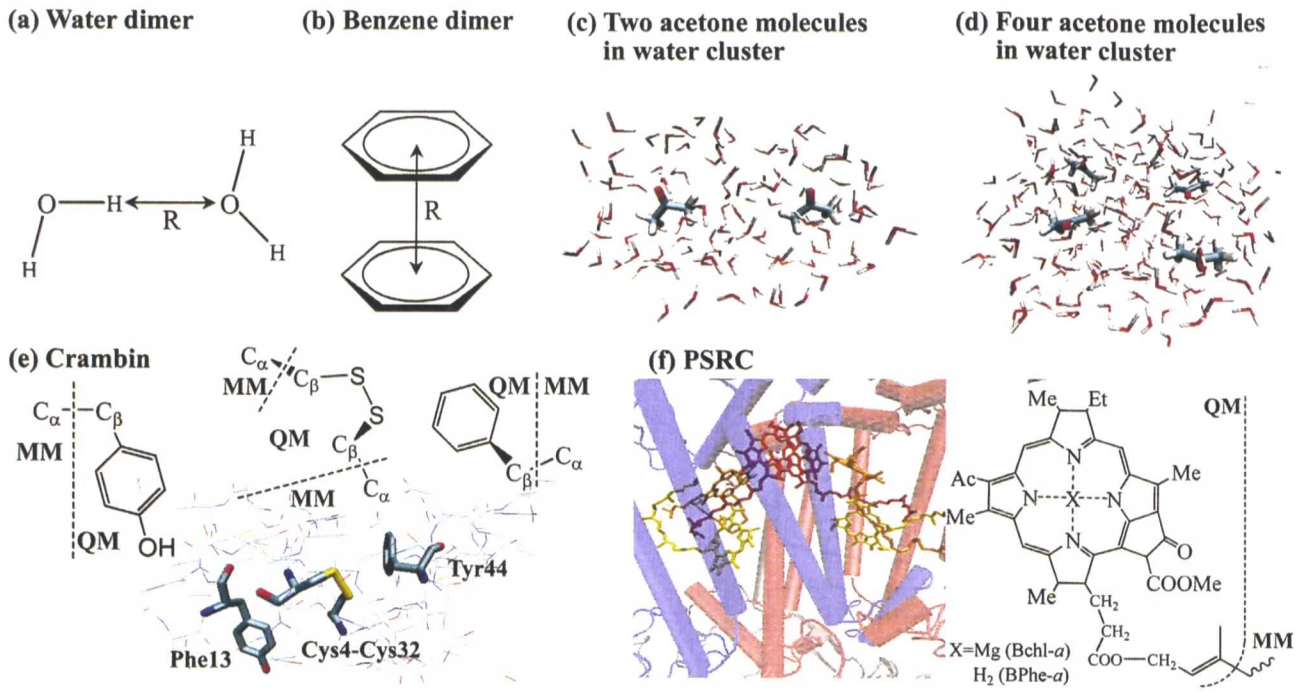

Figure 3. Sample molecules for test calculations. (a) Water dimer, (b) benzene dimer, (c) two acetone molecules in 100 water molecules, (d) four acetone molecules in 200 water molecules, (e) crambin, (f) and photosynthetic reaction center of *Rhodobactor sphaeroides*.

force field were used for the MM part. ESP charges were used for eq. (3).

In the acetone/water systems, acetone molecules are treated together as single QM region in the scQM/MM case, and each acetone molecule defines the QM subregion in the mcQM/MM case. Since the distances between the acetones are around 6.0 Å, the errors in the energy gradient is very small as shown in Table 1. Maximum and rms errors are at most 0.0003 and 0.0002 hartree/bohr, respectively, which are smaller than the "normal" convergence criteria in Gaussian 03 (0.00045 and 0.00030 hartree/bohr for max and rms errors, respectively).

