Theor Chem Acc (2006) 116: 43–50
DOI 10.1007/s00214-005-0008-x

REGULAR ARTICLE

Yingkai Zhang

Pseudobond ab initio QM/MM approach and its applications to enzyme reactions

Received: 29 March 2005 / Accepted: 19 May 2005 / Published online: 8 October 2005

Abstract This perspective article mainly focuses on the development and applications of a pseudobond ab initio QM/MM approach to study enzyme reactions. The following aspects of methodology development are discussed: the approaches for the QM/MM covalent boundary problem, an efficient iterative optimization procedure, the methods to determine enzyme reaction paths, and the approaches to calculate free energy change in enzyme reactions. Several applications are described to illustrate the capability of the methods. Finally, future directions are discussed.

Keywords Ab initio QM/MM methods · Density function theory (DFT) · QM/MM interactions · Free energy · Enzyme reaction · Transition state

1 Introduction

Enzymes, the remarkable catalysts provided by nature, play essential roles in every biological process. They catalyze a variety of chemical reactions with great efficiency and specificity. In order to fully understand inner workings of enzyme catalysis, information from experiments is necessary, but often insufficient. With the developments in computer technology and advances in computational chemistry, there is of great interest in computational studies of enzymes to complement experimental investigations.

The primary difficulties for computational methods to study enzyme reactions stem from the following three aspects:

1. Enzyme reactions, involving bond forming and breaking, requires the computational method to consider explicitly electronic configuration change during the process.
2. Enzymes are very large and heterogeneous, containing at least thousands of atoms. The remarkable capability of an enzyme is not only determined by its active site, but also controlled by its protein and solvent environment. Therefore, it requires the computational method to take the heterogeneous enzyme environment into account explicitly.
3. Under native conditions, an enzyme in solution undergoes a variety of constant random thermal motions and conformational transitions. To take account of enzyme dynamics, extensive statistical sampling should be carried out.

The high level quantum mechanical (QM) methods can describe chemical reactions, but computationally too expensive to study a system with thousands of atoms. The combined quantum mechanical and molecular mechanical (QM/MM) methods [1–3] and the ONIOM method [4] extend the realm of quantum mechanical calculations to large systems. In a QM/MM calculation, a small chemically active region is treated by a QM method, while the remainder of the system containing a large number of atoms is described by a molecular mechanical force field. Such a combined QM and MM approach can take advantage of the applicability and accuracy of the QM methods for chemical reactions and of the computational efficiency of the MM calculations. The total energy of this QM/MM system can be written as follows:

$$E_{\text{total}} = E_{\text{qm}}(\text{QM}) + E_{\text{mm}}(\text{MM}) + E_{\text{qm/mm}}(\text{QM/MM}). \tag{1}$$

where $E_{\text{qm}}(\text{QM})$ is the quantum mechanical energy of the QM sub-system, and $E_{\text{mm}}(\text{MM})$ the standard molecular mechanical interactions involving exclusively atoms in the MM sub-system. The QM/MM interaction between the QM sub-system and the MM sub-system can be divided into three terms: electrostatic contribution, van der Waals contribution and MM bonded interaction, as in the following equation:

$$\begin{aligned} E_{\text{qm/mm}}(\text{QM/MM}) = {} & E_{\text{electrostatics}}(\text{QM/MM}) \\ & + E_{\text{vdw}}(\text{QM/MM}) \\ & + E_{\text{MM-bonded}}(\text{QM/MM}), \end{aligned} \tag{2}$$

Y. Zhang
Department of Chemistry, New York University, New York, NY 10003, USA
E-mail: yingkai.zhang@nyu.edu

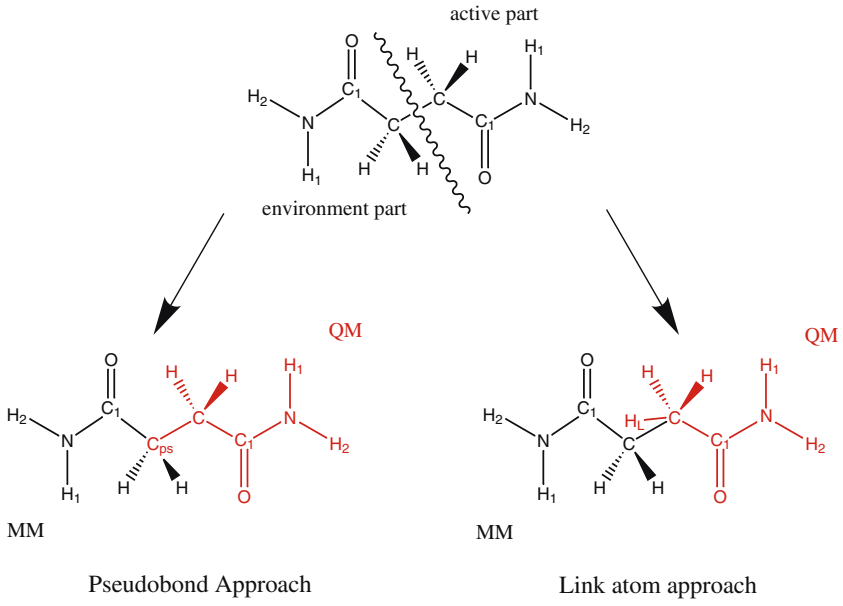

Fig. 1 Illustration of the difference between the pseudobond approach and the conventional link atom approach in the treatment of the QM/MM boundary problem

The sum of E_{qm}(QM) and $E_{\mathrm{electrostatics}}$(QM/MM) are calculated with a QM method, while the remaining terms are treated with an MM force field.

