New Sampling Scheme for Combined Quantum Mechanical and Molecular Mechanical Simulation

Tetsuya Sakata, Yukio Kawashima, and Haruyuki Nakano

Abstract
Solvent models and sampling schemes for obtaining average physical properties for molecules in solution utilizing combined quantum mechanical / molecular mechanical (QM/MM) method are reviewed and a new sampling scheme of ours is presented. In the new scheme, we sample the structure of the solute molecule and solvent molecules separately. First, we execute a QM/MM MD simulation, where we sample the solute molecule in solution. Next, we choose random solute structures from this simulation and perform classical MD simulation for each chosen solute structure with its geometry fixed. This new scheme allows us to sample the solute molecule quantum mechanically and sample many solvent structures. This scheme is combined with the generalized-multiconfigurational perturbation theory and applied to study the absorption spectra of Coumarin 120 in water. We succeeded in constructing the absorption spectra and in realizing the red shift of the absorption spectra found in polar solvents. We found that the umbrella-like open-close motion of the amino group of the coumarin plays an important role in the broadening of the spectra, whereas the red shift of the peak was caused by the solvation structure of the amino group.

Keywords
QM/MM; excited states; molecular dynamics simulation; GMC-QDPT; C120; 7-aminocoumarins

1. Background
Quantum mechanical calculations of electronic excited states enable assignments of observed absorption spectra and prediction of excited states that cannot be observed. Recent calculations succeed to calculate the vertical excitation energies of molecules in the gas phase within 0.2 eV, which allows the precise assignment of the absorption spectra. However, most of the molecules of interest exist in solution or in biochemical environment. Thus, the calculation of the electronic excited states of molecules in a real system is a new goal to achieve in quantum chemistry. To compute a molecule in solution, we need to compute the solvent molecule with many solvent molecules. Despite the increase of computer resources, quantum mechanical calculations including the surrounding solvent molecules are still computationally unrealistic. The inclusion of the solvent effect in electronic excited state calculations has attracted a great deal of attention in quantum chemistry for a long time. Mochizuki et al. have employed the multilayer fragment molecular orbital method with CIS methods to electronic excited states of microsolvated formaldehyde. Hirata et al. applied the pair-interaction method with the EOM-CCSD method to the formaldehyde with 81 water molecules. These two approaches seek possibilities for performing full quantum calculations of the entire system including solute and solvent molecules; however, these calculations are limited to snapshot calculations. In order to calculate the electronic excited states in solution, the solvent structures must be obtained stochastically, i.e., the electronic excited states calculation must be averaged over many snapshots of solvent structures. However, not enough quantum mechanical calculations of the entire system can be computed to obtain converged averaged properties. Moreover, quantitative excited state calculations are difficult compared to ground state calculations because of their multiconfigurational character, which leads to an increase in computational effort. Thus, employing an inexpensive model to describe the solvent is the key to compute quantum mechanical calculation of molecules in solution.

2. Solvent Models for Quantum Mechanical Calculation
The first model introduced was the dielectric continuum model. This model mimics the solvent as a dielectric continuum mean field with an inner cavity, in which the solute is placed. This model is commonly used because of the small computational cost. The self-consistent reaction field (SCRF) method is the oldest dielectric continuum model with a spherical cavity. The simplicity of handling of this method allowed many applications; however, this model is too crude for describing the solvent effect of complex molecules. The polarizable continuum model...
improved the SCRF method by introducing the molecular shape for the cavity. Moreover, in the framework of nonequilibrium regime, PCM can be applied to the electronic excited states. The drawback of these models is that they cannot describe the microscopic structure of the solute-solvent interaction.

Another model was introduced by combining quantum mechanical calculation and liquid theories. The most widely used method is the RISM-SCF method developed by Ten-no et al., combining quantum mechanical calculation and the reference interaction site-model (RISM) theory. The solvent effect is introduced as a mean field by describing the solute-solvent interaction in terms of the site-site radial distribution function. This model, which considers the microscopic solute-solvent interaction, overcomes the drawbacks of the dielectric continuum model. However, the use of radial distribution functions limits the application to small solute molecules. Recently, Sato et al. developed the 3D-RISM-SCF method, which generalizes the RISM integral equation in three dimensions, and Yokogawa et al. developed the RISM-SCF-SEDD method, which introduces the spatial electron density distribution to improve RISM-SCF. Yoshida et al. developed the cost consuming but very accurate MOZ-SCF method.

