QM/MM STUDIES OF PHOSPHORYL TRANSFER REACTIONS IN ALKALINE PHOSPHATASE SUPERFAMILY

by

Guanhua Hou

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Chemistry)

at the

UNIVERSITY OF WISCONSIN–MADISON

2012

Date of final oral examination: 05/31/12

The dissertation is approved by the following members of the Final Oral Committee:

Qiang Cui, Professor, Chemistry
Arun Yethiraj, Professor, Chemistry
J.R. Schmidt, Assistant Professor, Chemistry
Edwin Sibert, Professor, Chemistry
Wm Wallace Cleland, Professor, Chemistry

QM/MM STUDIES OF PHOSPHORYL TRANSFER REACTIONS IN ALKALINE PHOSPHATASE SUPERFAMILY

Guanhua Hou

Under the supervision of Professors Qiang Cui
At the University of Wisconsin-Madison

Members in the Alkaline Phosphatase (AP) superfamily demonstrate amazing catalytic specificity and promiscuity for a wide range of substrates. In particular, AP and Nucleotide Pyrophosphatase/Phosphodiesterase (NPP) feature very similar active site structures with an identical bi-metallo zinc site, analogous nucleophiles and hydrogen bond interactions, yet distinct substrate selectivities: AP catalyzes phosphate monoester hydrolysis reactions with remarkable proficiency while maintaining a lower reactivity for phosphate diester hydrolysis; NPP, conversely, favors phosphate diesters over monoesters. This project aims at understanding the molecular origin of these functional differences of this pair of enzymes by state-of-the-art computational techniques and improving theoretical tools for describing condense phase phosphoryl transfer reactions. This project also provides useful understandings of the principles that control enzyme promiscuity and offers guidance for enzyme engineering.

A semi-empirical Density Functional Theory, the Self-Consistent-Charge Density-Functional-Tight-Binding (SCC-DFTB) theory, with the parameters specifically developed for phosphate hydrolysis reactions is used in the Quantum Mechanics/Molecular Mechanics framework for enzyme catalysis. A Poisson-Boltzmann (PB) solvation model together with a charge-dependent radii scheme is developed for an efficient and semi-quantitative characterization of aqueous reactions involving highly charged species. The SCC-DFTB/PB model is used to study aqueous phosphoryl transfer reactions that serve as the reference for understanding enzyme catalysis. A state-dependent QM/MM interaction scheme is also developed

to better describe enzyme reactions with significant charge redistributions, which are common for phosphoryl transfers.

Equipped with these methods, we study the hydrolysis reactions of two phosphate esters, $pNPP^{2-}$ and $MpNPP^{-}$, in solution, an AP mutant (R166S) and the wild type NPP. Extensive comparisons and the general agreement with available experimental data and high level computational results highlight the semi-quantitative feature of our model. Our calculation results suggest that AP and NPP catalyze phosphate mono- and di-ester hydrolysis via a loose and a synchronous transition state (TS), respectively, similar to the reactions in solution. In addition, we discuss several ambiguous points regarding the interpretation of experiment techniques, e.g., the thio substitution effects and the vanadate TS analog.

Qiang Cui

To my parents, Yinghui Hou and Yindi Yang.

For your unconditional love and support.

ABSTRACT

Members in the Alkaline Phosphatase (AP) superfamily demonstrate amazing catalytic specificity and promiscuity for a wide range of substrates. In particular, AP and Nucleotide Pyrophosphatase/Phosphodiesterase (NPP) feature very similar active site structures with an identical bi-metallo zinc site, analogous nucleophiles and hydrogen bond interactions, yet distinct substrate selectivities: AP catalyzes phosphate monoester hydrolysis reactions with remarkable proficiency while maintaining a lower reactivity for phosphate diester hydrolysis; NPP, conversely, favors phosphate diesters over monoesters. This project aims at understanding the molecular origin of these functional differences of this pair of enzymes by state-of-the-art computational techniques and improving theoretical tools for describing condense phase phosphoryl transfer reactions. This project also provides useful understandings of the principles that control enzyme promiscuity and offers guidance for enzyme engineering.

A semi-empirical Density Functional Theory, the Self-Consistent-Charge Density-Functional-Tight-Binding (SCC-DFTB) theory, with the parameters specifically developed for phosphate hydrolysis reactions is used in the Quantum Mechanics/Molecular Mechanics framework for enzyme catalysis. A Poisson-Boltzmann (PB) solvation model together with a charge-dependent radii scheme is developed for an efficient and semi-quantitative characterization of aqueous reactions involving highly charged species. The SCC-DFTB/PB model is used to study aqueous phosphoryl transfer reactions that serve as the reference for understanding enzyme catalysis. A state-dependent QM/MM interaction scheme is also developed

to better describe enzyme reactions with significant charge redistributions, which are common for phosphoryl transfers.

Equipped with these methods, we study the hydrolysis reactions of two phosphate esters, $pNPP^{2-}$ and $MpNPP^{-}$, in solution, an AP mutant (R166S) and the wild type NPP. Extensive comparisons and the general agreement with available experimental data and high level computational results highlight the semi-quantitative feature of our model. Our calculation results suggest that AP and NPP catalyze phosphate mono- and di-ester hydrolysis via a loose and a synchronous transition state (TS), respectively, similar to the reactions in solution. In addition, we discuss several ambiguous points regarding the interpretation of experiment techniques, e.g., the thio substitution effects and the vanadate TS analog.

NOMENCLATURE

AP	alkaline phosphatase
DFT	density functional theory
DFTB	density functional tight binding
GBSW	generalized Born with a simple switch
GSBP	generalized solvent boundary potential
KIE	kinetic isotope effect
KO	Klopman Ohno
LFER	linear free energy relationship
MM	molecular mechanics
MMP	methyl monophosphate
MmNPP	methyl m-nitro phenyl phosphate
MpNPP	methyl p-nitro phenyl phosphate
MPP	methyl phenyl phosphate
NOE	nuclear Overhauser effect
NPP	nucleotide pyrophosphatase/phosphodiesterase
PB	Poisson Boltzmann
PMF	potential of mean force
pNPP	p-nitro phenyl phosphate
QM	quantum mechanics

QM/MM	quantum mechanical molecular mechanical
SASA	solvent accessible surface area
SCC-DFTB	self-consistent charge density functional tight binding
TMP	trimethyl monophosphate
vdW	van der Waals
WHAM	weighted histogram analysis method

LIST OF REFERENCES

[1] G. Hou, X. Zhu and Q. Cui, "An implicit solvent model for SCC-DFTB with Charge-Dependent Radii", *J. Chem. Theory Comput.*, vol. 6 pp. 2303–2314, 2010.

[2] C. Yi, G. Jia, G. Hou, Q. Dai, G. Zheng, X. Jian, C. Yang, Q. Cui and C. He, "Iron-Catalyzed Oxidation Intermediates Captured in A DNA Repair Monooxygenase", *Nature*, vol. 468 pp. 330–333, 2010.

[3] G. Hou and Q. Cui, "Alkaline Phosphatase and Nucleotide pyrophosphatase/phosphodiesterase do not alter phosphoryl transfer transition state for phosphate di-esters relative to solution: A QM/MM analysis", *J. Am. Chem. Soc.*, vol. 134 pp. 229–246, 2012.

[4] D. Riccardi, X. Zhu, P. Goyal, S. Yang, G. Hou and Q. Cui, "Toward molecular models of proton pumping: challenges, methods and relevant applications", *Sci. China Chem.*, vol. 55 pp. 3–18, 2012.

[5] G. Hou and Q. Cui, "QM/MM studies of Linear Free Energy Relationship of phosphate diesters in solution and Alkaline Phosphatase superfamily", *(In preparation).*

[6] G. Hou, X. Zhu, M. Elstner and Q. Cui, "Charge dependent QM/MM interactions with the Self-Consistent-Charge Tight-Binding-Density-Functional Theory", *(In preparation).*

[7] G. Hou and Q. Cui, "QM/MM studies of phosphate monoester hydrolysis reactions in Alkaline Phosphatase and Nucleotide pyrophosphatase/phosphodiesterase", *(In preparation).*

TABLE OF CONTENTS

Page

Appendix

Page

Appendix

Page

LIST OF TABLES

Appendix
Table Page

Appendix
Table

LIST OF FIGURES

Figure Page

Appendix
Figure Page

Appendix
Figure Page

Appendix
Figure Page

Figure Page

Appendix
Figure

Figure Page

Figure Page

ACKNOWLEDGMENTS

Through the past five years, many people helped me in different ways without which my graduate study and the finish of this thesis work would be impossible. Therefore I want to express my genuine thanks to all of them, for their constant support and generous assist.

First and foremost, I want to convey my most sincere gratitude to my research advisor Prof. Qiang Cui, for his patient coach, guidance and support over the past five years. As a young and energetic mentor, Qiang is always available whenever I need help; as a wise and knowledgeable teacher, Qiang always provides insightful opinions on tough problems; as a pure and enthusiastic scientist, Qiang always inspires me to strive for perfection and devote myself to science and research. Working with him is an enjoyable experience and a great honor that I will remember forever.

My research projects would not have been successful without the support from people in Prof. Dan Herschlag's group at Standford University and Prof. Chuan He's group at University of Chicago. I appreciate their invaluable discussions and comments on the research and the share of experimental data. I would also like to thank Prof. Arun Yethiraj for his mentor and assistance in my job searching process. In addition, I want to acknowledge Prof. J.R. Schmidt, Prof. Edwin Sibert and Prof. Wm Wallace Cleland to be in my defense committee and read through my thesis. Last but not least, I want to thank my former research advisor Prof. Xin Xu in China who introduced computational chemistry to me and introduced me to Qiang.

Far too many people to mention individually have assisted me in so many ways during my work at Madison. They all have my sincere gratitude. In particular, I would like to thank Dr. Xiao Zhu, Ms. Puja Goyal and Dr. Michael Gaus who shared the office with me. Xiao is like my elder brother, always taking care of me, teaching me and helping me out in research and life. He is not only my labmate and collaborator, but also my friend forever. Puja is a nice and smart girl with respectful diligence and sincere love of science. Our numerous discussions from QM/MM method development to applications on biological systems are tremendously enlightening and beneficial to me. Michael is a professional in SCC-DFTB method and long-distance running. I appreciate his perspicacious suggestions and support on my research and life.

Other former and current members in Cui group are also much appreciated, and to name a few: Ms. Junjun Yu, Dr. Jan Zienau, Dr. Nilanjan Ghosh, Dr. Liang Ma, Dr. Jejoong Yoo, Dr. Peter Koenig, Ms. Nihal Korkmaz, Ms. Xiya Lu, Ms. Xueqin Pang, Mr. Leili Zhang. Many friends in chemistry department are also very helpful, and to name a few: Dr. Yijie Li, Dr. Wei Xiong, Dr. Zhan Lu, Ms. Xin Chen and Ms. Tianning Diao, Mr. Yicun Ni. I also want to thank my friends outside the department: Difeng Zhu, Kai Wang, Yizhou Jiang, Shengxiang Ji and Yu Zhang.

Family is always the most important part of my life. I want to reserve my ultimate thank-you to my father Yinghui Hou and my mother Yindi Yang. For their unconditional love and support, always being there when I needed, and never once complaining about how infrequently I visit. They deserve far more credit than I can ever give them. Therefore I want to devote all my love and work to them.

Chapter 1

Introduction

Enzyme catalysis is appealing as tens of order magnitude rate acceleration can be achieved by the elegant assembly of the very basic biological parts, such as the amino acids and metal ions. The "lock and key" model has been the hallmark of enzyme catalysis for decades, highlighting the remarkable specificity toward cognate substrates. However, it is increasingly recognized that many enzymes have promiscuous catalytic activities in which the enzyme can catalyze a wide spectrum of substrates, besides their cognate substrates, with considerable proficiencies, challenging the traditional view of enzyme functions. [7–11] The enzyme promiscuity has been proposed to play an important role in evolution process since it can give an enzyme a "head start" by maintaining the old functions during the development of new functions, therefore providing a selective advantage. [12, 13] From an application point of view, a thorough understanding of the mechanisms of enzyme promiscuity helps glean precious insights and provide useful guidance to selectively tune enzyme reactivities or develop new catalytic reactions in enzyme engineering. [14–19] However, our knowledge of this emerging field is far from enough to even address the very basic questions, such as, to what extent can high catalytic proficiency and promiscuity be combined in one enzyme, or how do evolutionary pressures shape the level of promiscuity. Therefore, systematic efforts are imperative to broaden our knowledge and deepen understandings.