In the case of crambin, we examined two different QM regions: the first one is a S—S bond (Cys4-Cys32) and Phe13, and the second one is the S—S bond and Tyr44 (see Fig. 3e). The closest distance between the S—S bond and Phe13 is 3.6 Å (S—C distance), and the S—S axis and the benzene ring lie in parallel as shown in Figure 3e. On the other hand, the closest distance between the S—S bond and Tyr44 is 4.0 Å (H—H distance). The S—C distance is 6.0 Å, and the S—S bond locates on the molecular plane of the phenyl group. In the Phe13 case, the max and rms errors were 0.0089 and 0.0023 hartree/bohr, respectively, as shown in Table 2. These errors are larger than the Gaussian Loose criteria. As we show later, such errors in the energy gradient does not significantly affect the optimized geometry. On the other hand, the errors in the tyr44 case are 0.0003 and 0.00017 hartree/bohr, respectively, which are smaller than the Gaussian normal convergence criteria.

Next, we show the convergence profile of the mcQM/MM geometry optimizations of acetone/water system and amino-acid residues in crambin. In Figure 5, maximum force in each step is compared with those of the scQM/MM optimizations for the four acetones/water system. Geometry convergence of the mcQM/MM optimization is worse than the scQM/MM one. The mcQM/MM optimization with DFT (B3LYP) required 121 steps, while the scQM/MM required 90 steps. The similar behavior was observed, when MP2 method was used for the electronic structure calculations in the QM regions. Figure 6 shows the convergence behavior in the crambin case. Number of steps in the scQM/MM calculation is also smaller than the mc QM/MM case.

In the mcQM/MM calculations, some of the QM subregions were treated as a part of the MM region, when the QM subregion does not include the atoms of which the energy derivative is calculated. In such a case, the electron density is replaced by ESP charges. The convergence of the atomic charges may be one possibility to be a determinant of the geometry convergence. In Figure 7, the rms changes in the atomic charges are plotted against the step number. During the optimization steps, the amount of fluctuation in the atomic charges was around 0.0005–0.0010 in the case of four acetone/water system in the ground state. There is no remarkable difference between the mc and scQM/MM results, which indicates that the fluctuation does not significantly affect the number of steps until convergence.

In Figure 8, the mcQM/MM optimized structures (red) are superimposed with the scQM/MM ones (blue). As seen in this figure, the red and blue structures in acetone/water and crambin (Cys4Cys32-Phe13 case) overlap very well. Although the error in the gradient was relatively large in crambin, the optimized