The final model is the combined quantum mechanical/molecular mechanics method. Since the pioneering work of Warshel and Levitt, QM/MM methods have been successful in tackling the electronic structure calculation of large scale systems. Especially in the last decade, QM/MM methods have been applied to various problems such as chemical reactions in enzyme or solvent environment. In QM/MM methods, the calculated molecular system is divided into at least two different subsystems: a subsystem where the quantum effect that must be included is treated by QM, and the other, which includes the environmental effect that is treated by MM. Efforts have been made to obtain the average properties of the electronic excited states by using QM/MM methods and treating solvents by MM. Molecular dynamics (MD) simulation combined with QM/MM methods is a powerful tool for studying the electronic structure of molecules in solution. The solute and solvent molecules are sampled simultaneously; the former is sampled by quantum mechanical calculation and the latter by classical simulation. Both Car-Parrinello and Born-Oppenheimer MD employing QM/MM methods were applied to study the excited state structure in solvents; however, it is still difficult to sample enough solvent structures to obtain converged average properties.

In these methods, the excitation energy corresponding to the peak of the spectrum is obtained, while the entire spectra cannot be constructed. The solvent effect leads to a red or blue shift, and broadening of the spectra. Thus, to fully understand the effect of the environment, construction of the entire absorption spectra is necessary.

3. Sampling Scheme for QM/MM Calculation

In our previous work, we executed a classical Monte Carlo simulation and obtained enough snapshots to obtain average solvent properties. We next chose 400 configurations and performed quantum mechanical calculation for those configurations, which we used to calculate the average physical properties. We applied this scheme to the first excited state of formaldehyde and calculated the absorption spectrum in water. We succeeded in constructing the absorption spectrum and in obtaining the converged absorption spectrum peak and blue shift in water. The use of full classical simulation for sampling the solvent structure enables us to obtain the average property over many solvent structures. Many studies follow a similar scheme. Sanchez et al. use classical molecular dynamics simulation to calculate the averaged value of the solvent electrostatic potential, and perform quantum mechanical calculation using this averaged mean field. This reduces the number of quantum mechanical calculations; however, it cannot construct the entire spectrum. Pulay et al. recently developed a similar method, which uses expansion of electric potential generated by simulation. In previous sampling schemes, though the solvent structures are well sampled, the solute molecules are not sampled. In classical simulation, the solute molecules are fixed. This is not realistic, in particular for large solute molecules. Moreover, the geometry affects the absorption spectra, and thus quantum mechanical treatment is necessary to sample the solute structures.
4. Our New Sampling Scheme

To overcome the drawback of the previous method, we introduce a new efficient sampling scheme to obtain average physical properties. The new sampling scheme is described in Fig. 1. In our new sampling scheme, we sample the structure of the solute and solvent molecules separately. First, we execute a QM/MM MD simulation, where we sample the solute molecule in solution. Next, we choose random solute structures from this simulation and perform classical MD simulation for each chosen solute structure with its geometry fixed. In this classical simulation, we sample the solvent structures. Finally, we compute quantum chemical calculation with the randomly chosen solvent structures sampled from this classical simulation and the absorption spectrum is constructed. This sampling scheme is a combination of QM/MM MD simulation and the previous sampling scheme. This new scheme allows us to sample the solute molecule quantum mechanically and sample many solvent structures.