In this context, the members from Alkaline Phosphatase (AP) superfamily provide perfect examples for comprehensive studies. The AP superfamily contains a set of evolutionarily related enzymes that are structurally related to AP. [20, 21] They catalyze the hydrolytic

reactions of various substrates that differ in charge, size, intrinsic reactivities and nature of transition states, such as phosphoryl transfer reactions, which arguably represent the most important chemical transformation in biology. [22–24] For example, the *E. Coli* AP catalyzes the hydrolytic reactions of phosphate monoesters for its physiological functions but also exhibit promiscuous activities for the hydrolysis of phosphate diesters and sulfate esters. Similarly, although the main function of Nucleotide Pyrophosphatase/Phosphodiesterase (NPP) is to hydrolyze phosphate diesters, it can also cleave phosphate monoesters and sulfate esters with considerable acceleration. The catalytic efficiencies vary greatly, ranging from $> 10^{20}$ for the cognate activity to 10^{6-11} for the promiscuous activity. In other words, the selectivity of AP and NPP for phosphate mono- and di-esters differ by up to a remarkable level of 10^{15} fold. [25–28] These significant levels of differences are particularly striking in light of the fact that AP and NPP are very similar in their active site features, e.g., both enzymes have an identical bi-metallo zinc site, analogous nucleophiles and hydrogen bond interactions. Therefore, this pair of enzymes are ideal for in-depth comparative analyses.

Dan Herschlag's lab has made remarkable progress toward understanding the factors that dictate the AP and NPP catalysis. [29–33] Based on the extensive studies via spectroscopy, linear free energy relationship (LFER) and kinetic isotope effects (KIE), it has been proposed that AP and NPP do not alter the transition states of phosphate mono- and di-esters compared to aqueous reactions. Instead, the enzymes can recognize and catalyze the substrates via different pathways: for phosphate monoesters, a loose TS is employed while for phosphate diesters, a more synchronous TS is employed. However, these experimental techniques and conclusions have been challenged, [34,35] underscoring the contentious feature of this subject.

The controversy comes from the difficulty of characterizing transition states. It's well established that understanding catalytic characteristics of enzymes hinges on elucidating the relevant transition states at an atomic level. [36–41] However, the popular experimental techniques, such as LFER and KIE, can only explore transition states indirectly, [42–44] resulting in difficulties of data interpretations. Under this scenario, the computer simulation

can serve as an important supplement to experimental approaches by explicitly correlating experimental data with reaction mechanisms. Nevertheless, computational methods also need to be tested by the ability of reproducing crucial experimental observables and further improved if necessary, thus maximizing the complementarity between computation and experiment.

For studying chemical reactions, the quantum mechanics (QM) method is required to describe the breaking and formation of chemical bonds. Due to the large size of the enzyme system and the significant amount of samplings to obtain statistical meaningful results, semi-empirical QM method is typically used in computational framework. The Self-Consistent-Charge Density-Functional-Tight-Binding (SCC-DFTB) method has been used in this project to meet the requirement. [45] The SCC-DFTB method is an approximate method derived from density functional theory by neglect, approximation and parameterization of interaction integrals. Its reasonable balance between computational speed and accuracy makes it possible to carry out the large number of reaction path and potential of mean force calculations that are crucial to address the key questions. A version of SCC-DFTB method that has been developed by including the third-order on-site extension and fitted using a set of phosphate hydrolysis reactions in the gas phase, referred as SCC-DFTBPR, [46] is used in this project. Its good performance for phosphate hydrolysis has been demonstrated by numerous successful applications in previous work. [47–49]

Aqueous reactions are usually the reference for enzyme catalysis, therefore having a decent description of aqueous reactions serves as the cornerstone of understanding enzyme catalysis. Although significant amount of experimental and computational work has been carried out to determine mechanisms of phosphate hydrolysis in solution, the results are still not conclusive. [43, 50, 51] The difficulties come from two major reasons: due to the multiple covalencies of the phosphorus atom, various mechanisms are possible; the reaction energy barriers for different mechanisms are quite similar and sensitive to the environment. In Chapter 2, a recently developed implicit solvent model for SCC-DFTB is introduced to rapidly explore the potential energy surface of aqueous reactions that involve highly

charged species. [52] The solvent effect, described as solvation free energy, is calculated using a popular model that employs Poisson-Boltzmann equation for electrostatics and a surface-area term for nonpolar contributions. To balance the treatment of species with different charge distributions, we make the atomic radii that define the dielectric boundary and solute cavity depend on the solute charge distribution. This model can be effectively used, in conjunction with high-level QM calculations, to explore the mechanisms of aqueous reactions for phosphate hydrolysis.

For enzyme reactions, quantum mechanics/molecular mechanics (QM/MM) method [53] is the most popular simulation framework in which the important enzyme matrix effects are captured by MM method at modest cost. In conventional QM/MM implementations, [54,55] the QM/MM interaction contains electrostatic and van der Waals terms: the electrostatic term describes the interaction between the QM electrons and MM point charges and takes the simple Coulomb form; the van der Waals term is often modeled by the Lennard-Jones form with predetermined parameters that are fixed through chemical reactions. [56,57] When the charge distribution of the QM region changes significantly, such as in the AP and NPP catalysis, these simple functional forms can lead to large errors since changes in the effective size and polarzability of the QM region are poorly modeled. [46] In Chapter 3, we describe a state-dependent QM/MM interaction scheme based on a damped Coulomb (Klopman-Ohno) form that is able to improve the description for the effect of charge redistribution. This novel scheme successfully improves the calculation accuracy for condense phase chemical reactions using SCC-DFTB method and has been used in our enzyme studies.

Equipped with these methods, in Chapter 4 we first look at the hydrolysis of a phosphate diester, $MpNPP^-$, in solution, two experimentally well-characterized variants of AP (R166S AP, R166S/E322Y AP) and wild type NPP. [58] The general agreements of benchmark calculations with available experimental data for reactions in solution and enzyme support the use of SCC-DFTBPR/MM for a semi-quantitative analysis of the AP and NPP catalysis. Although phosphate diesters are cognate substrates for NPP but promiscuous substrates for AP, the calculations suggest that their hydrolysis reactions catalyzed by AP and NPP feature

similar synchronous transition states that are slightly tighter in nature than those in solution. Therefore, this study provides the first direct computational support to the hypothesis that enzymes in the AP superfamily do not significantly alter the nature of transition states of their substrates compared to aqueous reactions.

Following this study, in Chapter 5 we further apply the computation methods to studying the hydrolysis of two similar aryl phosphate diesters, $MmNPP^-$ and MPP^-. Together with the work of $MpNPP^-$, we successfully reproduce the general trend of reaction energetics in solution and enzymes. The transition states of the enzyme reactions are very similar to those in aqueous reactions, featuring the synchronous nature. To compensate the semi-empirical feature of the SCC-DFTB method and reduce the overestimation of the substrate substitution effects, we explore a correction scheme based on one-step free energy perturbation and the high level *ab initio* QM method. Our benchmarks indicate that the correction scheme can quantitatively improve the agreement with experimental data.

With the help of Klopman-Ohno scheme developed in Chapter 3, in Chapter 6 we study the hydrolysis reactions of a phosphate monoester, $pNPP^{2-}$, which is more challenging for QM/MM framework due to the large amount of charge redistributions in chemical reactions. With the inclusion of the one-step free energy perturbation corrections by a high level density functional, the calculated reaction energetics are in decent agreement with experimental results and consistent with our diester studies. Our results suggest that AP and NPP employ a similar loose transition state for $pNPP^{2-}$ hydrolysis, clearly different from the more synchronous nature of transition state for phosphate diesters hydrolysis and fundamentally distinct from the two-step mechanism reported in previous theoretical work for a alkyl phosphate monoester. Therefore, these results, together with the studies of phosphate diester reactions, render the complete view of AP and NPP catalysis which agrees with the experimental hypothesis that AP and NPP recognize and catalyze different substrates via similar mechanisms to their aqueous reactions.

Chapter 2

An implicit solvent model for SCC-DFTB with Charge-Dependent Radii

2.1 Introduction

Many chemical reactions take place in solution so a proper description for solvation effect is one of the most important challenges for computational chemistry. Although major progress has been made in QM/MM [59–64] and *ab initio* molecular dynamics [65] methods in which the solvent molecules are treated explicitly, the cost of such calculations is still rather high. Therefore, implicit solvent models remain an attractive choice for many studies. In the context of studying chemical reactions, the most commonly used framework for treating solvent implicitly is the dielectric continuum model [66, 67] in which the solvent is replaced by a homogeneous dielectric medium. More sophisticated treatments based on integral equations have also been developed, such as (MC)SCF-RISM [68], although they tend to be computationally more expensive than dielectric continuum models.

Over the past few decades, many different dielectric solvent models have been developed in the quantum chemistry community, such as the Self-Consistent Reaction Field (SCRF) model [69, 70], Polarized Continuum Model (PCM) [71–83], Generalized Born (GB) model [84–90], Conductor-like Screening Model (COSMO) [91–96] and the Langevin Dipole model [97]. For the application to chemical reactions involving large solutes, there are two practical issues. First, the computational cost of implicit solvent model calculations is still rather high, especially when used with a high level QM method. Therefore, it is fairly common to perform gas-phase optimization for stationary points and then carry out single point energy

calculations in solution using a dielectric continuum model. This can be problematic when there is significant difference between the gas phase and solution potential energy landscape [98], a scenario which is not uncommon when the solute is highly charged or zwitterionic. The second problem is that most implicit solvent models employ a set of fixed atomic radii to define the solvent/solute dielectric boundary, and these radii are typically pre-optimized based on the experimental solvation free energies of a set of small molecules [66, 67, 99] and limited by the quality of the training set. The use of fixed atomic radii causes additional errors in application to chemical reactions as the description of transition states is rarely included during parametrization stage. Methods have been developed in which the molecular cavity is determined based on the electron isodensity surface [100, 101], although an optimal value for the electron density cutoff is not always straightforward to determine [102].

Motivated by these considerations, we have implemented a dielectric solvent model for an approximate density functional theory, the Self-Consistent-Charge Density-Functional-Tight-Binding (SCC-DFTB) method [45]. SCC-DFTB is an approximation to Density Functional Theory (DFT) based on a second-order expansion of DFT total energy around a reference electron density. With respect to computational efficiency, SCC-DFTB is comparable to the widely used semi-empirical methods such as AM1 and PM3, i.e., being 2-3 orders of magnitude faster than popular DFT methods. In terms of accuracy, fairly extensive benchmark calculations have indicated that it is particularly reliable for structural properties, while energetics are generally comparable to AM1 and PM3 [103–105]. With recent developments of SCC-DFTB [106, 107] for metal ions [108–111] and a few other elements that require d orbitals for a reliable description (e.g., phosphorus [46]), an effective implicit solvent model for SCC-DFTB will be very useful and complementary to existing models based on other semi-empirical methods [84, 112, 113]. Our model takes advantage of the finite difference Poisson-Boltzmann approach [114, 115] implemented in CHARMM [116], and has analytic first derivatives [117]. This makes it possible to perform geometry optimization, reaction path searchers and vibrational frequency calculations (based on numerical finite difference of first derivatives).

Our main aim is to use SCC-DFTB for quickly exploring minimum energy paths for reactions in solution, and then refine selected results based on higher level of theories. To be able to describe transition state and stable structures on equal footing, it is desirable to determine the atomic radii in a self-consistent fashion based on the electronic structure of the solute. The simple model we have adopted is to make the atomic radii depend on the Mulliken charges, which are fundamental to SCC-DFTB [45] and are solved self-consistently via an iterative procedure (see **Methods**). The similar idea was explored in the context of an implicit solvent model for PM3 [118]. More recently, as this work was in progress, charge-dependent radii have been developed for a DFT based COSMO approach [119, 120], and much improved results (solvation free energies and chemical reactions) compared to fixed-radii models have been reported for small ions.