Over the last three decades, many combined QM/MM approaches have been developed, ranging from empirical valence bond (EVB) method [5–7], to semi-empirical QM/MM approaches [3, 8–11] , and to methods based on high level QM methods [2, 12–21]. The EVB method is based on the valence bond concept and provides a smooth interpolation of the potential energy surface for the reaction region. In semi-empirical QM/MM approaches, several QM methods have been employed, including AM1 [22], PM3 [23], a semi-empirical density functional method SCC-DFTB [24], etc. Both EVB and semi-empirical QM/MM approaches have been widely employed to investigate enzyme reactions with extensive statistical sampling, and have provided much insights into enzyme catalysis [5–7, 25–31]. However, the difficulty to calibrate the EVB parameters and the deficiency of currently available semi-empirical methods limit their applicability and predictive power. For example, they are generally not suitable for studying enzymes, which contain transition metal atoms at their active sites. With the increase of the speed of computers and the development of more efficient algorithms which make ab initio QM calculations more affordable, there is consequently a great deal of interest in developing QM/MM methods based on ab initio QM methods, including density functional theory (DFT) with gradient-corrected and hybrid functionals. With an adequate basis sets, ab initio QM methods are more accurate to describe chemical reactions, and are more widely applicable.

Development and applications of combined QM/MM methods have been extensively reviewed [7, 25–27, 31–36, 29], and a full review is not repeated here. In this perspective article, we mainly focus on the pseudobond ab initio QM/MM approach and its applications to enzyme reactions.

2 Methods

In this section, the following four aspects of methodology development are discussed: the description of the QM/MM boundary across covalent bonds, an efficient iterative optimization procedure, the methods to determine enzyme reaction paths, and the approaches to calculate the free energy change in enzyme reactions.

2.1 Description of the QM/MM boundary across covalent bonds

A critical issue underlying the accuracy and applicability of the combined QM/MM methods for studying enzyme reactions is how to describe the QM/MM boundary across covalent bonds [3, 25, 12, 10, 37, 33]. For example, when the molecule in Fig. 1 needs to be treated by a QM/MM method, the active part resulted from the cutting of the C–C bond is a radical and has a free valence. The behavior of the radical is clearly much different from the original closed-shell system. Thus, it is not appropriate to simply treat the active part quantum mechanically. Over the years, a number of groups have made efforts to develop solutions for this covalent-bond-cutting boundary problem.

Link atom approach is the most straightforward prescription to this boundary problem [2–4, 12, 16, 38–40]. In the link atom approach, link atoms, which are generally hydrogen atoms, are inserted to cap the free valence of the active part, except that in the HyperChem software [41] where pseudo-halogen atoms are used in their semi-empirical QM/MM program in order to mimic the effect of the fragments, which are removed from the QM treatment [25]. The link atoms and the atoms in the active part form the closed-shell QM region, which can be described quantum mechanically, while the

rest are treated molecular mechanically. One main drawback of the link atom approach is the introduction of additional degrees of freedom into the system, which complicates the expression of the energy and force, the geometry optimization and molecular dynamics simulation. Although a variety of approaches have been made to alleviate these complications within the link-atom framework [38–40], there is a great deal of interest in the search for approaches without introducing additional atoms into the system.

An alternative approach to describe QM/MM interface across covalent bonds is the use of bonding hybrid orbitals, including the hybrid orbital method [1,42], local self-consistent field (LSCF) method [43–46], generalized hybrid orbital (GHO) method [37,47] and frozen orbital method [48,18]. In the LSCF method, the localized bonding orbitals, which are obtained from separate QM calculations on small model compounds, are used to cap the free valences of the active part. Due to the use of hybrid orbitals, extensive theoretical formulation and substantial code development are required for the implementation of hybrid orbital methods. Meanwhile, it has been realized that the use of hybrid orbitals alone cannot lead to a satisfactory description of the QM/MM interface, and some specific parameterizations are needed [37,48,18, 46,47].