5. General Multiconfigurational Quasi-degenerate Perturbation Theory

We combine this new scheme with a multireference perturbation theory developed by us. The multireference perturbation theory based on multiconfigurational reference functions has become a practical tool for studying the electronic structures of low-lying excited states. The multireference Møller-Plesset perturbation theory \[ \text{MC-QDPT} \] and the multiconfigurational quasi-degenerate perturbation theory (MC-QDPT) \[ \text{GMC-QDPT} \] succeeded in describing the excited states of π-conjugated systems that take into account both static and dynamic electron correlation \[ \text{GMC-QDPT} \]. These methods include valence π and π* orbitals in the active space of the reference CASSCF wavefunction to perform calculations of the π→π* excited states. However, the use of the CASSCF wavefunction limits the application of these methods to large π-conjugated systems since the active space dimension grows enormously with the number of active orbitals and electrons. To avoid this drawback, a perturbation theory using general multiconfiguration (MC) SCF wavefunctions as reference functions (GMC-QDPT and GMC-PT) was developed \[ \text{GMC-QDPT} \]. This method enables us to calculate using larger active space, which leads to a broader application.

6. Application

We applied our new scheme to the excited state of Coumarin 120 (C120, 7-amino-4-methyl-1,2-benzopyrone, Fig. 2), which is the basic molecule in the 7-aminocoumarin family. The compounds 7-aminocoumarins, or 4-methyl-7-dielalamino-coumarins, are the most featured dyes among the coumarin family. The substitution at the 7-position with an electron-donating group enhances the fluorescence of the dye, which leads to the wide application in blue-green laser dyes and fluorescence probes. Furthermore, 7-aminocoumarin dyes are applied to study solvatochromic properties since the large Stokes’ shifts of these molecules are very sensitive to the polarity and viscosity of the surrounding solvent environment. Hence, the photophysics of these dyes has been studied intensively. In spite of the large amount of research and wide applications of 7-aminocoumarins, further investigation of the excited states is essential. The nonradiative deactivation mechanism of 7-aminocoumarins has been debated for the past two decades. The first model describes the nonradiative deactivation process as the 7-aminocoumarin forming the so-called twisted intramolecular charge transfer state from the S1 singlet excited state \[ \text{S1} \]. The second is the so-called open–closed umbrella-like motion mechanism \[ \text{GMC-PT} \]. This mechanism ascribes the internal conversion process to a structural change of the amino group from a planar N+-aromatic configuration (with sp² hybridization for the nitrogen atom) to a pyramidal N-aromatic configuration (with sp³ hybridization for the nitrogen atom). For coumarin 120 (C120) \[ \text{C120} \], it is known that the nonradiative deactivation process differs in polar and nonpolar solvents. Thus, it is essential to elucidate the solvent effect of the excited states. Here, we aim to elucidate the character of the first singlet excited state of C120 in water solution. Utilizing our new sampling scheme, we aim to construct the absorption spectra in water and understand the correlation.

---

Sampling Scheme for QM/MM
between the absorption spectra and solute and solvent structures.

![Figure 2 Skeletal Structure of C120](image)

6-1. Excited state calculation of C120 in gas phase

First, to check the accuracy of our calculation, we computed the excitation energy of the first singlet excited state of C120 in the gas phase utilizing the GMC-PT method. The first singlet excited state results from the HOMO → LUMO transition, which is a $\pi \rightarrow \pi^*$ excitation. The calculated excitation energy is 3.66 eV. This state has large oscillator strength, thus we have assigned this state to the first peak in the experimental spectrum. GMC-PT calculation result is slightly above the experimental value of 3.51 eV recorded in ethanol. Compared with the gas phase, absorption spectra in ethanol will shift to the red, hence GMC-PT calculation seems to be accurate. Geometry is optimized at MP2 level, and cc-pVDZ basis set is used as basis set for all calculations.

6-2. Excited state calculation of C120 in water

We applied QM/MM calculations to C120 in water, where the solute, C120, is treated quantum mechanically, and the solvent water molecules are treated molecular mechanically. First, we executed QM/MM MD simulation of the ground state of C120 in 417 water molecules with a spherical boundary for 40,000 steps, with 0.5 fs time step. C120 was treated at B3LYP level. We next choose 65 solute structures from QM/MM MD simulation and performed classical MD simulation for each chosen C120-water system fixing the geometry of the solute molecule for 1,000,000 steps, with 0.5 fs time step. Periodic boundary condition is applied to classical MD simulation utilizing the Particle Mesh-Ewald method for treating the long range Coulomb force with 808 water molecules. The leapfrog Verlet algorithm was used for integration of the equation of motion. Finally, we chose four C120-water snapshot structures from each classical MD simulation and computed the excitation energies of each snapshot structure. Thus, the total number of excited state calculations was 260. The GMC-PT method was used for the excited state calculations of the solute C120 molecule, whereas water molecules were treated molecular mechanically. Dunning’s cc-pVDZ basis set was used for all QM calculations, and CHARMM27 force field parameters and TIP3P water parameters were used for MM calculations.