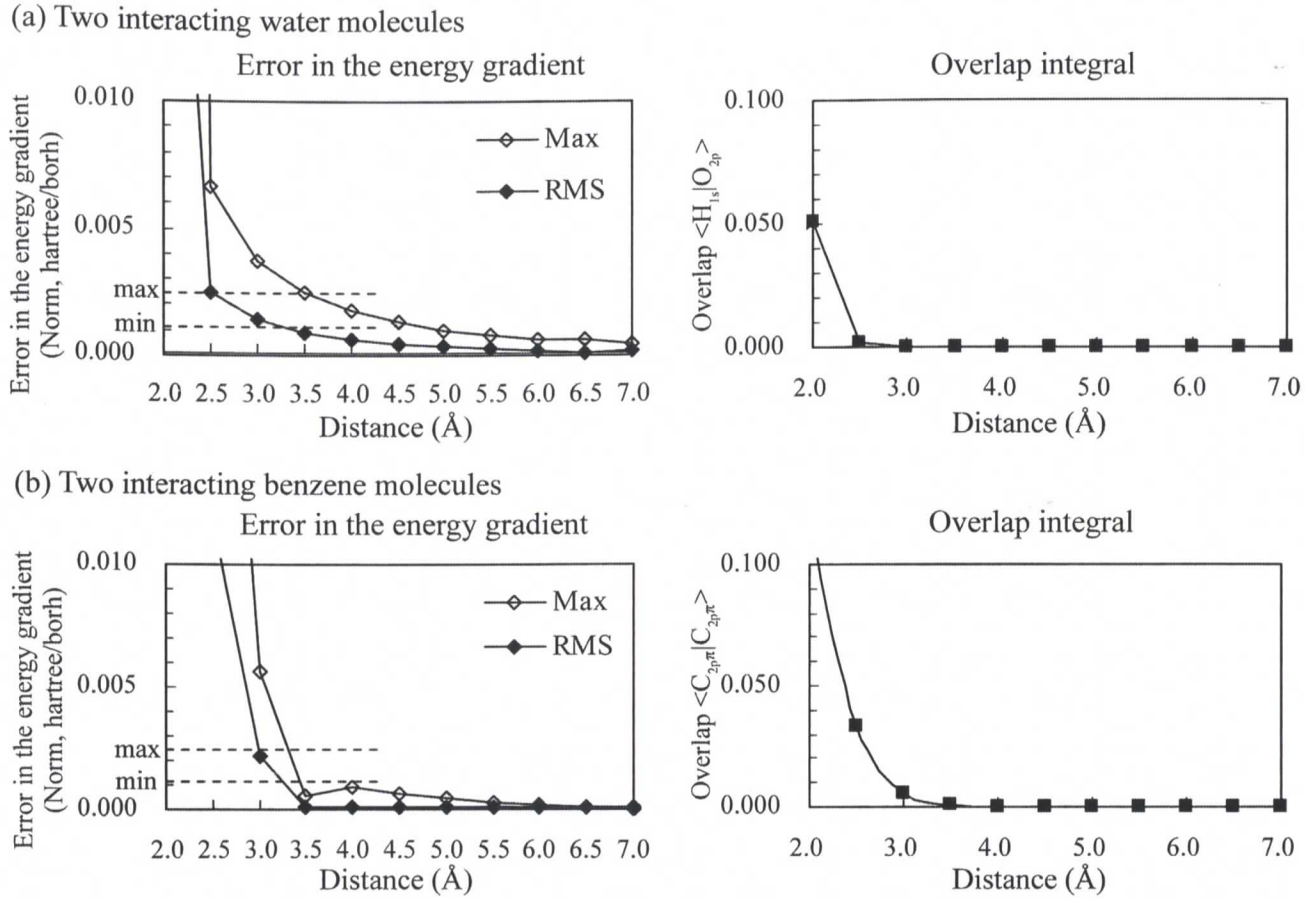

Figure 4. Intermolecule-distance dependence in the error of the energy gradient calculated by the multicore approach: (a) Two waters and (b) two benzene molecules in the ground states. Overlap integrals between the monomers are also compared.

Table 1. Error in the Energy Gradient (Norm, in hartree/bohr unit): B3LYP/6-31G(d,p):TIP3P Calculations were Performed for the Ground State of the Acetone/Water Systems.

(a) Two acetone molecules			(b) Four acetone molecules in a water molecule				
	Error in energy gradient (norm, hartree/bohr)			Error in energy gradient (norm, hartree/bohr)			
Atoms	Acetone 1	Acetone 2	Atoms	Acetone 1	Acetone 2	Acetone 3	Acetone 4
1 C	0.000198	0.000182	1 C	0.000261	0.000230	0.000341	0.000315
2 O	0.000140	0.000139	2 O	0.000278	0.000237	0.000192	0.000188
3 C	0.000097	0.000182	3 C	0.000161	0.000153	0.000322	0.000219
4 H	0.000029	0.000051	4 H	0.000044	0.000053	0.000141	0.000058
5 H	0.000017	0.000054	5 H	0.000034	0.000052	0.000104	0.000093
6 H	0.000029	0.000056	6 H	0.000052	0.000038	0.000060	0.000045
7 C	0.000197	0.000079	7 C	0.000168	0.000136	0.000301	0.000291
8 H	0.000055	0.000025	8 H	0.000061	0.000051	0.000117	0.000133
9 H	0.000055	0.000022	9 H	0.000053	0.000037	0.000068	0.000077
10 H	0.000057	0.000016	10 H	0.000036	0.000029	0.000091	0.000069
RMS	*0.000109*	*0.000101*		*0.000146*	*0.000128*	*0.000202*	*0.000176*

Table 2. Error in the Energy Gradient (Norm, in hartree/bohr unit): B3LYP/6-31G(d,p):AMBER Calculations were Performed for the Ground State of Crambin.