A third category of methods for handling the QM/MM boundary problem neither introduces additional atoms into the system nor employs bonding hybrid orbitals, which includes pseudobond method [15,49], connection-atom method [50], quantum capping potential method [51], effective group potential method [52], and minimum principle approach [53]. In this category, the pseudobond method developed by Zhang, Lee and Yang [15] was the first approach developed for ab initio QM/MM methods. The main idea of the pseudobond approach [15] is as follows: we consider that a large molecule is partitioned into two parts, an environment part and an active part, by cutting a covalent σ bond $Y-X$. Y and X refer to boundary atoms of the environment part and the active part, respectively. Instead of using a hydrogen atom to cap the free valence of X atom as in the conventional link atom approach, here a pseudobond $Y_{ps}-X$ is formed by replacing the Y atom with a one-free-valence boundary Y atom (Y_{ps}). The Y_{ps} atom is parameterized to make the $Y_{ps}-X$ pseudobond mimic the original $Y-X$ bond with similar bond length and strength, and also similar effects on the rest of the active part. In the pseudobond approach, the Y_{ps} atom and all atoms in the active part form a well-defined (often closed-shell) QM sub-system, which can be treated by quantum mechanical methods. Excluding Y atom, the rest atoms in the environment part form the MM sub-system, which will be represented by a molecular mechanical force field. As illustrated in Fig. 1, the pseudobond approach offers a smooth connection at the QM/MM interface and does not introduce additional atoms into the system as the link-atom approach. In comparison with hybrid orbital methods, the formalism of the pseudobond approach is simpler and it does not necessitate extensive changes to an existing QM source code.

Very recently, a new formulation has been employed to develop the pseudobonds [49], in which the boundary carbon (C_{ps}) atom has (1) seven valence electrons, (2) nuclear charge seven, (3) an angular-momentum-independent effective core potential, (4) an STO-2G basis set. Seven valence electrons are just enough to doubly fill three out of the total four valence orbitals and leave the remaining one singly occupied; the C_{ps} atom thus has a free valence to make the pseudobond. Since the effect of core electrons has been included in the effective core potential, there is no core electron needed. Thus the total number of the electrons as well as the nuclear charge for this atom C_{ps} are seven. By parameterizing both the effective core potential and the basis set which have a total of six parameters, not only has the $C_{ps}(sp^3)$-$C(sp^3)$ pseudobond been further improved, but also have the accurate $C_{ps}(sp^3)$-$C(sp^2$, carbonyl) and $C_{ps}(sp^3)$-$N(sp^3)$ pseudobonds been developed for the cutting of protein backbones and nucleic acid bases with the 6-31G* basis set. The developed pseudobonds are independent of the molecular mechanical force field. Although the parameterization is performed with density functional calculations using hybrid B3LYP exchange-correlation functional, it is found that the same set of parameters is also applicable to Hartree–Fock and MP2 methods, as well as DFT calculations with other exchange-correlation functionals. Tests on a series of molecules yield very good structural, electronic and energetic results in comparison with the corresponding full ab initio QM calculations.

2.2 An efficient iterative optimization procedure

With a well-defined potential energy surface (PES), the next major task for computational chemistry is to locate the stationary points (minima and saddle points) on the PES, which corresponds to equilibrium geometries and transition states. There have been great efforts devoting to this subject and many successful methods have emerged [54–61]. However, the efficiency of an algorithm is highly dependent on the size of the system and the cost to calculate the PES. Those algorithms developed for pure QM studies, such as the quasi-Newton minimizer in redundant coordinates [55,56], are inefficient to tackle a system with thousands of moving atoms. On the other hand, some powerful minimization methods have been developed for the MM PES, such as the truncated newton method [59–61]. However, the requirement for the second order derivatives or for the large number of energy and gradient evaluations makes these methods impractical for ab initio QM/MM studies of large systems.

Here, an efficient iterative optimization procedure has been developed for the pseudobond ab initio QM/MM calculations [62]. The main idea is to optimize the small QM sub-system using the quasi-newton minimizer in redundant internal coordinates with QM/MM calculations, while minimizing the large MM sub-system with the truncated newton method in Cartesian coordinates with only MM calculations. Thus we can take the advantages of both minimizers and

avoid their disadvantages. With the smooth connection between the QM and MM regions offered by the pseudobond QM/MM approach [15], this iterative optimization procedure turns out to be very efficient. It should be noted that besides this efficient iterative optimization procedure [62], several other micro-iterative optimization approaches [18, 63,64] have been developed for QM/MM optimizations. All these methods differ in specific implementation details, but share the very similar spirit.

2.3 Methods to determine enzyme reaction paths

The characterization of enzyme reaction mechanisms is one of the primary goals of chemical biology. For many enzymes, in spite of extensive investigations, the detailed reaction mechanism is often unclear or unsettled due to experimental challenges in providing direct information regarding transition states and intermediates during the reaction process. Meanwhile, the capability to determine enzyme reaction mechanisms is one of the main strengths of ab initio QM/MM methods, and is a prerequisite to understand the structure-reactivity relationship. However, to locate transition states using an ab initio QM/MM method for an enzyme system with thousands of atoms is very challenging, considering that transition state search is far from a routine task even for a small chemical system. Over the years, there have been great efforts devoting to this subject. Here we focus on the approaches which have been adapted to study enzyme reactions.