We have developed two sets of solvation radii parameters for SCC-DFTB. The first set is for the standard second-order SCC-DFTB [45] with parameters for C, H, O, and N. We recommend to use this set for general applications to molecules consisting of these elements. The second set is for SCC-DFTBPR [46], which is a specific version parameterized for phosphate hydrolysis reaction and includes third order on-site terms for C, H, O, and P; this set can be useful for studying phosphate hydrolysis reactions, although we caution that SCC-DFTBPR has been parameterized mainly for monoanionic phosphates and a limited set of hydrolysis reactions. Two rather large training sets for solvation free energy with the emphasis on bio-related molecules (including 103 and 57 solutes for SCC-DFTB and SCC-DFTBPR, respectively) are used to develop the solvation radii parameters. Calculations on two additional sets of test molecules shows that the performance for neutral and charged species is rather well balanced and the error is comparable to the SM6 model [89], which is more sophisticated yet also much more expensive computationally. To illustrate the applicability of our model to chemical reactions in solution, we briefly study the hydrolysis of Mono-methyl Mono-phosphate ester (MMP) and Trimethyl Monophosphate ester (TMP). The results from the current implicit solvent model are generally consistent with previous *ab initio* calculations in conjunction with PCM [3, 121] or the Langevin dipole solvation

models [4], as well as with our explicit solvent simulations using SCC-DFTBPR/TIP3P [46]. Compared to the latter, however, the significant over-stabilization of the zwitterionic intermediate is avoided, which highlights the complementary value of implicit solvent models to explicit solvent methods for studying reactions that involve highly charged species.

The paper is organized as follows: in Sect. II we summarize the key theoretical foundation for our implicit solvent model for SCC-DFTB; details for the parameterization and benchmark calculations are also included. In Sect. III, we present results and discussions of the parameterization and benchmark data, including the overall performance for both the training and test sets of molecules, and results for the hydrolysis of MMP/TMP. Finally, we summarize in Sect. IV.

2.2 Methods

2.2.1 SCC-DFTB

Here we briefly recall the basic elements of SCC-DFTB [45, 108] that are important to the development of an implicit solvent model. The SCC-DFTB approach is based on a second-order expansion of the DFT total energy around a reference density, ρ_0,

$$\begin{aligned} E &= \sum_{i}^{occ} < \Psi_i|\hat{H}^0|\Psi_i > + \frac{1}{2}\iint (\frac{1}{|\vec{r}-\vec{r'}|} + \frac{\delta^2 E_{xc}}{\delta\rho\delta\rho'}|_{\rho_0)}\delta\rho\delta\rho' - \\ &\quad \frac{1}{2}\iint \frac{\rho_0'\rho_0}{|\vec{r}-\vec{r'}|} + E_{xc}[\rho_0] - \int V_{xc}[\rho_0]\rho_0 + E_{cc}, \end{aligned} \tag{2.1}$$

where $\hat{H}^0 = \hat{H}[\rho_0]$ is the effective Kohn-Sham Hamiltonian evaluated at the reference density ρ_0, and the Ψ_i are the Kohn-Sham orbitals. E_{xc} and V_{xc} are the exchange-correlation energy and potential, respectively, and E_{cc} is the core-core repulsion energy. With a minimal basis set, a monopole approximation for the second-order term and the two-center approximation to the integrals, the SCC-DFTB total energy is given in the following form,

$$E = \sum_{i\mu\upsilon} c_\mu^i c_\upsilon^i H_{\mu\upsilon}^0 + \frac{1}{2}\sum_{\alpha\beta}\gamma_{\alpha\beta}\Delta q_\alpha \Delta q_\beta + \frac{1}{2}\sum_{\alpha\beta} U[R_{\alpha\beta}; \rho_0^\alpha, \rho_0^\beta], \tag{2.2}$$

where the $c^i_{\mu/\upsilon}$ are orbital coefficients, $\Delta q_{\alpha/\beta}$ are the Mulliken charges on atom α/β, and $\gamma_{\alpha\beta}$ is the approximate second-order kernel derived based on two interacting spherical charges. The last pairwise summation gives the so-called repulsive potential term, which is the core-core repulsion plus double counting terms and defined relative to infinitely separated atomic species.

As discussed in our recent work [60, 106, 107], it was found that further including the third-order contribution can substantially improve calculated proton affinity; for a set of biologically relevant small molecules, significant improvements were observed even with only the on-site terms included. The corresponding expression for the SCC-DFTB total energy is,

$$E = \sum_{i\mu\upsilon} c^i_\mu c^i_\upsilon H^0_{\mu\upsilon} + \frac{1}{2}\sum_{\alpha\beta}\gamma_{\alpha\beta}\Delta q_\alpha \Delta q_\beta + \frac{1}{2}\sum_{\alpha\beta} U[R_{\alpha\beta};\rho^\alpha_0,\rho^\beta_0] + \frac{1}{6}\sum_\alpha U^d_\alpha \Delta q^3_\alpha, \tag{2.3}$$

where U^d_α is the derivative of the Hubbard parameter of atom α with respect to atomic charge. For the development of SCC-DFTBPR for phosphorus-containing systems [46], we found it was useful to adopt an empirical Gaussian functional form for the Hubbard charge derivative; i.e.

$$U^d_\alpha(q) = U^d_{0\alpha} + D_0 exp[-\Gamma_0(\Delta q_\alpha - Q_0)^2], \tag{2.4}$$

where the charge-independent parameter ($U^d_{0\alpha}$) is dependent on the element type, whereas the three parameters associated with the Gaussian (D_0, Γ_0, Q_0) are taken to be independent of element type to minimize the number of parameters.

2.2.2 The solvation model based on Surface area and Poisson-Boltzmann

The implicit solvent framework that we adapt is based on the popular formulation [122] that includes a surface-area-dependent non-polar component and an electrostatic component,

$$\Delta G_{sol} = \Delta G_{np} + \Delta G_{elec}, \tag{2.5}$$

where

$$\Delta G_{np} = \gamma S; \tag{2.6}$$

here S is the Solvent Accessible Surface Area (SASA) which is dependent on atomic radii and γ is a phenomenological surface tension coefficient.

The electrostatic solvation free energy ΔG_{elec} for a given charge distribution $\rho(\mathbf{r})$ is generally given by,

$$\Delta G_{elec} = \frac{1}{2} \iint d\mathbf{r}\, d\mathbf{r}' \rho(\mathbf{r}) G(\mathbf{r}, \mathbf{r}') \rho(\mathbf{r}'), \tag{2.7}$$

where $\frac{1}{2}$ reflects the linearity of the dielectric medium [123] and the reaction field Green's function $G(\mathbf{r}, \mathbf{r}')$ corresponds to the reaction field potential at $\mathbf{r}$ due to a unit charge at $\mathbf{r}'$ [124],

$$\phi_{rf}(\mathbf{r}) = \int d\mathbf{r}' G(\mathbf{r}, \mathbf{r}') \rho(\mathbf{r}'). \tag{2.8}$$

For a set of point charges, $\rho(\mathbf{r}) = \sum_\alpha q_\alpha \delta(\mathbf{r} - \mathbf{r}_\alpha)$, ΔG_{elec} is simplified to

$$\Delta G_{elec} = \frac{1}{2} \sum_\alpha q_\alpha \phi_{rf}(\mathbf{r}_\alpha) \tag{2.9}$$

The reaction-field potential $\phi_{rf}(\mathbf{r})$ is obtained by subtracting a reference electrostatic potential computed in vacuum, $\phi_v(\mathbf{r})$, from the electrostatic potential computed in the dielectric solvent medium, $\phi_s(\mathbf{r})$. The electrostatic potentials are determined as solutions of the (linearized) Poisson-Boltzmann (PB) equation [115, 125],

$$\nabla \cdot [\epsilon(\mathbf{r}) \nabla \phi(\mathbf{r})] - \overline{\kappa}^2(\mathbf{r}) \phi(\mathbf{r}) = -4\pi \rho(\mathbf{r}) \tag{2.10}$$

with the appropriate dielectric boundary ($\epsilon(\mathbf{r})$) and charge distributions in finite difference (FD) form using iterative numerical techniques. The solution yields the electrostatic potential at every grid point and the total electrostatic solvation free energy is given by

$$\Delta G_{elec} = \frac{1}{2} \sum_i q_i (\phi_{s,i} - \phi_{v,i}), \tag{2.11}$$

where q_i and ϕ_i are the charge and calculated potential at the ith gridpoint, for the cases of vacuum (v) and solution (s).

In SCC-DFTB, ΔG_{elec} in Eq.2.7 is also simplified by the fact that the charge (electrons plus nuclei) density is represented by a collection of atom-centered Mulliken charges, [45, 55]

$$\rho(\mathbf{r}) = \sum_\alpha \Delta q_\alpha \delta(\mathbf{r} - \mathbf{R}_\alpha), \tag{2.12}$$

where Δq_α is the Mulliken charge of atom α. Thus calculating ΔG_{elec} is a straightforward extension of the classical expression,

$$\begin{aligned}\Delta G_{elec} &= \frac{1}{2}\iint d\mathbf{r}\, d\mathbf{r}'\rho(\mathbf{r})G(\mathbf{r},\mathbf{r}')\rho(\mathbf{r}') \\ &= \frac{1}{2}\int d\mathbf{r}\rho(\mathbf{r})\phi_{rf}(\mathbf{r}) \\ &= \frac{1}{2}\sum_\alpha \Delta q_\alpha \phi_{rf}(\mathbf{R}_\alpha),\end{aligned} \tag{2.13}$$

Using variational principle, the solvation contribution to the total solute energy leads to additional terms in the SCC-DFTB matrix elements during SCF iterations:

$$\frac{1}{2}S_{\mu\nu}[\phi_{rf}(\mathbf{R_C}) + \phi_{rf}(\mathbf{R_D})] \qquad \mu \in C, \nu \in D, \tag{2.14}$$

where μ and ν run over a minimal set of localized pseudo-atomic Slater orbitals located on atoms C and D, respectively, and $S_{\mu\nu}$ is the overlap integral associated with the two basis functions.

Additional analytical gradient components from the solvation are calculated based on the finite difference force proposed by Im, et al. [117] They used a continuous, spline-based dielectric boundary, which has been shown to give accurate and numerically stable forces for PB calculations. The total solvation force acting on atom α is given by,

$$\begin{aligned}\mathbf{F}_\alpha^{sol} &= -\frac{\partial \Delta G_{sol}}{\partial \mathbf{R}_\alpha} \\ &= -\frac{\partial \Delta G_{elec}}{\partial \mathbf{R}_\alpha} - \frac{\partial \Delta G_{np}}{\partial \mathbf{R}_\alpha} \\ &= \mathbf{F}_\alpha^{RF} + \mathbf{F}_\alpha^{DB} + \mathbf{F}_\alpha^{IB} + \mathbf{F}_\alpha^{NP}\end{aligned} \tag{2.15}$$

This method calculated the electrostatic solvation force as a sum of individual terms [117]: reaction field force ($\mathbf{F}_\alpha^{RF}$) arising from the variation of atomic positions assuming the dielectric boundary remains constant, dielectric boundary force ($\mathbf{F}_\alpha^{DB}$) caused by the spatial variations of the dielectric function $\epsilon(\mathbf{r})$ from the solvent to the solute interior and ionic boundary force ($\mathbf{F}_\alpha^{IB}$) resulting from spatial variations of the modified Debye-H*ü*ckel

screening factor $\overline{\kappa}(\mathbf{r})$. In SCC-DFTB/PB approach, for the atom α located at position $\mathbf{R}_\alpha$, the three terms in the limit of infinitesimal grid spacing are

$$\begin{aligned}
\mathbf{F}_\alpha^{RF} &= -\int_V d\mathbf{r}\,[(\phi_s-\phi_v)\frac{\partial\rho(r)_\alpha}{\partial\mathbf{R}_\alpha}] \\
\mathbf{F}_\alpha^{DB} &= -\frac{1}{8\pi}\int_V d\mathbf{r}\,\phi_s\nabla\cdot[(\frac{\partial\epsilon}{\partial\mathbf{R}_\alpha}+\frac{\partial\epsilon}{\partial\Delta q_\alpha}\frac{\partial\Delta q_\alpha}{\partial\mathbf{R}_\alpha})\nabla\phi_s] \\
\mathbf{F}_\alpha^{IB} &= \frac{1}{8\pi}\int_V d\mathbf{r}\,(\phi_s)^2\frac{\partial\overline{\kappa}^2}{\partial\mathbf{R}_\alpha}
\end{aligned} \tag{2.16}$$

Calculations for the derivative of Mulliken charge, dielectric function and modified Debye-H*ü*ckel screening factor have been discussed in previous studies [117]. As preliminary tests indicate, the contribution from the second term in $\mathbf{F}_\alpha^{DB}$ is rather small, therefore we omit it to simplify calculation (i.e., to avoid solving the coupled-perturbed KS equations [126] for the derivative of the MO coefficients).

2.2.3 Charge-dependent Radii Scheme

To establish a simple relationship between the dielectric boundary and the electronic structure of the solute, we take the atomic radius of a solute atom α to be linearly dependent on its Mulliken charge, Δq_α,

$$R_\alpha = A_{i(\alpha)} + B_{i(\alpha)}\Delta q_\alpha \tag{2.17}$$

where $A_{i(\alpha)}, B_{i(\alpha)}$ are element type dependent parameters that need to be determined based on a training set (see below). Higher-order polynomials have also been tested although no systematic improvement in the results is observed.