(Cys4-Cys32)-(Phe13)				(Cys4-Cys32)-(Tyr44)			
Cys4-Cys32		Phe13		Cys4-Cys32		Tyr44	
Atoms	Error	Atoms	Error	Atoms	Error	Atoms	Error
1 C	0.001886	1 C	0.008918	1 C	0.000026	1 C	0.000252
2 H	0.000037	2 H	0.000550	2 H	0.000072	2 H	0.000047
3 H	0.000047	3 H	0.000382	3 H	0.000006	3 H	0.000078
4 H	0.000458	4 H	0.001753	4 H	0.000075	4 H	0.000046
5 S	0.001569	5 C	0.000267	5 S	0.000082	5 C	0.000161
6 S	0.003577	6 C	0.000579	6 S	0.000075	6 C	0.000061
7 C	0.005432	7 C	0.001556	7 C	0.000182	7 C	0.000291
8 H	0.000334	8 C	0.000111	8 H	0.000045	8 C	0.000092
9 H	0.000150	9 C	0.000843	9 H	0.000015	9 C	0.000156
10 H	0.000844	10 C	0.000244	10 H	0.000006	10 C	0.000126
		11 H	0.000186			11 H	0.000042
		12 H	0.001171			12 H	0.000247
		13 H	0.000104			13 H	0.000036
		14 H	0.000116			14 H	0.000070
		15 H	0.000104			15 O	0.000075
						16 H	0.000060
RMS	*0.002222*		*0.002424*		*0.000140*		*0.000194*

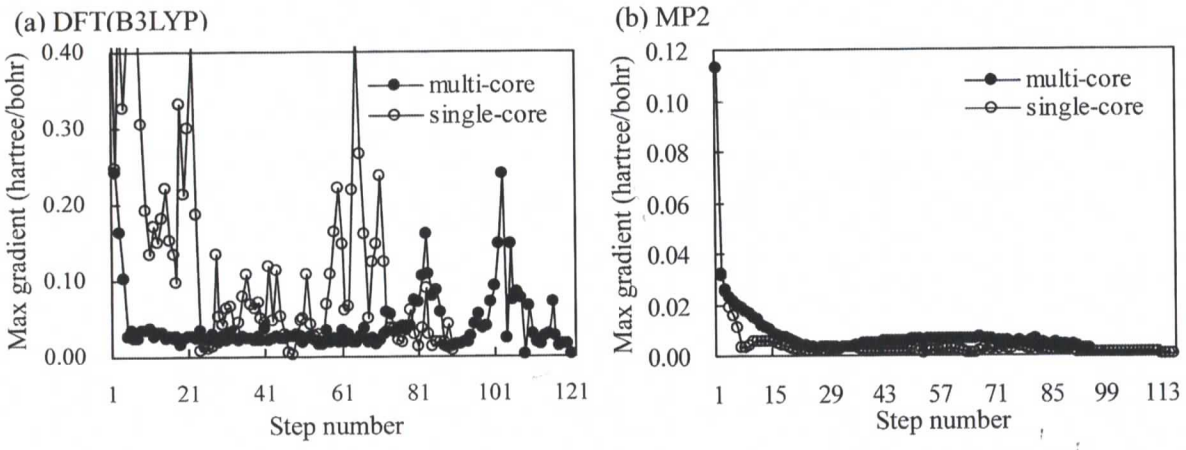

Figure 5. Maximum energy gradient (hartree/bohr) in each step of the optimization calculated by (a) DFT(B3LYP) and (b) MP2 methods. Calculations were performed for four acetone molecules in water cluster in the ground state.