The most intuitive and so far most commonly used approach to study enzyme reactions is the reaction coordinate driving (RCD) method: stepping along a proposed reaction coordinate, and performing the energy minimization with respect to the remaining coordinates [65,66,54,12,67]. Provided that the obtained minimum energy path is smooth and continuous, it has been shown that the energy maximum along the path is a good approximation for the transition state [68,69]. In applications, the initial choice of a reaction coordinate, which is a linear combination of internal coordinates, is often based on the determined reactant structure and the specific proposed reaction mechanism. In calculations, the constrained optimization procedure can be applied to different points along the reaction path by changing the value of the chosen reaction coordinate gradually from the initial state to the final state. Then for stationary points along a smooth and continuous path, frequency calculations can be carried out. The energy maximum on the path with one and only one imaginary frequency can be located as the transition state. Although the RCD method seems to be conceptually simple and can be easily implemented, it is actually not so easy to apply due to the difficulty to obtain a smooth and continuous path. Except for some simple and well-studied enzyme reactions, it often requires a lot of experience and many tries to choose a "perfect" reaction coordinate to achieve the goal. Therefore, there is a lot of interest to adapt more sophisticated methods, including both of chain-of-state methods and eigenvector-following methods.

Nudged elastic band (NEB) method [70–73] are among the most successful chain-of-state methods to determine chemical reaction paths. Very recently, by resolving issues coming from the large number of soft and floppy degrees of freedom in enzymes, Xie, Liu and Yang [74] have adapted the NEB method to study multi-step enzyme reactions with semi-empirical QM/MM approaches. Meanwhile, Liu et al. [75] developed a parallel iterative path optimization method for ab initio QM/MM modeling of enzyme reactions based on the adaption of the path optimization procedure [57]. In this method [75], an initial reaction path is defined by a discrete set of structures from interpolation using trial reactant/product structures, or trial reactant/transition state (TS)/product structures, or from previous path relaxation state. Then by the introduction of a new metric defining the distance between different structures in the configuration space, optimization can be performed on each structure along the path, which can be carried out in an embarrassingly parallel fashion. The method has been demonstrated its efficiency and applicability in testing on two enzymes TIM and 4OT, with the results in consistent with previous reaction paths obtained from the reaction coordinate driving method. Very recently, a two-step procedure [76] has been developed to determine enzyme reaction paths by combining the above two approaches: the NEB method [74] and the parallel iterative path optimization method [75]. Its efficiency and applicability were tested on two enzymes in which this two-step procedure [76] can reduce the number of path optimization cycles by half in comparison with using the parallel iterative path optimization method [75] alone.

Besides the chain-of-state methods, the eigenvector-following methods [54,77] have also been adapted to locate or refine the transition state using QM/MM methods [78,79, 64]. Although it is much more expensive, it has been shown to be useful in cases that reaction coordinate driving method has difficulty to obtain a smooth and continuous path but can find the structure in the quadratic region. Meanwhile, the results indicate that the RCD method is in general good enough to locate the transition state if an appropriate reaction coordinate can be chosen [79].

2.4 Free energy calculations

For an enzyme system with thousands of atoms, extensive sampling on the potential surface is needed to obtain reliable results. The free energy changes associated with the reactions are not only better-defined for such systems, but also characterize the reactions better than the potential energies. Ideally, extensive molecular dynamics or Monte Carlo simulations should be carried out on an entire QM/MM system to take care of the statistical mechanical sampling. In practice, it is computationally very expensive to directly perform MD or MC simulations on an ab initio QM/MM potential energy surface. There is a lot of interest in developing more efficient approaches.

Based on the pseudobond approach [15] and the efficient iterative optimization procedure [67], an ab initio QM/MM

free energy (QM/MM-FE) approach [67] was developed by employing two main approximations: (1) the dynamics of the QM subsystem and MM subsystem is independent of each other, (2) the fluctuation of QM subsystem is estimated with a harmonic approximation. In this approach, first the enzyme reaction path is determined with ab initio QM/MM method. Then the free energy change associate with QM/MM interactions along the reaction path can be calculated with free energy perturbation calculations. The contribution of the QM fluctuation to free energy change is estimated with frequency calculations. The method has been applied to study various enzymes [67,80–82]. Unlike the QM-FE approach [83,84,14] which combines gas phase QM calculations with free energy calculations, the advantage of the QM/MM-FE approach is that it determines the reaction path in enzyme environment with ab initio QM/MM approaches.

In the QM/MM-FE approach, the independent dynamics assumption may not be appropriate if the coupling between reaction dynamics and protein dynamics is strong. In order to address this issues, very recently Lu and Yang [85] have developed a novel formulation of reaction path potential with parameters which can be determined by ab initio QM/MM calculations. With the constructed reaction path potential, which employs a linear response formula for the QM charges and a second order approximation for the QM internal energy, the extensive molecular dynamics simulation on an enzyme can be efficiently carried out to calculate the free energy of enzyme reactions as well as the reaction rate. For the initial proton transfer step catalyzed by triosephosphate isomerase, this new method has been demonstrated to be able to calculate the reaction free energy barrier [85]. Furthermore, it has been employed to compute the transmission coefficients and examine the dynamic effects on the enzyme reaction rate [86].

3 Applications

In this section, we describe the applications of the pseudobond ab initio QM/MM approach to study several enzymes, including enolase [80], acetylcholinesterase [87,88], 4-oxalocrotonate tautomerase [81,82] and kinase [89]. These studies demonstrate that the method is powerful in providing detailed insights into enzyme reactions. Some theoretical predictions [80,81] were subsequently confirmed by experimental studies [90,82,91].