6-3. QM/MM simulation results of C120 in water

We analyzed the QM/MM MD run and found that the motion of the amino group shows the largest deviation. The amino group moves between planar structure and pyramidal structure, the so-called open–closed umbrella-like motion.

We constructed the absorption spectra applying QM/MM calculation to obtained snapshots from simulation. The spectrum is a histogram of the calculated excitation energies of chosen snapshots. The constructed absorption spectrum illustrated in Fig. 3 ranges from 3.15 eV to 4.48 eV. The peak of this histogram is found at between 3.80 and 3.90 eV and agrees well with the average excitation energy of 3.86 eV. The spectrum is well described as a Gaussian distribution with the peak at the average excitation energy. The calculated excitation energy in gas phase is 3.89 eV. Our calculation obtained the red shift in water solution. Hereafter, we focus on this red shift of the absorption spectrum in water to discuss the solvent effect.
6-4. Decomposition of the solvent effect in the red shift of the Absorption Spectrum

In our sampling scheme, we sampled the solute structure and solvent structure, separately. This enabled us to decompose the solvent effect into two components. First is the solvent effect, which is caused by the geometry distortion of the solute from the gas phase structure. Second is the field effect of the solvent molecules. Here, we decomposed the red shift of the absorption spectrum into two components. We denote the shift from the former as $\Delta E_1$ and the latter as $\Delta E_2$. This decomposition allowed us to seek the origin of the shape of the spectrum, which cannot be done in the previous QM/MM sampling scheme. The average of $\Delta E_1$ and $\Delta E_2$ is 0.00 eV and 0.03 eV, respectively. Thus, we found that the red shift originates from the field effect of the solvent molecules, and the effect of the solvent molecules does not affect the spectrum. We depict the plots of $\Delta E_1$ and $\Delta E_2$ over the total shift $\Delta E$ in Fig. 4(a) and 4(b), respectively. The plot in Fig. 4(a) shows a large correlation between $\Delta E_1$ and $\Delta E$. Thus, the broadening of the spectrum comes from the deviation of the solute geometry. In Fig. 4(b), $\Delta E_2$ is mainly found in the upper region of the plot, which shows that the field effect leads to red shift, whereas the effect of the solute geometry leads to both red and blue shifts.

6-5. Analysis of the correlation between the red shift and solute-solvent structures

To elucidate the correlation between the red shift of the spectrum and the snapshot structures, we categorized the snapshot structures according to the value of the red shift. The use of snapshot structures is a major advantage in QM/MM simulation, where dielectric continuum models and liquid theory-based methods do not have detailed snapshot structure information. We categorized snapshots with $\Delta E_1$ red shift smaller than $-0.10$ eV as $\Delta E_1^{\text{small}}$, snapshots with $\Delta E_1$ red shift larger than $-0.10$ eV as $\Delta E_1^{\text{large}}$, and $\Delta E_1^{\text{medium}}$. The black line represents the RDF of all snapshots.
ΔEᵢ red shift larger than 0.10 eV as ΔEᵢ \text{large}_i, and snapshots with ΔEᵢ red shift between −0.10 and 0.10 eV as ΔEᵢ \text{medium}_i. Using these snapshots, we focused on the solute-solvent structure. The radial distribution functions between the nitrogen of the solute amino group and oxygen of water, and hydrogen of water are illustrated in Fig. 5(a) and (b), respectively. As the red shift ΔEᵢ becomes larger, the first peak of both functions grows. This shows that snapshots with large ΔEᵢ red shift have the tendency to form a pyramidal coumarin amino group (Fig. 6(a)), which constructs hydrogen bonds with solvents at the nitrogen atom of the amino group, whereas snapshots with small ΔEᵢ red shift have the tendency to form a planar amino group (Fig. 6(b)), which does not construct hydrogen bonds at the nitrogen atom. Our results suggest that the umbrella open-close motion, the largest motion found in QM/MM MD simulation, plays an important role in the broadening of the calculated absorption spectrum.