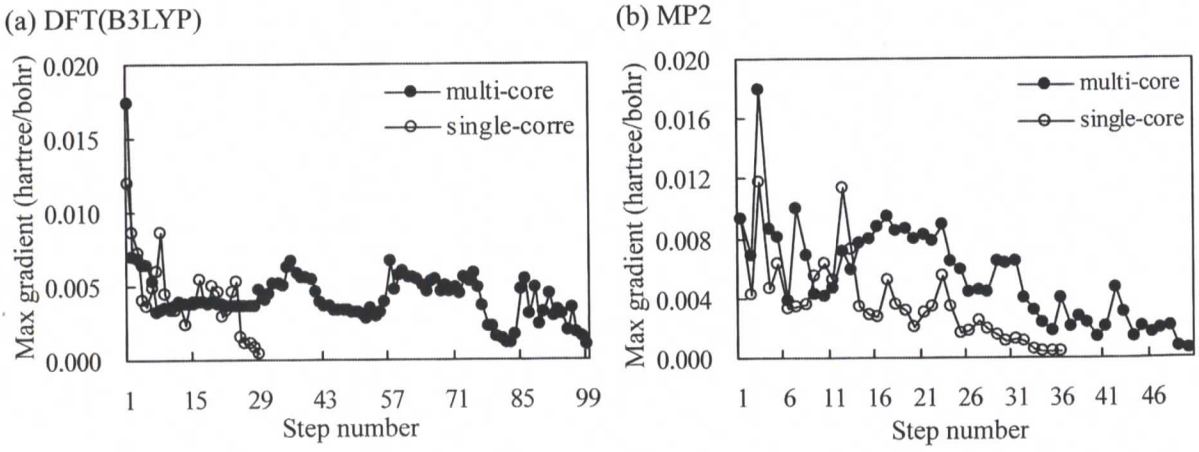

Figure 6. Maximum energy gradient (hartree/bohr) in each step of the optimization calculated by (a) DFT(B3LYP) and (b) MP2 methods. Calculations were performed for crambin in the ground state.

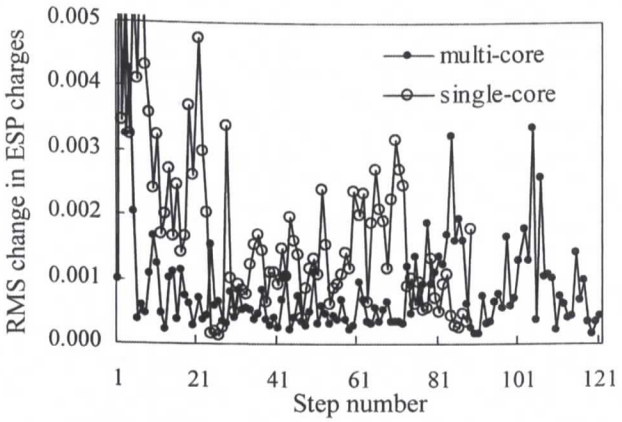

Figure 7. RMS Change in the ESP charges in each step of the optimization. DFT(B3LYP) calculations for four acetone molecules in water cluster.

structures obtained by the mc and scQM/MM calculations are very similar to each other. In Figure 8d, the mcQM/MM structure of the ground state of the photosynthetic reaction center is compared with the X-ray structure. We could find minor modifications in the chromophore structures by quantum mechanical calculations. The averaged changes in the interchromophore distances were around 0.07 Å.

Next, the structural relaxation energy in the geometry optimization was compared for acetone/water systems and crambin. We checked several states including the ground, excited, and charge-transfer states. The relaxation energy was defined as the energy difference between the initial and final structures. In the mcQM/MM calculation, the charge-transfer between acetone molecules is represented by the pair of acetones in ionized and anionized states. On the other hand, the CT state had to be described as one of the excited states in the scQM/MM calculations. The results are summarized in Table 3. The rms errors in the mcQM/MM optimizations are 2.4, 3.7, and 1.2 kcal/mol in