3.1 Enolase [80]

Enolase catalyzes the dehydration of 2-phospho-d-glycerate (PGA), to form phosphoenolpyruvate (PEP). Previous studies have suggested a stepwise mechanism. The first step involves the abstraction of the proton on carbon-2 of PGA (the α-proton) by a general base of the enzyme, and the second step is the leaving of the hydroxyl group on carbon-3 (the β-hydroxyl group) from an enolic intermediate, with the assistance of a general acid. Using the pseudobond ab initio QM/MM approach and free energy calculations, the reaction paths and free energy barriers for the two steps of the reaction have been obtained. The calculated free energies of activation are consistent with the reaction rates measured from experiments. To understand how the enzyme environment can favor two reaction steps which result in opposite changes of charge on the substrate, the energy decomposition analysis of the activation barriers of the two reaction steps were carried out. The analyses are well correlated with available site-directed mutegenesis studies on enolase. More importantly, combining the three-dimensional structure of the active site and the theoretical results (especially, the analyses on the roles of individual residues in transition state stabilization) brought about an important insight into the structure-function relationship of this enzyme. That is, the polar and charged residues at the active site are organized in a distinctive manner so that they do not interfere with (only modestly enhance) the transition state stabilizing effects of the metal cations in the α-proton abstraction step. At the same time, they strongly compensate for the transition-state destabilizing effects of the same metal cations in the β-hydroxyl group leaving step. It was found that to achieve overall catalytic efficiency, the structure of the enolase active site takes advantage of the fact that the charge reorganization procedures accompanying the two reaction steps take place in two different directions in space.

3.2 Acetylcholinesterase [87,88]

Acetylcholinesterase (AChE) is a serine hydrolase responsible for the termination of impulse signaling at cholinergic synapses. It catalyzes the hydrolysis of the neurotransmitter acetylcholine (ACh) with a remarkably high catalytic efficiency and it is also a promising drug-design target for the treatment of Alzheimer's disease. The essential catalytic functional unit of AChE is the catalytic triad consisting of Ser203, His447 and Glu334.

The mutation of any of these three residues to alanine leads to a decrease in activity of at least 3,300-fold. Although the catalytic roles of Ser203 and His447 are quite clear, several different roles proposed for Glu334 in the literature, including electrostatic stabilization, "charge-relay" mechanism and the "low-barrier hydrogen bond" mechanism. Besides the catalytic triad, another important component of the active center in AChE is the oxyanion hole. The x-ray structure of a transition state analog complex with torpedo AChE revealed that there exists a three-pronged oxyanion hole in the active site of AChE, in contrast to the two-pronged oxyanion hole in most serine proteases. The respective O to N hydrogen bond distances are 2.9, 2.9 and 3.2 $\mathring{A}$, respectively. The question is whether these three hydrogen bonds always formed in enzyme reactions, and whether the weaker hydrogen bond is less important than the other two.

From the ab initio initio QM/MM studies of first step of the acylation reaction, it is found that the catalytic role of the third residue of the catalytic triad, Glu334, is to stabilize the transition state through electrostatic interactions.

The calculations did not support the "charge-relay" mechanism or the "low-barrier" hydrogen bond mechanism. For the oxyanion hole, the calculations indicate that in the AChE-ACh Michaelis complex, there are only two hydrogen bonds formed between the carbonyl oxygen of ACh and the peptidic NH groups. As the reaction proceeds, the distance between the carbonyl oxygen of ACh and NH group of third residue becomes smaller, and the third hydrogen bond is formed both in the transition state and the tetrahedral intermediate. Such a progressive hydrogen bond formation makes that weak hydrogen bond at least as important as the other hydrogen bonds in enzyme catalysis. In order to take account of protein dynamics and to investigate the effect of different conformations on the enzyme reaction energy barrier, this reaction step has also been studied with a multiple QM/MM reaction path approach. The approach consists of two main components: generating enzyme-substrate conformations with classical molecular dynamics simulation, and mapping out the minimum reaction energy path for each conformational snapshot with combined QM/MM calculations. It is found that enzyme-substrate conformation fluctuations lead to significant differences in the calculated reaction energy barrier; however, the qualitative picture of the role of the catalytic triad and oxyanion hole in AChE catalysis is very consistent.

3.3 4-Oxalocrotonate tautomerase [81, 82]

4-Oxalocrotonate tautomerase (4-OT) catalyzes the isomerization of unconjugated α-keto acids, and is part of a degradative metabolic pathway that converts various aromatic hydrocarbons to intermediates in the citric acid cycle. Although experimental results suggest that this is a general acid-base mechanism, with Pro-1 acting as the general base, it is not clear which residue acts as the general acid in the reaction. One possible suggestion is about Arg39″, which is in the active site. Here ab initio QM/MM calculations have been carried out examine different proposed reaction schemes. The calculations clearly show that there is no general acid in the reaction. Arg-39″, which is hydrogen bonded to the carboxylate group of the substrate, and an ordered water, which moves closer to the site of the charge formed in the transition state and intermediate, play the main role in transition state/intermediate stabilization without acting as general acids in the reaction. This theoretical prediction about the role of Arg-39″ has recently been confirmed by experimental studies [91]. In addition, a combined ab initio QM/MM and experimental studies have shown that the protein backbone make important contributions to 4-OT catalysis [82].