Since the atomic radii have an impact on the solvation free energy and therefore on the solute wavefunction and the Mulliken charges, R_α and Δq_α need to be determined self-consistently through an iterative scheme:

1. Perform a gas phase SCC-DFTB energy calculation to obtain the initial solute wavefunction and Mulliken charges;

2. Substitute Mulliken charges into Eq. 2.17 to obtain the atomic radii and establish the dielectric boundary;

3. Solve the PB equation (Eq. 2.10) to obtain the reaction field, $\phi_{rf}(\mathbf{R}_\alpha)$;

4. Re-solve SCC-DFTB in the presence of reaction field perturbation (Eq.2.14) to obtain a new set of Mulliken charges;

5. Check the convergence of energy (0.001 kcal/mol used for this work), if the convergence criterion is not met, return to Step 2;

6. Based on converged atomic radii, calculate SASA, the nonpolar contribution and the total energy of the solute in solution.

For most molecules tested here, it requires less than 10 iterations (typically 4-8) of atomic radii/Mulliken charges update for each geometry.

2.2.4 Parameter Optimization

The new parameters in the SCC-DFTB/PB based solvation model are the $A_{i(\alpha)}, B_{i(\alpha)}$ in Eq.2.17, which are dependent only on the element type. Although in principle the surface tension parameter in Eq.2.6 can also be optimized, we have not done so because for the systems of interest, the non-polar contribution tends to be overwelmed by the electrostatic component; the value of γ adopted is 0.005 $kcal/(mol \cdot \mathring{A}^2)$, which is commonly used in protein simulations using implicit solvent models [127]. For optimizing $A_{i(\alpha)}, B_{i(\alpha)}$, two training sets with molecules of broad chemical compositions have been constructed (see **Supporting Information**), for which the experimental solvation free energies are taken from Ref. [84,89,128]. Set 1 is used for parameterizing the solvation model with the standard (second-order) SCC-DFTB method and includes 103 species that contain C, H, O, N; the list includes alkane, alkene, alkyne, arene, alcohol, aldehyde, carboxylic acid, ketone, ester, amine, amide and other bio-related molecules and ions. Set 2 is used for parameterizing the solvation model with SCC-DFTBPR and includes 57 species that contain C, H, O, P; the list includes representative species from Set 1 plus phosphorus-containing molecules. Both sets contain a large number of charged species (57 in Set 1 and 24 in Set 2), which is essential for parameterizing the charge dependence of atomic radii.

The parameters are optimized using a Genetic Algorithm (GA) [129] in which the "fitness" (ξ) is defined as the inverse of a weighted sum of difference between solvation free energies determined from calculation and experiment:

$$\xi^{-1} = \frac{\sum_{i=1} w_i [\Delta G_i^{solv}(exp) - \Delta G_i^{solv}(calc)]^2}{\sum_{i=1} w_i}, \tag{2.18}$$

where i is the index of species in the training set and the sum is over all molecules in the training set. For the weighting factors (w_i), 1.0 and 0.1 are used for the neutral molecules and ions according to the typical uncertainties in the experimental values; as analyzed by Kelly, et al, [89] the typical uncertainties in *experimental data* for neutral molecules and ions are 0.2 kcal/mol and 3 kcal/mol, respectively. During optimization, a micro-GA technique with a population of 10 chromosomes that is allowed to operate for 500 generations with uniform crossovers; see Ref. [130] for detailed descriptions and recommendations for GA options.

In principle, geometry change upon solvation should be taken into consideration for a meaningful comparison to experiment. In practice, this is very time-consuming for parameter fitting even with the semi-empirical QM method (SCC-DFTB) we employ here. Several authors discussed this point [89,119] and concluded that the change in geometry is generally small. However, in several cases, such as alcohol anions, we have observed significant structural changes upon solvation that have a substantial influence on the calculated solvation free energy. Therefore, a compromise is adopted: the gas phase geometries are used to obtain the initial set of solvation parameters ($A_{i(\alpha)}, B_{i(\alpha)}$); with this set of parameters, solutes that have solvation free energy changes larger than 5 kcal/mol upon geometry optimization in solution are identified and their geometries in solution are updated for the optimization of a new set of $A_{i(\alpha)}, B_{i(\alpha)}$; this cycle continues until all cases with major structural changes upon solvation have been taken into account.

It is worth of mentioning that systematic optimization of surface tension coefficient γ (Eq.2.6) results in negligible improvements for both neutral molecules alone and the overall

training sets. The possible reason is that the nonpolar contribution is also made charge-dependent due to the correlation between SASA and charge-dependent atomic radii. So compared with the fixed-radii scheme, its dependence on γ is much less.

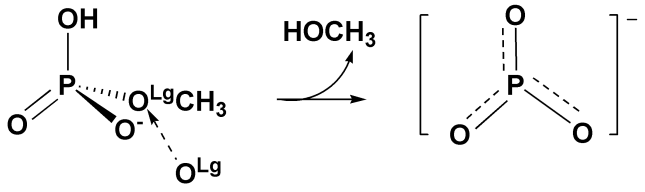
2.2.5 Additional Benchmark Calculations and studies of (H)MMP/TMP Hydrolysis

To test the transferability of the optimized parameters, test sets are constructed (see **Supporting Information**), which contain 32 for SCC-DFTB and 22 for SCC-DFTBPR. The calculated solvation free energies (including full geometry optimization in solution) are compared to the experimental values; similar to the training sets, the test cases contain a significant number of ionic species. As a comparison to popular and well-established solvation models, we also studied the same sets of molecules with the SM6 model of Cramer and Truhlar [89].

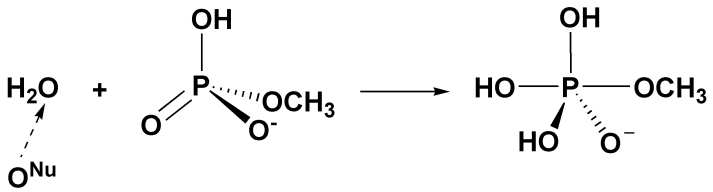
In addition, we have studied the mechanism [131, 132] (first steps of both dissociative and associative pathways, see Scheme 1) of Mono-methyl Mono-phosphate ester (MMP) hydrolysis using the SCC-DFTBPR/PB model. The potential energy surface is first explored by adiabatic mapping; the reaction coordinates include the $P-O^{Lg/Nu}$ distance (where O^{Lg} is the oxygen atom of the leaving group, methanol, and O^{Nu} is the oxygen in the nucleophilic water) and the anti-symmetric stretch that describes the relevant proton transfers that involve $O^{Lg/Nu}$. The anti-symmetric stretch is defined as the distance of donor-proton minus the distance of acceptor-proton. Each point in the 2D-adiabatic map is obtained by starting the constrained optimization from several different initial structures and taking the lowest energy value. Following the adiabatic mapping calculations, the structures along the approximate reaction path are examined carefully to ensure that the change of geometry is continuous along the path; in addition, the saddle point is optimized by Conjugated Peak Refinement (CPR) [133]. Finally, frequency calculations are carried out to confirm the nature of the stationary points and to compute the vibrational entropy and zero point energies. The results are compared to previous calculations with *ab initio* QM based

implicit solvent model calculations [3, 121, 134], SCC-DFTBPR/MM calculations by us [46] and available experimental data. To correct for intrinsic errors of SCC-DFTBPR, we also explore corrections based on single point energy calculations with B3LYP/6-311++G(d,p) at SCC-DFTBPR geometries in the gas phase; this level of theory was found to give very similar results for the reactions of interest compared to MP2 and large basis sets [46]. As discussed in the literature [135], such a simple correction may not always improve the energetics for semi-empirical methods given the errors in the geometries; however, our previous tests [46] indicated that this correction scheme appears useful for SCC-DFTBPR since the method gives fairly reliable structures, even for transition states.

Scheme 1



Dissociative Mechanism

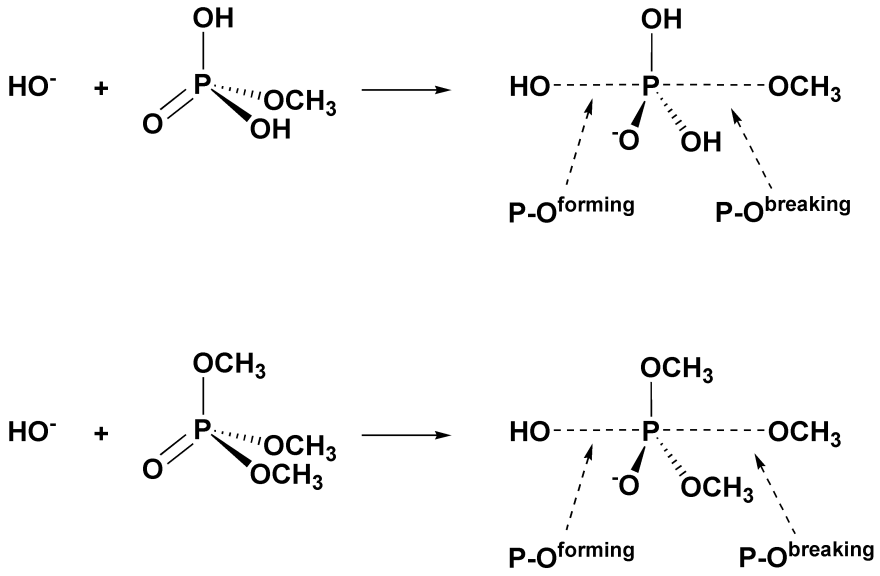


Associative Mechanism

Finally, we briefly compare the energetics of protonated MMP (HMMP) and Trimethyl Monophosphate ester (TMP) hydrolysis with OH^- as the nucleophile (see Scheme 2). This is motivated by the previous work of Warshel and co-worker [136], who discussed the roles of neutral water vs. OH^- as the nucleophile in MMP hydrolysis. Since SCC-DFTBPR was developed based on MMP hydrolysis with water as the nucleophile [46], this study helps to gain initial insights into the transferability of SCC-DFTBPR and lies the ground for

possible future developments. To better compare to previous calculations [4, 136], we follow the same 2-dimensional adiabatic mapping calculations with the bond lengths for the forming and breaking P-O bonds as the reaction coordinates. Single point B3LYP/6-311++G(d,p) calculations in the gas phase are used as an attempt to correct for intrinsic errors of SCC-DFTBPR.

Scheme 2



2.3 Results and Discussions

2.3.1 Performance for the training and test sets

The trends in optimized atomic radii (see Table 2.1) are consistent with other implicit solvent models and chemical intuition. For example, P has the largest charge-independent radius ($A_{i(\alpha)}$), while C, O, and N have comparable values, leaving H as the smallest. The absolute values are larger than those in SM6 and also the Bondi radii [137]. Compared with the atom type based charge-dependent radii in CD-COSMO by Dupuis et al. [119], comparable values are found for nitrogen and oxygen in our model and the "internal -N", "terminal oxygen" and "internal -O" in CD-COSMO. The hydrogen radius ($\sim$1.4 Å) in our

Table 2.1: Optimized atomic radii parameters and comparison to other values from the literature.

	SCC-DFTB		SCC-DFTBPR		SM6 [89]	Bondi [137]
Element	$A_{i(\alpha)}$	$B_{i(\alpha)}$	$A_{i(\alpha)}$	$B_{i(\alpha)}$		
C	1.85	-0.24	2.07	-0.05	1.57	1.70
O	1.70	-0.11	1.87	-0.07	1.52	1.52
N	1.94	-0.01	N/A	N/A	1.61	1.55
P	N/A	N/A	2.47	-0.10	1.80	1.80
H	1.47	-0.11	1.41	-0.25	1.02	1.20

a. $A_{i(\alpha)}$ in Å, $B_{i(\alpha)}$ in Å per charge. The values shown are fitted with solution geometry optimization (see **Methods**).

model is larger than that (polar hydrogen) in CD-COSMO (1.202Å). In terms of the charge-dependence, the typical $B_{i\alpha}$ values are around -0.10, although they are substantially larger ($\sim$-0.2) for C in SCC-DFTB and H in SCC-DFTBPR. Even the latter are nearly half of the values in CD-COSMO, which probably due to the use of different charges in SCC-DFTB (Mulliken) and CD-COSMO (CHELPG). It is worth emphasizing that the parameters in our model depend only on element type, rather than atom type as in CD-COSMO; therefore, CD-COSMO probably tends to be more accurate (see below for some comparison) while our scheme tends to be less problematic for studying transition states, which likely involve change in atom types.