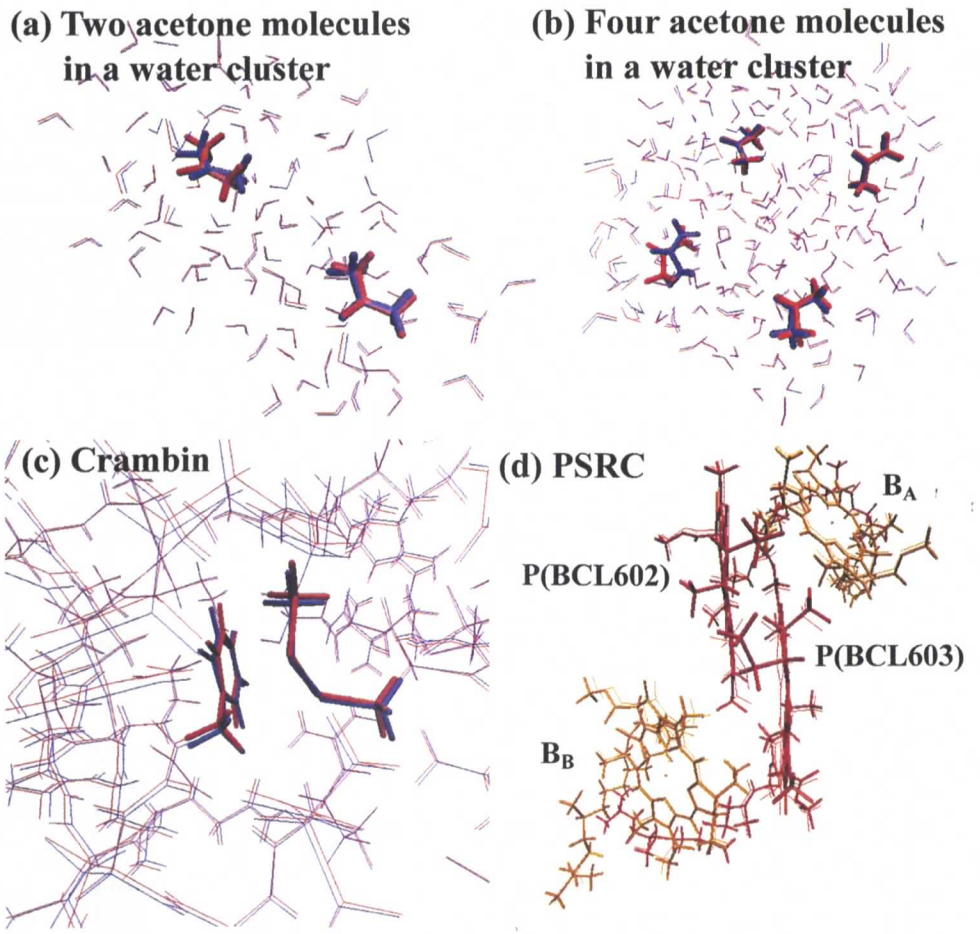

Figure 8. (a–c) Superposition of the structures optimized by the sc and mcQM/MM calculations. The structures shown in blue and red were obtained by the single-core and multicore approaches, respectively. (d) Superposition of the mcQM/MM (thick lines) and X-ray structures (thin lines) of the photosynthetic reaction center protein. Only P, B_A, and B_B are shown for the simplicity.

Table 3. Comparison of the Structural Relaxation Energy (kcal/mol).

Target systems and electronic states	mc[a]	sc[b]	Δ[c]
(a) Two acetone molecules cluster			
Ground state (B3LYP)	38.3	37.5	0.8
Ground state (MP2)	42.8	39.7	3.1
Locally excited state	47.0	45.9	1.1
Charge-transfer state	48.7	44.6	4.1
(b) Four acetone molecules cluster			
Ground state (B3LYP)	53.9	57.3	−3.4
Ground state (MP2)	61.4	60.7	0.7
Locally excited state	58.2	56.6	1.6
Charge-transfer state	61.3	67.7	−6.4
(d) Crambin			
Ground state (B3LYP)	13.8	15.9	−2.1
Ground state (MP2)	14.2	14.7	−0.5
Locally excited state	24.1	23.6	0.5
Charge-transfer state	48.1	47.2	0.9
(d) Photosynthesis reaction center			
Locally excited state	26.4	–	
Charge-transfer state	36.1	–	

Single point energy calculation using structures optimized by multicore and single-core QM/MM calculations. Energies of the initial structures were used as the reference.
[a]mcQM/MM geometry.
[b]scQM/MM geometry.
[c]Δ = "mc" − "sc."

two acetones/water, four acetones/water, and crambin, respectively. The error is relatively larger in the solution system than in the protein, which would be related to the flexibility of the system.