3.4 Protein kinase [89]

Protein kinase (PKA) catalyzes the protein phosphorylation, which regulates all aspect of cell life. In spite of extensive experimental studies, some detailed questions regarding how PKA catalyzes the phosphorylation reaction remain unsettled, including whether the reaction is associative or dissociative, and the respective catalytic roles of specific residues, metal ions and structural elements. There are particularly intriguing questions regarding the catalytic roles of Asp166, Lys 168, glycine rich loop, etc.

Density functional theory QM/MM calculations have been carried out on the catalytic subunit of cAMP-dependent PKA. The QM/MM calculations indicate that the phosphorylation reaction catalyzed by PKA is mainly dissociative, and Asp166 serves as the catalytic base to accept the proton delivered by the substrate peptide. Among the key interactions in the active site, the Mg^{2+} ions, glycine rich loop, and Lys72 are found to stabilize the transition state through electrostatic interactions. On the other hand, Lys168, Asn171, Asp184, and the conserved waters bound to Mg^{2+} ions do not directly contribute to lower the energy barrier of the phosphorylation reaction, and the possible roles for these residues are proposed. The QM/MM calculations with different QM/MM partition schemes or different initial structures yield consistent results. In addition, 12 ns molecular dynamics simulations on both wild type and K168A mutated PKA, respectively, have been performed to demonstrate that the catalytic role of Lys168 is to keep ATP and substrate peptide in the near-attack reactive conformation.

4 Perspectives

The field of the development and application of ab initio QM/MM methods is expanding rapidly. The methods have become increasingly powerful in complementing experimental methods to elucidate the chemistry of the complex biological process and to investigate chemical reactions in the condensed phase. Meanwhile, it has been recognized that in order for ab initio QM/MM methods to become an equal partner of experimental methods, significant efforts are still needed to further improve their accuracy and applicability. The following are some thoughts on future directions.

1. One of the key directions to develop the next generation of ab initio QM/MM methods is to develop a more accurate, transparent and systematic treatment for the QM/MM covalent boundary problem. Despite that many efforts have been devoted to this subject, as it is discussed previously, this problem remains one of the major challenges for ab initio QM/MM methods. For example, in the pseudobond approach, although the $C_{ps}(sp^3)$-$C(sp^3)$, $C_{ps}(sp^3)$-$C(sp^2$, carbonyl) and $C_{ps}(sp^3)$-$N(sp^3)$ pseudobonds for 6-31G* basis set have been developed, the pseudobonds for other basis sets and many other pseudobonds have not been developed. Meanwhile, these developed pseudobond parameters are semi-empirical in nature because they are parameterized against molecular properties with a limited set of molecules. It would be ideal to design the boundary atom to mimic atomic properties of the corresponding atom instead of molecular properties.

2. Besides the QM/MM covalent boundary problem, another key aspect which limits the accuracy of ab initio QM/MM potential energy surface is the treatment of QM/MM interactions. In most ab initio QM/MM methods, the non-bonded QM/MM interactions consist of the van der Waals part through Lennard-Jones potential and the electrostatic part calculated through a coulombic term in an effective Hamiltonian with the MM atoms as fixed point charges. The energy of the effective Hamiltonian, which is obtained by QM calculations, is the sum of the QM energy of the QM subsystem (E_{QM}) and the electrostatic interaction between QM and MM subsystems. Although such a formulation suffices for many QM/MM applications, it only takes into account the polarization effect of the MM charges on QM subsystem but does not include the polarization effect of the QM subsystem on MM atoms or the polarization interaction among MM atoms. Coupling the QM method with a polarizable force field is certainly desirable and has been developed [92, 10, 93, 94]. However, improving the description of the polarization effect alone may not necessarily lead to a better description of the QM/MM interactions since polarization effect and exchange interaction are often coupled and compensated. The only repulsive term in the most QM/MM formulations, which partly takes account of the exchange effects, is the very short-ranged $1/r^{12}$ term of the Lennard–Jones potential. Such a short-ranged repulsive potential may become inadequate when a polarizable MM force field is employed. With the recent development in MM polarizable force fields [95, 96] and effective fragment potential method [97], it is expected that new coupling schemes are likely to emerge to take account of both polarization and exchange interactions, which can lead to a more accurate treatment of QM/MM interactions in the near future.
3. In comparison with EVB and semi-empirical QM/MM methods, one of the main drawback of ab initio QM/MM method is that it is often computationally too expensive to carry out conventional MD or MC simulations on an entire QM/MM system to take care of the statistical mechanical sampling. Much more efforts are clearly needed to develop more efficient approaches to sample the ab initio QM/MM potential energy surface extensively and reliably. Meanwhile, the appropriate treatment of boundary condition and electrostatic interaction is also very important in QM/MM calculations. This subject has been discussed in more details by Cui in this special issue.
4. For applications, one possible niche for density functional theory QM/MM methods is the study on enzymes containing transition metal at its active site [98, 36]. Meanwhile, ab initio QM/MM methods are expected to have significant impact for those enzymes which have been little understood experimentally.