6-6. Analysis of two representative snapshots

To discuss the difference of the two solvent structures in detail, we selected two solute structures, the pyramidal amino group structure at step number 1080 and the planar structure at step number 1260, from QM/MM MD simulation. The structures depicted in Fig. 7. Fig. 7(a) and (b) corresponds to structure of step numbers 1080 and 1260, respectively. The calculated excitation energies and red shift are listed in Table 1. The excitation energies calculated with pyramidal structure show a large ΔEᵢ red shift, while there is no ΔEᵢ red shift found. On the other hand, the planar structure shows a larger ΔEᵢ red shift than ΔEᵢ. To understand the difference in the ΔEᵢ red shift between the two structures, we focus on the solute-solvent interactions. The Radial distribution function between the nitrogen of the amino group and the oxygen and hydrogen of water are illustrated in Fig. 8(a) and (b), respectively. The hydrogen bond from the amino nitrogen to the water hydrogen leads to the peak found in Fig. 8(b) for the pyramidal structure. The first peak for the planar structure is the hydrogen bond from the hydrogen of the amino group to the oxygen of water described in Fig. 6(b). The solvation structures contribute to the electrostatic effect of the electronic
structure of the solute. Thus, the radial distribution function indicates that hydrogen bond between nitrogen of the amino group and hydrogen of water leads to a large decrease in the ΔE₂ red shift, and the hydrogen bond found in the planar structure leads to a large ΔE₂ red shift. In Table 1, the dipole moments between the two structures are compared. The planar structure has a larger dipole moment than the pyramidal. This stabilizes the charge-transfer excited state, which leads to the red shift of the absorption spectrum. We here suggest that the solvation structure of the amino group of C120 plays a key role for the red shift.

7. Conclusion and Outlook

We developed a new sampling scheme to obtain average physical properties in solution for QM/MM calculations. We sampled the structure of the solute molecule and solvent molecules separately. First, we carried out a QM/MM MD simulation, where we sampled the solute molecule in solution. Next, we chose arbitrary solute structures from this simulation and performed classical MD simulation for each chosen solute structure with its geometry fixed. This new scheme allows the sampling of the solute molecule quantum mechanically and the sampling of many solvent structures. We applied this new scheme to the absorption spectrum of C120 in water. We succeeded in constructing the absorption spectrum and obtained the red shift of the absorption spectrum found in polar solvents. Taking advantage of our scheme, the solvent effect of the absorption spectrum was analyzed in three steps. First, we analyzed the red shift by decomposing the solvent effect of the absorption spectra into two components: the effect from the distortion of the solute molecule and the field effect from the solvent molecules. We next analyzed the solvent effect of the red shift by classifying snapshots into groups with large, medium, and small shifts and studied the correlation between the red shift and the solute-solvent structures. Finally, we chose representative solute-solvent structures and discussed the origin of the red shift in the absorption spectra. We suggest that the umbrella-like open-close motion of the amino group of the coumarin governs the broadening of the spectra, whereas the red shift of the peak is caused from the solvation structure of the amino group.

QM/MM MD simulation of the C120-water system in the excited state is now in process. We aim to elucidate the solvent effect in the excited state of C120. This leads to the understanding of the nonradiative deactivation process of the 7-aminocoumarins. It is possible to carry out the simulation in nonpolar solvents as well and study the nonradiative deactivation process, which differs from the process in polar solvents. Our new scheme is not limited to excitation energy calculations and various physical properties can also be obtained. Furthermore, our scheme can be also applied to quantum mechanical calculation in biochemical environment. Thus, in the near future, it is possible to obtain well-converged physical properties in biological systems.

Acknowledgements

YK was supported by the Program for Improvement of Research Environment for Young Researchers from Special Coordination Funds for Promoting Science and Technology (SCF), Japan. HN is grateful to CREST, Japan Science and Technology Agency (JST) for funding this research.
References