As shown in the **Supporting Information**, the absolute value of solvation free energy is usually less than 10 kcal/mol for neutral molecules but larger than 60 kcal/mol for ions. Therefore, it is generally challenging to reproduce the solvation free energy of ions in a reliable fashion. Nevertheless, as shown in Table 2.2, the overall performance of our SCC-DFTB(PR) based solvation model is very encouraging. For example, for ions, the Mean Unsigned Error (MUE) for SCC-DFTB is $\sim$3 kcal/mol either without or with geometry optimization in

solution. For SCC-DFTBPR, the error is slightly larger, with the corresponding MUE values of 5 and 4 kcal/mol. These values can be compared to results from the SM6 model [89], which is one of the most sophisticated and well-calibrated models developed with *ab initio* DFT methods; the MUE values are 4 and 5 kcal/mol for the first (for SCC-DFTB) and second (for SCC-DFTBPR) training sets, respectively, which are even slightly larger than the values for our SCC-DFTB(PR) based solvation model.

The level of performance deteriorates slightly for the test sets. As shown in Table 2.3, for example, the MUE for the ions in the first and second test sets is 3 and 5 kcal/mol, respectively, when geometry optimization in solution is carried out; without solution geometry optimization, the MUE values are 4 and 6 kcal/mol. By comparison, the SM6 MUE values are 5 and 7 kcal/mol, again slightly larger than the SCC-DFTB(PR) values. These benchmark calculations indicate that the good performance of our model is fairly transferable. This is very encouraging since the SCC-DFTB(PR) based calculations are much faster than the DFT (MPW1PW91/6-31+G(d,p)) based SM6 calculations. Compared with CD-COSMO [119], which is also DFT based and involves more elaborate parameterization of charge-dependence of atomic radii, it is again encouraging to see that for the three ions tested by both models, the performance is comparable. For example, for hydroxide SCC-DFTB with or without solution geometry optimization gives an error of 2 kcal/mol while CD-COSMO gives 3 kcal/mol; for ammonium SCC-DFTB has an error of -3 kcal/mol while CD-COSMO gives -2 kcal/mol; for methylamine(+1), the corresponding values are -3 kcal/mol and -4 kcal/mol, respectively.

We note that, relatively speaking, the performance of our model for neutral molecules is less stellar. In fact, for both the training and test cases, the SM6 model consistently outperforms the SCC-DFTB(PR) solvation model; e.g., the MUE is typically smaller by $\sim$ 1 kcal/mol with SM6 (see Tables 2.2,2.3). This is likely because parameters in the non-polar component, which makes a significant (relative to ions) contribution to the total solvation free energy of neutral molecules, we have not optimized in the current model. Indeed, in the work of Xie et al. [128], who have implemented a GBSA model with SCC-DFTB, a Root

Table 2.2: Error (in kcal/mol) Analysis of Solvation Free Energies for Training Set 1 and 2[a]

	Single Point[b]			Optimization[c]			SM6[d]		
	RMSE	MUE	MSE	RMSE	MUE	MSE	RMSE	MUE	MSE
Neutral	2.0	1.7	0.6	2.1	1.7	0.4	0.8	0.7	0.4
Ions	4	3	2	3	3	0	4	4	2
All data	3	3	1	3	2	0	3	2	1
Neutral	1.6	1.3	-0.5	2.0	1.9	-1.3	1.5	0.9	0.6
Ions	4	5	5	4	4	2	4	5	5
All data	4	3	2	4	3	0	4	3	2

a. First three rows are for the first training set (for SCC-DFTB), and the three bottom rows are for the second training set (for SCC-DFTBPR). RMSE: Root-Mean-Square-Error; MUE: Mean-Unsigned-Error; MSE: Mean-Signed-Error. All errors measured against experimental solvation free energies, which have typical uncertainties of 0.2 kcal/mol and 3 kcal/mol for neutral molecules and ions, respectively. b. With gas-phase geometries. c. With solution phase geometry optimizations (see **Methods**). d. Results are obtained by MPW1PW91/6-31+G(d,p).

Table 2.3: Error Analysis (in kcal/mol) of Solvation Free Energies for Test Set 1 and 2[a]

	Single Point			Optimization			SM6		
	RMSE	MUE	MSE	RMSE	MUE	MSE	RMSE	MUE	MSE
Neutral	2.2	1.8	0.7	2.3	1.9	0.2	1.0	0.8	-0.2
Ions	5	4	2	4	3	1	6	5	2
All data	4	3	1	3	3	0	4	2	1
Neutral	1.5	1.4	-1.2	2.1	2.1	-2.0	0.9	0.7	-0.1
Ions	7	6	2	7	5	0	7	7	5
All data	4	3	0	4	3	-1	5	3	2

a. See Table 2.2 for format.

Mean Square Error (RMSE) of 1.1 kcal/mol was obtained for 60 neutral molecules containing C, H, O, N and S when the non-polar parameters were optimized. On the other hand, we note that for most chemical reactions of biological relevance, the non-polar contribution likely plays a much less significant role compared to the electrostatic component. Finally, as shown in **Supporting Information**, our solvation model gives rather large errors for amine and amide molecules; for example, the error for ammonia is more than 3.2 kcal/mol with or without solution geometry optimization, which is more than 70% off the experimental value. This behavior was noted in previous analysis of implicit solvation models [70], and it was argued that hydrogen-bonding energies are poorly correlated with classical electrostatic interaction energies and therefore more sophisticated treatments are needed for such short-range interactions.

2.3.2 MMP hydrolysis reaction with neutral water as nucleophile

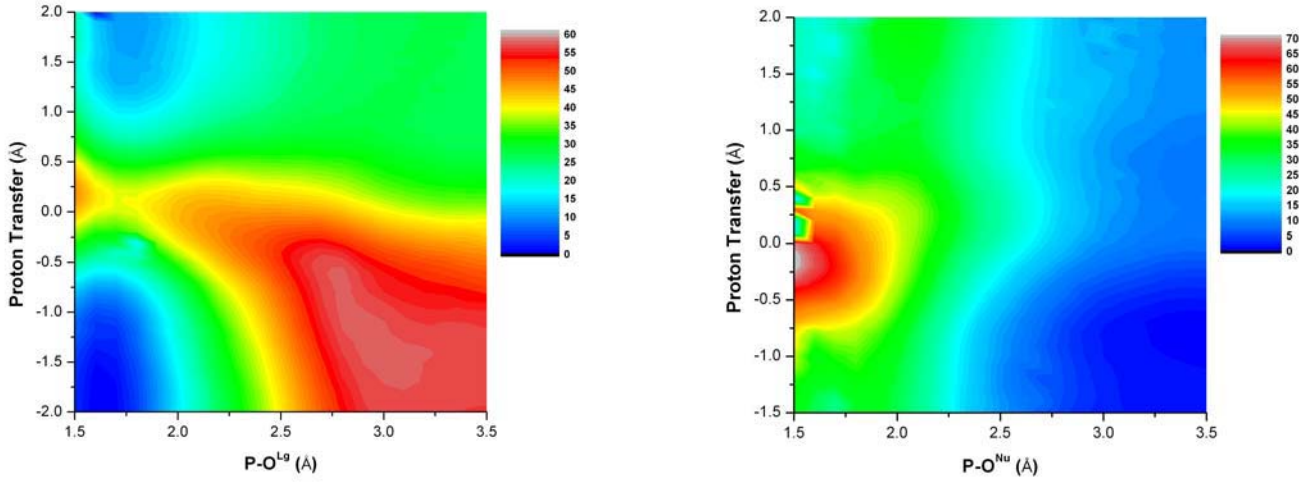
Experimental studies of MMP hydrolysis reaction [138–140] determined that the reaction rate peaks at pH 4-5 with activation energy of 31 kcal/mol. The reaction mechanism is traditionally regarded as dissociative though dispute still exists. [34] Here as a benchmark calculation for the new solvation model we investigate the first steps of both dissociative and associative pathways (see Scheme 1) and compare the results with previous theoretical studies [3, 4, 46].

For the dissociative pathway, the adiabatic map in solution with our new solvation model (Fig. 2.1a) is qualitatively consistent with previous PMF result obtained using explicit solvent SCC-DFTBPR/MM simulations [46]. The transition state region involves largely an intramolecular proton transfer from the protonated oxygen in MMP to the oxygen in the leaving group (O^{Lg}), and the $P-O^{Lg}$ bond is only slightly stretched compared to MMP. As discussed in Ref. [46], the $P-O^{Lg}$ bond in the transition state decreases significantly from the gas phase (∼2.1 Å) to solution (∼1.7-1.8 Åin SCC-DFTBPR/MM PMF simulations); thus our model has captured this solvation effect adequately. Following the proton transfer,

a zwitterionic intermediate is formed, which is again in qualitative agreement with both SCC-DFTBPR/MM PMF calculations [46] and previous DFT-PCM study [3].

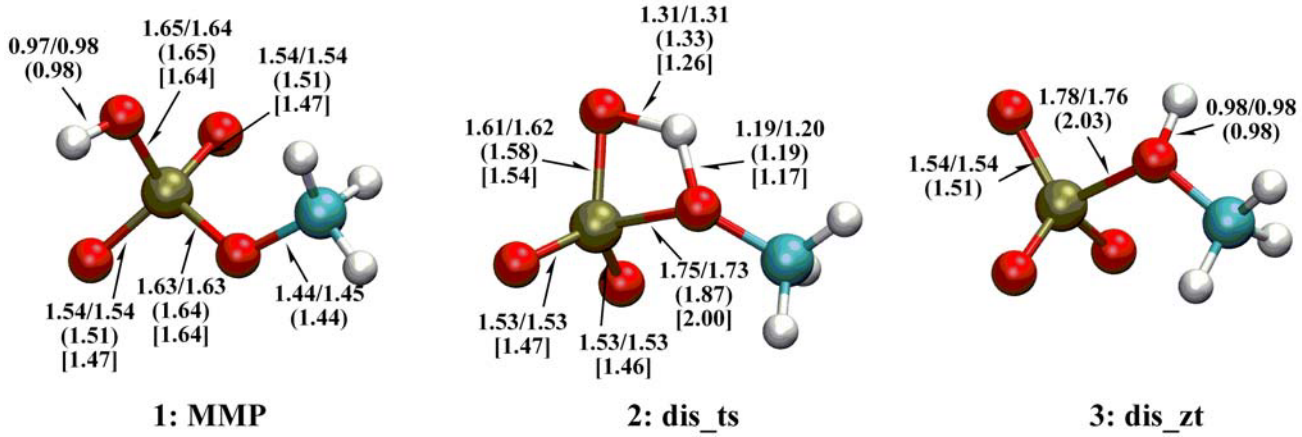
More quantitatively, the fully optimized structures for MMP, the transition state (**dis_ts**) and the zwitterionic intermediate (**dis_zt**) at the SCC-DFTBPR level are in decent agreement with previous calculations; the optimized structure does not depend sensitively on the grid size in the PB calculations (for comparison of 0.2 vs. 0.4 Å grid sizes, see Fig.2.2, which also contain an illustration for the imaginary mode in the optimized transition state, **dis_ts**, with a frequency of $1742i\text{cm}^{-1}$). Compared to the work of Vigroux et al. [3], in which the structures were optimized at the level of B3LYP-PCM and a double-zeta quality basis set plus diffuse and polarization functions, and pseudo-potential for non-hydrogen atoms, the only major difference is that their optimized $P-O^{Lg}$ distances in **dis_ts** and **dis_zt** are longer by ∼0.1 Å and 0.25 Å, respectively. The study of Florian et al. [4] did not examine the zwitterionic intermediate, and the $P-O^{Lg}$ distance in their transition state is substantially longer than both values from this work and from Ref. [3]; this is likely because geometries of Florian et al. [4] were mainly optimized in the gas-phase and the transition state in solution was only approximately located by single point Langevin dipole calculations along the minimum energy path from gas phase calculations.