Finally, it can be concluded that the future of ab initio QM/MM methods is very bright, considering the robustness and accuracy of ab initio QM methods, the advantage of QM/MM framework, the central importance of enzymes and complex chemical systems, as well as the continuing increase of the speed of computers. With further methodology development to improve their accuracy, efficiency and applicability, ab initio QM/MM methods would become equal partners of experimental methods in the not too distant future.

5 Acknowledgement

I thank Professors Weitao Yang for his encouragement and discussions. The support from National Science Foundation (CHE-CAREER-0448156) is gratefully acknowledged.

References

1. Warshel A, Levitt M (1976) J Mol Bio 103:227
2. Singh UC, and Kollman P (1986) J Comp Chem 7:718
3. Field MJ, Bash PA, and Karplus M (1990) J Comp Chem 11:700
4. Maseras F, and Morokuma K (1995) J Comp Chem 16:1170
5. Warshel A (1991) Computer modeling of chemical reactions in enzymes, Wiley, New York
6. Aqvist J, and Warshel A (1993) Chem Rev 93:2523
7. Warshel A (2003) Annu Rev Biophys Biomolec Struct 32:425
8. Gao J, and Xia X (1992) Science 258:631
9. Liu H, Muller-Plathe F, and van Gunsteren WF (1996). J Mol Biol 261:454
10. Bakowies D, and Thiel W (1996) J Phys Chem 100:10580
11. Cui Q, Elstner M, Kaxiras E, Frauenheim T, Karplus M (2001) J Phys Chem B 105:569
12. Eurenius KP, Chatfield DC, Brooks BR, Hodoscek M (1996) Int J Quantum Chem 60:1189
13. Bentzien J, Muller RP, Florian J, Warshel A (1998) J Phys Chem B 102:2293
14. Stanton RV, Perakyla M, Bakowies D, Kollman PA (1998) J Am Chem Soc 120:3448
15. Zhang Y, Lee TS, Yang W (1999) J Chem Phys 110:46
16. Lyne PD, Hodoscek M, Karplus M (1999) J Phys Chem A 103:3462
17. Eichinger M, Tavan P, Hutter J, Parrinello M (1999) J Chem Phys 110:10452
18. Murphy RB, Philipp DM, Friesner RA (2000) Chem Phys Lett 321:113
19. Yarne DA, Tuckerman ME, Martyna GJ (2001) J Chem Phys 115:3531
20. Laio A, Vandevondele J, Rothlisberger U, (2002) J Chem Phys 116:6941
21. Sherwood P, et al (2003) Theochem J Mol Struct 632:1
22. Dewar MJS, Thiel W (1977) J Am Chem Soc 99:4899
23. Stewart JJP (1989) J Comp Chem 10:209
24. Elstner M, et al (1998) Phys Rev B 58:7260
25. Gao J (1995) Methods and applications of combined quantum mechanical and molecular mechanical potentials, In: Review in computational chemistry, vol 7, p 119–185, VCH, New York
26. Monard G, Merz KM (1999) Accounts Chem. Res. 32:904
27. Gao J, Truhlar DG (2002) Annu Rev Phys Chem 53:467
28. Benkovic SJ, Hammes-schiffer S (2003) Science 301:1196
29. Shurki A, Warshel A, (2003) Adv Protein Chem 66:249
30. Hammes-schiffer S (2004) Curr Opin Struct Biol 14:192
31. Garcia-viloca M, Gao J, Karplus M, Truhlar DG (2004) Science 303:186
32. Gogonea V, Suarez D, van der Vaart A, Merz KW (2001) Curr Opin Struct Biol 11:217
33. Field MJ (2002) J Comput Chem 23:48
34. Zhang Y, Liu H, Yang W (2002) Ab initio qm/mm and free energy calculations of enzyme reactions, In: Schlick e.a.t (ed) Methods for macromolecular Modeling 332–354
35. Colombo MC, et al (2002) Chimia 56:13