Finally, we discuss the computing time and efficiency of mcQM/MM geometry optimization. In Table 4, total number of steps required for the convergence, CPU time per step, and efficiency ratio were compared between mcQM/MM and scQM/MM optimizations. The efficiency ratio was defined as the ratio of CPU time per step as shown in Table 4. If the ratio is larger, the mcQM/MM is more efficient than scQM/MM in the CPU time. In general, number of steps in mcQM/MM is larger than that in scQM/MM method. One of the reasons would be the lack of the self-consistency in the electron density during the optimization as mentioned in the Method section. The error in the gradient could increase number of iteration to achieve the convergence. However, CPU time in mcQM/MM is usually smaller than that in scQM/MM method, even in such small test examples. The efficiency ratio becomes significantly large, when the electronic structure method for the QM region needs longer CPU time and when the size of the QM region becomes larger. In the two acetone/water case, the ratios are similar in the ground, excited, and charge-transfer states. In the four acetone/water case, the ratios for the excited and charge-transfer states becomes more than 10 times larger than the two-acetone case. We performed the mcQM/MM calculations for the excited and charge-transfer states of the photosynthetic reaction center of a purple bacteria, *Rhodobactor sphaeroide*. B3LYP[29,30] and CIS calculations were performed for the ground and excited states, respectively. The 6-31G[31] basis sets were employed. AMBER and TIP3P force field were used for the MM part. Special pair was treated as one QM subregion, while each bacteriochlorophyll and bacteriopheophytin was treated as single QM subregion. It was impossible to perform scQM/MM geometry optimization even for the ground state. The CPU time for the mcQM/MM optimization is around 18 and 13 h/steps in the local excited and charge-transfer states, respectively. The most time-

Table 4. Total Number of Steps, CPU Time per Step, and Efficiency Ratio (Single/Multi) in the Geometry Optimizations Using Single- and Multicore QM/MM Calculations.

	Total number of steps		CPU time per step (s)		
Target systems and electronic states	Multicore	Single-core	Multicore	Single-core	Ratio (single/multi)
(a) Two acetone molecules in water cluster					
Ground state (B3LYP)	120	17	152.1	176.5	1.16
Ground state (MP2)	70	85	138.5	249.9	1.80
Locally excited state	50	20	201.0	312.2	1.55
Charge-transfer state	132	109	225.9	318.0	1.41
(b) Four acetone molecules in water cluster					
Ground state (B3LYP)	121	45	316.6	425.1	1.34
Ground state (MP2)	95	116	459.8	1134.3	2.47
Locally excited state	72	29	272.7	5414.1	19.85
Charge-transfer state	109	190	354.9	3873.4	10.92
(d) Crambin					
Ground state (B3LYP)	99	29	269.5	458.9	1.70
Ground state (MP2)	50	36	266.9	1506.6	5.64
Locally excited state	88	79	382.5	2037.7	5.33
Charge-transfer state	65	22	449.1	1414.4	3.15
(d) Photosynthesis reaction center					
Locally excited state	27	–	65022.3	–	–
Charge-transfer state	90	–	47280.6	–	–

consuming step was to calculate the energy gradient of the special pair in the excited states. The optimized structures obtained in this study will be used for analyzing the origin of the unidirectional electron transfer in terms of the potential energy profile.

Conclusions

In this study, we proposed a feasible approach to optimize geometry of very large molecular systems. The present approach, mcQM/MM, is a kind of QM/MM method. Instead of treating a large QM region at once, the QM region is treated as the sum of the QM subregions. The scQM/MM calculations are performed for each QM subregion. 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.

In the mcQM/MM calculation, quantum mechanical intermolecular (inter-QM subregions) interactions are approximated to those used in QM/MM method (electrostatic, through-bond, and vdW interactions). The accuracy of the mcQM/MM energy gradient was examined by comparing with the scQM/MM one. We concluded that the mcQM/MM energy gradient reproduced the scQM/MM one in the intermolecular distances more than 3.5–4.0 Å. The error in the energy gradient rapidly decreases in a vdW complex, benzene dimer, when the distance became longer. On the other hand, in the case of a hydrogen-bonding complex, water dimer, the error decreases very slowly at a large intermolecular distance. Regarding the optimized structure in water solution and in protein, the mcQM/MM structures are similar to the scQM/MM ones. The error in the optimized structure in water solution is larger than that in protein as seen in Figure 8, which indicates that the structural flexibility relates the applicability of the mcQM/MM approach. In addition, we clearly found the advantage of the mcQM/MM geometry optimization in the computation effort. When the size of the QM region is bigger and the electronic structure method used is computationally demanding, the computing time of the mcQM/MM optimization becomes much smaller than the scQM/MM one.