For the energetics, the free energy barrier estimated with the current SCC-DFTBPR based solvation model is 34.8 kcal/mol; including single point B3LYP/6-311++G(d,p) gas-phase correction lowers the barrier to be 31.3 kcal/mol. As shown in Table 2.4, these values are consistent with previous calculations [3, 4] and experimental studies [141], which range from 30.7 to 34 kcal/mol. For the zwitterionic intermediate, which was first discussed in the work of Bianciotto et al. [3, 121] the current solvation model with SCC-DFTBPR predicts a free energy of 13.7 kcal/mol above the MMP reactant; with the B3LYP correction, the value becomes 21.1 kcal/mol. The large magnitude of the gas-phase correction was discussed in our previous study [46], which emphasized that the SCC-DFTBPR model was developed without any information concerning the zwitterionic region of the potential energy surface. The B3LYP corrected free energy value is in close agreement with the DFT-PCM study

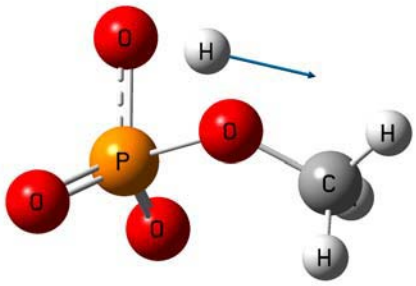


(a) (b)

Figure 2.1: Adiabatic mapping results (energies in kcal/mol) for the first step of (a) the dissociative (b) associative pathway for the hydrolysis of Monomethyl Monophosphate ester (MMP). The O^{Lg} stands for the oxygen in the leaving group (see Scheme 1), which is methanol in this case; O^{Nu} stands for the oxygen in water (see Scheme 1). In (a) the proton transfer coordinate is the antisymmetric stretch that describes the intramolecular proton transfer between the protonated oxygen in MMP and O^{Lg}; in (b), the proton transfer coordinate is the antisymmetric stretch that describes the proton transfer between the nucleophilic water and the basic oxygen in MMP.



(a)



(b)

Figure 2.2: Geometries of reactant, transition state and the zwitterionic intermediate for the first step of the dissociative pathway for the hydrolysis of Monomethyl Monophosphate ester (MMP). (a) Values (in Å) without parentheses are from the current SCC-DFTBPR based solvation model calculations with a grid size of 0.2/0.4 Å; values with parentheses are from Ref. [3], which were obtained with B3LYP-PCM and a double-zeta quality basis set plus diffuse and polarization functions; values with brackets are from Ref. [4], which were obtained with HF/6-31G(d) in the gas phase with approximate adjustments for solvation using the Langevin dipole model. (b) An illustration of the imaginary vibrational mode in **dis_ts**.

Table 2.4: Energetics for the first step of the dissociative pathway of MMP hydrolysis from current[a] and previous studies[b]

Species	ΔE^c	$T\Delta S^c$	ΔZPE^c	ΔG^c	Ref. [46]	Ref. [3]	Ref. [4]	Exp. [141]
MMP	*-11774.6*	*24.5*	*39.8*	*-11759.2*				
dis_ts	39.0/35.5	1.4	-2.8	34.8/31.3	32	33.5	34	30.7
dis_zt	12.6/20.0	-0.1	1.0	13.7/21.1	-3	21.2		

a. For MMP, the total energies are given (in italics), for other species, energetics relative to MMP are given in kcal/mol. The entropic contribution ($T\Delta S$, T=373K in all Tables, including for the experimental rate constants) and zero-point energy correction (ZPE) are calculated with the SCC-DFTBPR based solvation model and harmonic-oscillator-rigid-rotor approximation. b. Ref. [46] employs explicit solvent SCC-DFTBPR/MM PMF simulations; Ref. [3] used B3LYP-PCM and a double-zeta quality basis set plus diffuse and polarization functions, and pseudo-potential for non-hydrogen atoms; in Ref. [4], geometries were obtained with HF/6-31G(d) in the gas phase with approximate adjustments for solvation using the Langevin dipole model, single point calculations are performed at the MP2/6-31+G(d,p) level with Langevin dipole for solvation. c. Numbers before slash are SCC-DFTBPR results; numbers after slash are results after single point gas-phase correction at the level of B3LYP/6-311++G(d,p).

of Bianciotto et al. [3], who predicted a value of 21.2 kcal/mol. Most importantly, our solvation model does not suffer from the unphysically large stabilization found in explicit solvent SCC-DFTBPR/MM simulations, which predicted that the zwitterionic intermediate is *lower* than the reactant (MMP) by ∼3 kcal/mol. As discussed in Ref. [46], such significant overstabilization of the zwitterionic intermediate highlighted the need of improving QM/MM interactions beyond the typical form with parameters that do not reflect the electronic structure of the QM region [142]. The success of the current solvation model, on the other hand, illustrates the charge dependence required in QM/MM interactions can be effectively covered by the charge dependent radii when studying solution reactions that involve large charge redistribution.

For the associative pathway, the adiabatic map (Fig.2.1b) is qualitatively similar to the PMF from explicit solvent SCC-DFTBPR/MM simulations [46]. For example, the potential energy surface is rather flat in regions with long $P-O^{Nu}$ distances but positive proton transfer coordinate, which suggests that proton transfer from the nucleophillic water to MMP can occur prior to the nucleophillic attack. Indeed, we obtained a local minimum with geometry optimization that corresponds to a molecular complex between OH^- and protonated MMP (HMMP) on the potential energy surface. Compared to the reaction complex between water and MMP (**asc_pre**), this complex (**asc_hydro**) is substantially higher in energy by ∼15 kcal/mol; including the B3LYP/6-311++G(d,p) gas-phase correction further increases the value to ∼26.2-3.6=22.6 kcal/mol (see Table 2.5). Once again, the large magnitude of the correction reflects deficiency in the current SCC-DFTBPR approach for balancing proton affinity of phosphate and non-phosphate species, which remains an interesting challenge for future improvement [46].

Both the adiabatic mapping and saddle point optimization point to an associative transition state in which the $P-O^{Nu}$ distance is ~ 2Å and the water proton is already transferred to the phosphate oxygen (see Fig.2.3 for the structure of the transition state, **asc_ts**). Compared to the structure optimized by Florian et al. [4] with the Langevin dipole model, the key difference is that the proton transfer is halfway in their structure, with a $O^{Nu}-H$ distance

Table 2.5: Energetics for the first step of the associative pathway of MMP hydrolysis[a]

Species	ΔE	TΔS	ΔZPE	ΔG	Ref. [46]	Ref. [4]	Exp. [141]
MMP + H_2O	*-14345.9*	*38.3*	*53.1*	*-14331.2*			
asc_pre	-8.8/-7.0	-9.3	1.3	1.8/ 3.6			
asc_hydro	6.8/16.5	-9.1	0.6	16.5/26.2			
asc_ts	22.6/27.0	-9.6	0.8	33.1/37.5	34	35	30.7
asc_int	20.6/23.0	-10.4	1.4	32.5/34.9		29	

a. Same format as in Table 2.4; the reference is infinitely separated MMP and H_2O.

of 1.44 Å, compared to the value of 2.18 Å in our case. Since our structure is consistent with the previous PMF results based on SCC-DFTBPR/MM simulations, we suspect that the difference is again due to the limited solution geometry optimization in the work of Florian et al. [4] (see discussions above for the **dis_ts**). The agreement in the optimized structures for the penta-valent intermediate, **asc_int**, from the two sets of studies is much better, as expected (see Fig.2.3).

As for the energetics for the associated pathway, the SCC-DFTBPR based solvation model gives a free energy barrier of 33.1 kcal/mol, which increases slightly to 37.5 kcal/mol when gas-phase B3LYP correction is included. These values, especially the one with B3LYP correction, are close to previous computational studies (see Table 2.5) but somewhat higher compared to the experimental value of 30.7 kcal/mol [141]. The pentavalent species, **asc_int**, is also less stable by a few kcal/mol compared to the study of Florian et al. [4]. We note that all calculations found that the barrier for the associative pathway is higher than that in the dissociative pathway, although the difference is fairly small (∼1-2 kcal/mol) with either SCC-DFTBPR/MM or the Langevin dipole model, while SCC-DFTBPR based solvation model gives the largest difference (∼6 kcal/mol) when B3LYP correction is included. Before more systematic analysis into the quantitative nature of B3LYP correction, it remains premature to conclude that MMP hydrolysis strongly prefers a dissociative pathway.

2.3.3 HMMP and TMP hydrolysis with OH^- as nucleophile

A long-standing mechanistic postulate for MMP hydrolysis is that it is possible to exclude the nucleophilic attack of OH^- on the neutral phosphate. The argument was based on the high activation energy measured for the OH^- attack of trimethyl monophosphate (TMP) at high pH, which is around 25 kcal/mol (at 373K) [143], and the underlying assumption was that HMMP and TMP hydrolysis reactions have similar activation barriers. However, as pointed out by Warshel et al. [136], this analogy was not necessarily valid, and their calculations based on MP2 and Langevin dipole solvation model found that the barriers for OH^- attack of HMMP and TMP differ by more than 10 kcal/mol. Moreover, the barrier

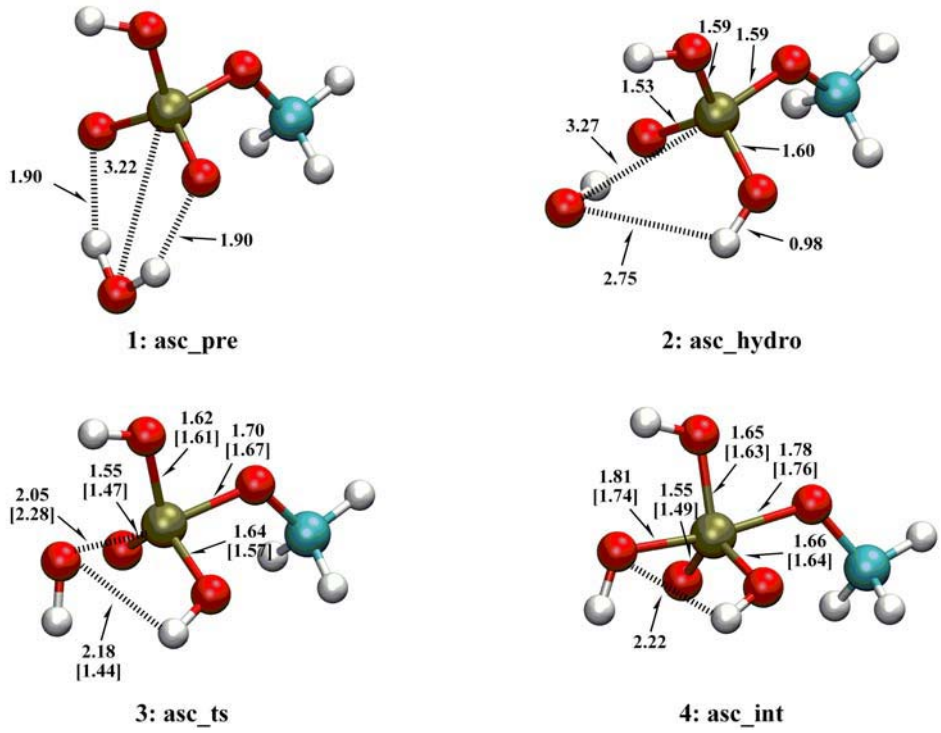


Figure 2.3: Similar to Fig.2.2, but for structures along the the first step of the associative pathway for MMP hydrolysis.

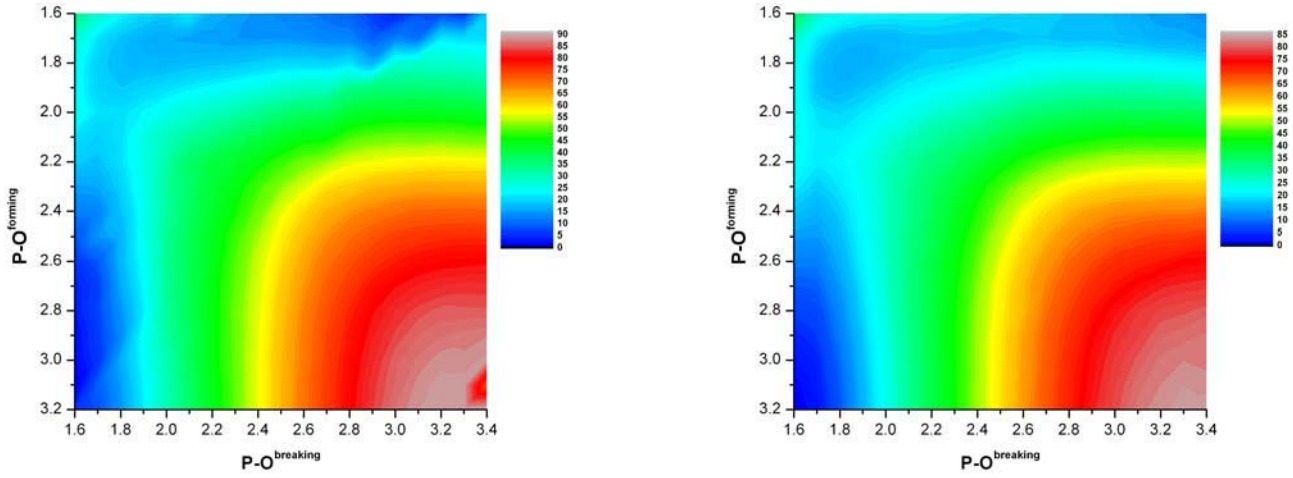
Table 2.6: Relative free energies of key species for the hydrolysis of MMP and TMP along associative pathway with hydroxide as the nucleophile[a].