36. Friesner RA, et al (2003) Coord Chem Rev 238:267
37. Gao J, Amara P, Alhambra C, Field MJ (1998) J Phys Chem A 102:4714
38. Das D, et al (2002) J Chem Phys 117:10534
39. Amara P, Field MJ (2003) Theor Chem Acc 109:43
40. Ferre N, Olivucci M (2003) Theochem-J Mol Struct 632:71
41. Hypercube, Inc. (1994) HyperChem users manual, computational chemistry, Hypercube, Inc., Waterloo, Ontario 1994
42. Kairys V, Jensen JH (2000) J Phys Chem A 104:6656
43. Thery V, Rinaldi D, Rivail J-L (1994) J Comp Chem 15:269
44. Monard G, Loos M, Thery V, Baka K, Rivail J-L (1996) Int J Quantum Chem 58:153
45. Assfeld X, Rivail J-L (1996), Chem Phys Lett 263:100
46. Ferre N, Assfeld X, Rivail JL (2002) J Comput Chem 23:610
47. Pu JZ, Gao J, Truhlar DG (2004) J Phys Chem A 108:632
48. Philipp DM, Friesner RA (1999) J Comput Chem 20:1468
49. Zhang Y (2005) J Chem Phys 122:024114
50. Antes I, Thiel W (1999) J Phys Chem A 103:9290
51. Dilabio GA, Hurley MM, Christiansen PA(2002) J Chem Phys 116:9578
52. Poteau R et al (2001) J Phys Chem A 105:198
53. Yasuda K, Yamaki D (2004) J Chem Phys 121:3964
54. Schlegel HB (1987) Optimization of equilibrium geometries and transition structures, In: Lawley KP (ed) Ab initio methods in quantum chemistry. Advances in chemical physics, vol 67, p 249–286, Wiley, New York
55. Pulay P, Fogarasi G (1992) J Chem Phys 96:2856
56. Peng C, Ayala PY, Schlegel HB, Frisch MJ (1996) J Comp Chem 17:49
57. Ayala PY, Schlegel HB (1997) J Chem Phys 107:375
58. Paizs B, Fogarasi G, Pulay P (1998) J Chem Phys 109:6571
59. Ponder FM, Richards FM (1987) J Comp Chem 8:1016
60. Derreumaux P, Zhang G, Schlick T, Brooks B (1994) J Comp Chem 15:532
61. Dembo RS, Steihaug T (1983) Math Program 26:190
62. Zhang Y, Liu H, Yang W (2000) J Chem Phys 112:3483
63. Vreven T, Morokuma K, Farkas O, Schlegel HB, Frisch MJ (2003) J Comput Chem 24:760
64. Prat-resina X, Bofill JM, Gonzalez-lafont A, Lluch JM (2004) Int J Quantum Chem 98:367
65. Dewar MJS, Kirschner S (1971) J Am Chem Soc 93:4291
66. Williams IH, Maggiora GM (1982) J Mol Struc 89:365
67. Zhang Y, Liu H, Yang W (2000) J Chem Phys 112:3483
68. Rothman MJ, Lohr LL (1980) Chem Phys Lett 70:405
69. Scharfenberg P (1981) Chem Phys Lett 79:115
70. Henkelman G, Jonsson H (1999) J Chem Phys 111:7010
71. Henkelman G, Uberuaga BP, Jonsson H (2000) J Chem Phys 113:9901
72. Chu JW, Trout BL, Brooks BR (2003) J Chem Phys 119:12708
73. Trygubenko SA, Wales DJ (2004) J Chem Phys 120:2082
74. Xie L, Liu H, Yang W (2004) J Chem Phys 120:8039
75. Liu H, Lu Z, Cisneros GA, Yang W (2004) J Chem Phys 121:697
76. Cisneros GA, Liu H, Lu Z, Yang W (2005) J Chem Phys 122:114502
77. Schlegel HB (2003) J Comput Chem 24:1514
78. Prat-resina X, et al (2002) Theor Chem Acc 107:147
79. Prat-resina X, Gonzalez-lafont A, Lluch JM (2003) Theochem J Mol Struct 632:297
80. Liu H, Zhang Y, Yang W (2000) J Am Chem Soc 122:6560
81. Cisneros GA, Liu H, Zhang Y, Yang W (2003) J Am Chem Soc 125:10384
82. Cisneros GA, Wang M, Silinski P, Fitzgerald MC, Yang W (2004) Biochemistry 43:6885
83. Chandrasekhar J, Smith SF, Jorgensen WL (1985) J Am Chem Soc 107:154
84. Jorgensen WL (1989) Acc Chem Res 22:184
85. Lu Z, Yang W (2004) J Chem Phys 121:89
86. Wang M, Lu Z, Yang W (2004) J Chem Phys 121:101
87. Zhang Y, Kua J, McCammon JA (2002) J Am Chem Soc 124:10572
88. Zhang Y, Kua J, McCammon JA (2003) J Phys Chem B 107:4459
89. Cheng Y, Zhang Y, McCammon JA (2005), J Am Chem Soc 127:1553
90. Poyner RR, Larsen TM, Wong SW, Reed GH (2002) Arch Biochem Biophys 401:155
91. Metanis N, Brik A, Dawson PE, Keinan E (2004) J Am Chem Soc 126:12726
92. Thompson MA (1996) J Phys Chem 100:14492
93. Gao J (1997) J Comput Chem 18:1061
94. Dupuis M, Aida M, Kawashima Y, Hirao K (2002) J Chem Phys 117:1242
95. Rick SW, Stuart SJ (2002) Potentials and algorithms for incorporating polarizability in computer simulations, In: Review in computational chemistry, vol 18 p 89–146, VCH, New York
96. Ponder JW, Case DA (2003) Adv Protein Chem 66:27
97. Gordon MS, et al (2001) J Phys Chem A 105:293
98. Ryde U (2003) Curr Opin Chem Biol 7:136