Acknowledgment

A portion of the computations was carried out at RCCS (Okazaki, Japan).

References

1. Merz, K. M.; Stanton, R. V. In Encyclopedia of Computational Chemistry; Schleyer, P. v. R., Ed.; Wiley: Chichester, 1998; pp. 2330–2343.
2. Tomasi, J.; Pomelli, C. S. In Encyclopedia of Computational Chemistry; Schleyer, P. v. R., Ed.; Wiley: Chichester, 1998; pp. 2343–2350.
3. Hasegawa, J.; Fujimoto, K.; Swerts, B.; Miyahara, T.; Nakatsuji, H. J Comput Chem 2007, 28, 2443.
4. Fujimoto, K.; Hasegawa, J.; Hayashi, S.; Nakatsuji, H. Chem Phys Lett 2006, 432, 252.
5. Fujimoto, K.; Hayashi, S.; Hasegawa, J.; Nakatsuji, H. J Chem Theory Comput 2007, 3, 605.
6. Deisenhofer, J.; Michel, H. Science 1989, 245, 1463.
7. Roth, M.; Lewit-Bentley, A.; Michel, H.; Deisenhofer, J.; Huber, R. Nature 1989, 340, 659.
8. Allen, J. P.; Feher, G.; Yeates, T. O.; Komiya, H.; Rees, D. C. Proc Natl Acad Sci USA 1987, 84, 5730.
9. Stowell, M. H. B.; McPhillips, T. M.; Rees, D. C.; Soltis, S. M.; Abresch, E.; Feher, G. Science 1997, 276, 812.
10. Yang, W. Phys Rev Lett 1991, 66, 1438.
11. Yang, W.; Lee, T.-S. J Chem Phys 1995, 103, 5674.
12. Dixon, S. L.; Merz, K. M. J Chem Phys 1997, 107, 879.
13. Akama, T.; Kobayashi, M.; Nakai, H. J Comput Chem 2007, 28, 2003.
14. Kitaura, K.; Ikeo, E.; Asada, T.; Nakano, T.; Uebayasi, M. Chem Phys Lett 1999, 313, 701.
15. Kitaura, K.; Sugiki, S.; Nakano, T.; Komeiji, Y.; Uebayasi, M. Chem Phys Lett 2001, 336, 163.
16. Morita, S.; Sakai, S. J Comput Chem 2001, 22, 1107.
17. Li, W.; Li, S. J Chem Phys 2004, 121, 6649.
18. Kashiwagi, H.; Iwai, H.; Tokieda, K.; Era, M.; Sumita, T.; Yoshihiro, T.; Sato, F. Mol Phys 2003, 101, 81.
19. Singh, B. H.; Kollman, P. A. J Comput Chem 1984, 5, 129.
20. Besler, B. H.; Merz, K. M.; Kollman, P. A. J Comput Chem 1990, 11, 431.
21. Mulliken, R. S. J Chem Phys 1955, 23, 1833.
22. Mulliken, R. S. J Chem Phys 1955, 23, 1841.
23. Foster, J. P.; Weihold, F. J Am Chem Soc 1980, 102, 7211.
24. Reed, A. E.; Weinstock, R. B.; Weihold, F. J Chem Phys 1985, 83, 735.
25. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. J Am Chem Soc 1995, 117, 5179.
26. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, M.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03; Gaussian, Inc.: Pittsburgh, 2003.
27. Dunning, T. H.; Hay, P. J.In Modern Theoretical Chemistry;Schaefer, H. F., Ed.; Plenum: New York, 1976; pp. 1–28.
28. Jorgensen, W. L.; Chandreskhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. J Chem Phys 1982, 79, 926.
29. Lee, C.; Yang, W.; Parr, R. G. Phys Rev B 1988, 37, 785.
30. Becke, A. D. J Chem Phys 1993, 98, 5648.
31. Hehre, W. J.; Ditchfield, R.; Pople, J. A. J Chem Phys 1972, 56, 2257.
32. Hariharan, P. C.; Pople, J. A. Theor Chim Acta 1973, 28, 213.