Species	ΔE	TΔS	ΔZPE	ΔG	Ref. [136]	Exp. [141, 143]
asc_ts	15.6/ 7.0	-6.9	1.4	24.0/15.4	11.7	
tmp_ts	21.9/19.4	-7.9	1.2	30.9/28.5	24.7	24.6

a. Same format as in Table 2.4; the reference is infinitely separated HMMP/TMP and hydroxide.

of $\sim$ 12 kcal/mol found for HMMP was sufficiently low to make the OH^- attack pathway a competing mechanism of MMP hydrolysis. As an interesting benchmark of our solvation and the transferability of SCC-DFTBPR, we compare the barriers for the hydrolysis of HMMP and TMP with OH^- as nucleophile (see Scheme 2).

As shown in Fig.2.4, the overall energy landscapes are quite similar for HMMP and TMP, both undergoing an associative mechanism with the new P-$O^{forming}$ bond largely formed before the P-$O^{breaking}$ break. The transition state from the adiabatic mapping for HMMP is very consistent with the optimized saddle point **asc_ts**, which clearly is more appropriately classified as the transition state for OH^- attack of HMMP. According to Table 2.6, the corresponding energy barriers are 24.0 and 30.9 kcal/mol, with the TMP case higher by $\sim$ 7 kcal/mol. Including single point B3LYP/6-311++G(d,p) gas phase correction further increases the gap to $\sim$ 13 kcal/mol, which agrees very well with the result of Warshel and Florian [136]. This is a satisfying observation since SCC-DFTBPR was mainly parameterized based on MMP and Di-methyl monophosphate ester (DMP) hydrolysis; as speculated in our original work [46], however, the parameters are likely transferrable to other phosphates that follow similar reaction mechanisms because the number of parameters is fairly small. On the absolute scale, it appears that our estimates (for both HMMP and TMP) are systematically higher, by $\sim$ 4 kcal/mol, than the results of Warshel et al. [136] and the experimental barrier for TMP [143].



(a) (b)

Figure 2.4: Adiabatic mapping results (energies in kcal/mol) for the hydrolysis of (a) Hydrogen Methyl Monophosphate ester (HMMP) and (b) Trimethyl Monophosphate ester (TMP) by hydroxide. See Table 2.6 for the summary of the barrier heights, in which the reference is infinitely separated reactant molecules.

2.4 Conclusion

We report the development of an implicit solvent model for SCC-DFTB(PR) in which the solvation free energy is computed based on Poisson-Boltzmann for electrostatics and a surface area term for non-polar contributions. The unique aspect of our model is that the atomic radii that define the dielectric boundary of the solute are dependent on the solute charge distribution and are determined in a self-consistent fashion with the electronic structure of the solute. This self-consistency makes it possible to balance the solvation treatment of species with different charge distributions, such as neutral vs. ionic species and structures along a chemical reaction pathway. Indeed, benchmark calculations have shown that, even for ions, our model leads to results of comparable accuracy to the much more sophisticated SM6 model; this is very encouraging since SCC-DFTB(PR) calculations are at least hundreds of times faster than the DFT calculations required in the SM6 model.

Since our implementation has analytic first derivatives, the solvation model can be readily used to explore potential energy surfaces for solution reactions. This is demonstrated with a brief study of dissociative and associative pathways of MMP hydrolysis, as well as the hydrolysis of protonated MMP and TMP with OH^- as the nucleophile. The results (geometries and energetics) are largely in good agreement with previous computational studies using QM/MM or *ab initio*/DFT in conjunction with dielectric continuum models, as well as with available experiments. In particular, the solvation model avoids the overstabilization of the zwitterionic species along the dissociative pathway as found in explicit solvent SCC-DFTBPR/MM simulations [46]. This highlights the complementary nature of implicit solvent model to explicit solvent approaches for studying solution reactions that involve significant charge reorganizations.

Due largely to the computational efficiency of SCC-DFTB(PR), we anticipate that the current solvation model can be effectively used in semi-quantitative exploration of mechanisms for solution reactions, such as ruling out certain reaction pathways and obtaining approximate structures of key transition states and intermediates, which can be further refined

with higher-level calculations. As further developments, it would be interesting to extend the formulation of charge-dependent radii to more approximate solvation models such as Generalized Born [127], which can be computationally more efficient than Poisson-Boltzmann; this is particularly true in molecular dynamics simulations, which can be effective for estimating entropic contribution to reaction energetics in the framework of quasiharmonic analysis. Along this line, as extensively discussed in the literature, the first solvation shell of the solute can be treated explicitly, either at the same level of QM theory [67, 70] or with a Molecular Mechanics model [144, 145]. Since SCC-DFTB(PR) is fast, making such extension of the molecular model for better treatment of solvation is likely more cost effective than with *ab initio*/DFT methods.

Chapter 3

Charge-dependent QM/MM interactions with the Self-Consistent-Charge Tight-Binding-Density-Functional Theory

3.1 Introduction

With the increase of computational power, the analysis of chemical events in complex systems attracts more and more interests, e.g., the study of enzyme catalysis, enzyme engineering and redesign, which further pushes the development of de novo computational techniques for better accuracy and efficiency. In the presence of chemical reactions, quantum mechanics (QM) is required to describe the breaking and formation of chemical bonds. Despite the remarkable efforts and progress of new computation algorithm, large scale paralle

The total Hamiltonian for the molecular system under consideration in the QM/MM framework is

$$\hat{H} = \hat{H}^{QM} + \hat{H}^{QM/MM} + \hat{H}^{MM} \tag{3.1}$$

where $\hat{H}^{QM/MM}$ describes the interaction between the QM and MM atoms governed by $\hat{H}^{QM}$ and $\hat{H}^{MM}$, respectively. The $\hat{H}^{QM/MM}$ typically contains terms for the electrostatic, van der Waals (vdW), and bonded interactions

$$\hat{H}^{QM/MM} = \hat{H}^{QM/MM}_{vdW} + \hat{H}^{QM/MM}_{elec} + \hat{H}^{QM/MM}_{bonded} \tag{3.2}$$

The major contributions for long range interactions usually come from the $\hat{H}^{QM/MM}_{elec}$ while $\hat{H}^{QM/MM}_{vdW}$ plays an important role in the short range to estimate dispersion attractions that

fall off as r^{-6} and to prevent molecular collapse being strongly repulsive at short interaction distances. The $\hat{H}_{bonded}^{QM/MM}$ is required when partitioning a single molecule into quantum and molecular mechanics regions, whereas the valency of the QM region is satisfied with the addition of link atom [146] or frontier bonds. [147, 148]

In spite of the tremendous success of the conventional QM/MM interaction scheme, some limitations also exist and need to be improved for better performance. The first is that the vdW parameters are typically assigned based on pre-defined atomic types and fixed through chemical reactions, even though the chemical properties of the system can undergo drastic change, which is very common for highly charged systems, such as phosphate hydrolysis reactions. For example, when a water goes to attack a phosphate ester, it can lose its proton to the nonbridging phosphate oxygen first to form a hydroxide, then forms the P-O bond and finally transfers the other hydrogen to become a nonbridging phosphate oxygen. The chemical properties of the water oxygen experience drastic changes and are quite problematic to be described by a single set of vdW parameters. Element type of vdW parameters can avoid the trouble of pre-assignment but the performance is typically compromised (see the result part for some examples). The second problem is related to the semi-empirical QM method we use in the QM/MM framework. The Self-Consistent-Charge Density-Functional-Tight-Binding (SCC-DFTB) theory [45] is an approximation to Density Functional Theory with balanced performance and efficiency. The $\hat{H}_{elec}^{QM/MM}$, in the SCC-DFTB framework, is modeled by point charge interactions, i.e., the Mulliken charges of QM atoms and atomic charges of MM atoms, instead of solving one-electron integrals rigorously which is typically adopted by *ab initio* QM methods. Therefore the spatial distributions of the electron density are poorly modeled and result in increased errors at the short range.

In order to solve the first problems, the York group made impressive pioneering work of developing a charge-dependent vdW interaction model. [149] But the method has a number of parameters and has only been applied to simple systems. Alternatively, we are inspired by the popular way of treating two-center two-electron integrals in semi-empirical QM field where the Klopman-Ohno (KO) type of scaling [150,151] is usually applied. This scaling form

smoothly connects the classic electrostatic interaction in the long range limit with the self interaction in one-center limit and lead to improved performance in intermediate distance. [152] Along this line, the KO scheme can be used for a better description of the deviations from classical point charge interactions due to the interactions of electronic orbitals when a QM atom and a MM atom are close to each other. With a set of element type dependent vdW parameters, the KO algorithm adds little extra cost, yet is able to significantly improve the QM/MM descriptions of chemical reactions.

In this work, we implement and parametrize the KO scheme with the SCC-DFTB method which is based on a second-order expansion of DFT total energy around a reference electron density. With respect to computational efficiency, SCC-DFTB is comparable to the widely used semi-empirical methods such as AM1 and PM3, i.e., being 2-3 orders of magnitude faster than popular DFT methods. In terms of accuracy, fairly extensive benchmark calculations have indicated that it is particularly reliable for structural properties, while energetics are generally comparable to AM1 and PM3 [103–105]. There are several recent developments of SCC-DFTB [106, 107, 153] for metal ions [108–111] and a few other elements that require d orbitals for a reliable description (e.g., phosphorus [46]).

The paper is organized as follows: in Sect.3.2 we summarize computational methods and simulation setup. In Sect.3.3, we first present results for simple cluster model, and then demonstrate the performance for phosphate monoester dianion hydrolysis reactions in solution. Finally we draw some conclusions.

3.2 Theory and Methods

3.2.1 Conventional QM/MM Energy Evaluation.

According to eq 3.1, the energy of QM/MM simulations is determined by combining the Hamiltonians of the quantum mechanical and molecular mechanical regions with a QM/MM coupling term composed of electrostatic, bonded, and vdW contributions

$$U_{tot} = \langle \Psi | \hat{H}^{QM} + \hat{H}^{QM}_{elec} | \Psi \rangle + U^{QM/MM}_{vdW} + U^{QM/MM}_{bonded} + U^{MM} \tag{3.3}$$

The QM approach used here is SCC-DFTB, [45] which is very efficient due mainly to approximations to the two-electron integrals. This method introduces the charge self-consistency at the level of Mulliken population and, accordingly, the QM atoms interact with the MM sites electrostatically through Mulliken partial charges [55]

$$U_{elec}^{QM/MM} = \sum_{A \in MM} \sum_{B \in QM} \frac{Q_A \Delta q_B}{|\mathbf{R_A} - \mathbf{R_B}|} \tag{3.4}$$

where Q_A and Δq_B are the MM partial charges and Mulliken partial charges, respectively. We note that although other definitions of charges in SCC-DFTB and SCC-DFTB/MM calculations can in principle be used instead of the simple Mulliken charges, important parameters in SCC-DFTB (e.g., repulsive potentials) were optimized within the Mulliken framework.

The vdW term consists of predetermined parameters described by

$$U_{vdW}^{QM/MM} = \sum_{A \in MM} \sum_{B \in QM} \epsilon_{AB} [(\frac{\sigma_{AB}}{R_{AB}})^{12} - 2(\frac{\sigma_{AB}}{R_{AB}})^{6}] \tag{3.5}$$

where A and B are the indices for the MM and QM nuclei, respectively, and R_{AB} is the distance between QM and MM nuclei. The vdW parameters are defined by the standard combination rules: $\epsilon_{AB} = (\epsilon_A \epsilon_B)^{1/2}$ and $\sigma_{AB} = 1/2(\sigma_A + \sigma_B)$, where ϵ and σ describes the well depth and atomic radius, respectively. These parameters are typically atomic type based, therefore could be problematic for describing chemical reactions.

3.2.2 Klopman-Ohno type of QM/MM interaction scheme

The Klopman-Ohno (KO) formula was originally developed for evaluating s-orbitals interactions and later widely used in semi-empirical QM methods, such as MNDO, [152] as the damping function for two-center two-electron integrals. The original functional form is

$$\hat{H}_{elec,KO}^{QM/MM} = \sum_{\alpha I} \frac{\Delta q_\alpha Q_I}{\sqrt{R_{\alpha I}^2 + 0.25(1/U_\alpha + 1/U_I)^2}} \tag{3.6}$$

U_α is the Hubbard parameter which is related to chemical hardness η_α: $U_\alpha \approx I_\alpha - A_\alpha \approx 2\eta_\alpha$ and proportional to the atomic radii assuming a spherical charge density. [45] Therefore, the KO functional form allows an empirical damping of point charge interaction scheme in the short distance and effectively accounts for the deviations due to the increasing of electronic orbital interactions. When used in QM/MM framework, the MM "Hubbard" parameter is not well defined, although it can be taken from atomic electronic structure calculations or treated as a parameter similar to the width of the "Gaussian blur" in the approach introduced by Brooks and co-workers [154], or simply set to zero.

In this work, the KO functional form is further modified to include more flexibility,

$$H_{elec,KO}^{QM/MM} = \sum_{\alpha I} \frac{\Delta q_\alpha Q_I}{\sqrt{R_{\alpha I}^2 + a_\alpha (\frac{1}{U_\alpha(\Delta q_\alpha)} + \frac{1}{U_I})^2 e^{-b_\alpha R_{\alpha I}}}} \tag{3.7}$$

In this expression, $U_\alpha(\Delta q_\alpha)$ takes a linear relationship with atomic Mulliken charge via $U_\alpha(\Delta q_\alpha) = U_\alpha^0 + \Delta q_\alpha U_\alpha^d$ and U_α^d is Hubbard derivative with respect to atomic charge. For specific parametrization, see our previous work. [46] It is worth mentioning that by including the charge dependence into the Hubbard parameter, the modified KO functional form explicitly introduces the state dependence into the scaling of QM/MM interactions. The parameters a_α and b_α are based on element type so the current scheme only introduces two extra parameters for each element. With the inclusion of charge dependence into KO expression, the actually pair-wise functional form is determined self-consistently and can be adjusted with respect to different circumstances. Correspondingly, the SCC-DFTB interaction energy is slightly modified as

$$\begin{aligned} E^{SCC} &= \sum_i^{occ} \langle \phi_i | H^0 | \phi_i \rangle + \frac{1}{2} \sum_{A,B \in QM} \gamma_{AB} \Delta q_A \Delta q_B + \sum_{\substack{A \in MM, \\ B \in QM}} \gamma_{fit,AB} Q_A \Delta q_B \\ &\quad + \frac{1}{6} \sum_{A \in QM} \Delta^3 q_A U_A^d + E_{rep} \end{aligned} \tag{3.8}$$

where

$$\gamma_{fit} = \frac{1}{\sqrt{R^2 + a(\frac{1}{U(\Delta q)} + \frac{1}{U_A})^2 e^{-bR}}} \tag{3.9}$$

The matrix element also needs to be modified correspondingly as

$$\begin{aligned} H_{\mu\upsilon} &= H^0_{\mu\upsilon} + \frac{1}{2}S_{\mu\upsilon}\sum_{B\in QM}(\gamma_{CB}+\gamma_{DB})\Delta q_B + \frac{1}{2}S_{\mu\upsilon}\sum_{A\in MM}[(\gamma_{fit,AC}+\gamma_{fit,AD}) \\ &+(\frac{\Delta q_C \gamma^3_{fit,AC} U^d_C a_C e^{-b_C R_{AC}}}{U^3_C} + \frac{\Delta q_D \gamma^3_{fit,AD} U^d_D a_D e^{-b_D R_{AD}}}{U^3_D})]Q_A \\ &+\frac{1}{2}S_{\mu\upsilon}\sum_{A\in QM}\frac{\partial U_A}{\partial q_A}\Delta q^2_A \end{aligned} \tag{3.10}$$

where $\mu \in C; \upsilon \in D$

The force expression also needs to be modified accordingly.

Besides the improvement of electrostatic interactions, the vdW interactions in principle can also be made state dependent. For example, since the Hubbard parameter is directly related to the chemical hardness, including the charge dependence in the Hubbard parameter would also make the chemical hardness charge dependent. As discussed before, [107,155–158] the correlation between atom size and chemical hardness can be adopted as inversely related as $U = \frac{1}{R}$. Therefore, it is conceivable to use this relationship to make the radii of the vdW interaction charge dependent as well. However, the inclusion of charge dependence in vdW interactions requires extra work in the SCF calculations, therefore, can increase the computational overhead a lot based our test calculations. Alternatively, by adopting a set of element type dependent vdW parameters with the KO scheme, we are already able to achieve significant improvement compared with the conventional QM/MM interaction scheme. Thus, we leave the development of the state dependent vdW interactions as further work.

3.2.3 Parameter Optimization

To summarize, the new parameters in the KO interaction scheme are the a_i, b_i in Eq.3.7. In addition, the vdW parameters are made to depend only on the element type and hence need to be reparametrized. In principle, the MM Hubbard parameters can also be optimized to allow additional flexibility. In this work, we test two approaches: simple set the MM Hubbard parameters as zero which is referred as KO or use the atomic electronic structure calculation results which is referred as KO-MM. The solute-solvent (water in this work) interaction energy is used as the target property. Because our main interests are for condense phase performance which involves important multibody interactions, a cluster type of training set model is adopted in which we include the solute and all its nearby water, instead of the pair-wise training set model used in Ref. [57]. Based on our test, it is crucial to include the multibody interactions in parametrization as the pair-wise model fails to produce satisfactory parameter sets for solution reactions. The training set includes 23 molecules containing C, H, O, P, mimicking protein sidechains and phosphate species with various charge states. Each molecule in the training set is solvated with a water sphere of 25 Å radii, followed by 50 ps MD at 300 K from which 10 snapshots are taken out with even interval. For each snapshot, the solute and water molecules that are within 4 Å are kept while the rest are deleted to obtain the final cluster model with typically 15 water. The binding energy between solute and water molecules by full SCC-DFTB calculations serves as the reference. In particular, a special version of SCC-DFTB which is developed for phosphate hydrolysis reactions is used and referred as SCC-DFTBPR. In addition, a test set of 12 different molecules are also constructed via a similar fashion to evaluate the transferability of parameters in different QM/MM interaction schemes.

The parameters are optimized using a Genetic Algorithm (GA) [129] in which the "fitness" (ξ) is defined as the inverse of a weighted sum of difference between binding energies determined from full SCC-DFTB calculation and SCC-DFTB/MM calculation:

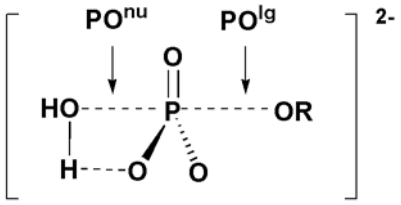
$$\xi^{-1} = \frac{\sum_{i=1} w_i[\Delta E_i^b(SCC) - \Delta E_i^b(QM/MM)]^2}{\sum_{i=1} w_i}, \tag{3.11}$$

where i is the index of species in the training set and the sum is over all molecules in the training set. During optimization, a micro-GA technique with a population of 10 chromosomes that is allowed to operate for 500 generations with uniform crossovers; see Ref. [130] for detailed descriptions and recommendations for GA options. For a fair comparison, we also reparametrized the vdW parameters via a similar fashion for the conventional QM/MM interaction scheme.

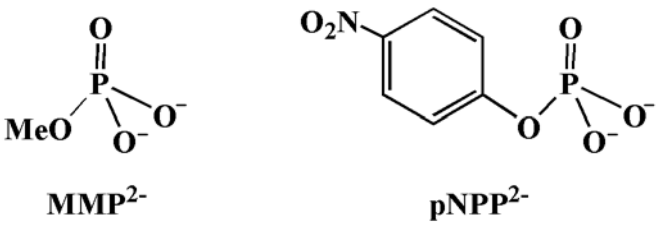
3.2.4 Potential of mean force (PMF) simulations for aqueous phosphate hydrolysis reactions

In order to evaluate the performance of different QM/MM interaction schemes for condense phase reactions, we study the aqueous hydrolysis reactions of two phosphate monoesters, methyl monophosphate^{2-} (MMP^{2-}) and p-nitrophenyl phosphate^{2-} ($pNPP^{2-}$) (see Fig. 3.1), with the water molecule as the nucleophile. These reactions serve as perfect examples for benchmark purpose as there are extensive previous experimental [159] and computational [4, 5] studies. In addition, these phosphate monoesters are the typical substrates of phosphatase, [27, 160] therefore the results also provide important reference for future enzyme studies.

The solute (MMP^{2-} or $pNPP^{2-}$) is solvated by the standard protocol of superimposing the system with a water droplet of 25 Å radius and removing water molecules within 2.8 Å from any solute atoms. [161] Water molecules are described with the TIP3P model [162] without any modifications. The QM region includes the solute and the nucleophile water. The generalized solvent boundary potential (GSBP) [124, 163] is used to treat long range electrostatic interactions in MD simulations. To be consistent with the GSBP protocol, the extended electrostatic model [164] is used to treat the electrostatic interactions among inner region atoms in which interactions beyond 12 Å are treated with multipolar expansions, including the dipolar and quadrupolar terms. The deformable boundary forces [165] are



(a)



(b)

Figure 3.1: The phosphate monoester dianions hydrolysis reactions studied in this work.

added in the boundary region to constrain water molecules within the sphere. An additional weak GEO type of potential is added the the QM region to keep it in the center of the water sphere. An angle constraint potential is added to the nucleophile water, the phosphate atom and the leaving group oxygen to guarantee the "in line" attacking. All bonds involving hydrogen in MM water are constrained using the SHAKE algorithm, [166] and the time step is set to 1 fs.

The 2D PMF calculations are carried out for the aqueous reactions. The whole system is optimized and slowly heated to 300 K and equilibrated for 50 ps. The reaction coordinate is defined as PO^{lg}-PO^{nu} and OH^{wat}-OH^{po}. The umbrella sampling approach [167] is used to constrain the system along the reaction coordinates. In total, more than 250 windows are used for each PMF and 50 ps simulations are performed for each window. The first 10 ps trajectories are discarded and only the last 40 ps are used for data analysis. Convergence of the PMF is monitored by examining the overlap of reaction coordinate distributions sampled in different windows and by evaluating the effect of leaving out segments of trajectories. The probability distributions are combined together by the weighted histogram analysis method (WHAM) [168] to obtain the PMF along the reaction coordinate.

As additional benchmarks focusing on the quality of the QM method rather than other technical details such as QM/MM coupling and sampling, we use a previous developed implicit solvent model [52] to study these aqueous reactions of phosphate monoesters. In this model, the solute radii are dependent on the charge distribution, which makes it particularly useful for studying solution reactions that involve highly charged species; our previous benchmark calculations suggest that the method has comparable accuracy as the SM6 model [89], while being much more efficient (due to the use of SCC-DFTB) and having only a small number of parameters. The reaction coordinates are similar to QM/MM simulations. Each point in the 2D PES is obtained by starting the constrained optimization from several different initial structures and taking the lowest energy value. The initial grid size is 0.2 Å due to the large number of points that need to optimize. Later a finer grid size (0.1 Å) is used to scan the TS region and locate the TS structure. Finally, frequency calculations are

Table 3.1: Optimized parameters for different QM/MM interaction schemes

	vdW opt[a]		KO				KO-MM[b]			
	ϵ	σ	ϵ	σ	a	b	ϵ	σ	a	b
O	-0.18	1.92	-0.03	2.05	0.068	0.017	-0.06	1.88	0.042	0.021
C	-0.06	2.15	-0.07	2.11	0.046	0.059	-0.05	2.15	0.026	0.069
P	-1.23	2.36	-0.52	2.39	0.060	0.001	-0.26	2.42	0.054	0.001
H	-0.02	0.82	-0.04	0.76	0.211	0.055	-0.02	0.81	0.066	0.053

a. Optimized vdW parameters for conventional QM/MM interaction scheme; b. KO scheme with MM Hubbard parameters included.

carried out to confirm the nature of the stationary points and to compute the vibrational entropy and zero point energies to obtain approximate activation free energy; although using a harmonic approximation to estimate activation entropy is known to be of limited accuracy, previous studies of phosphate diester hydrolysis found that activation entropy does not differ much between different diesters [6, 169].

To account for intrinsic errors of SCC-DFTBPR energies, we explore corrections based on gas phase single-point energy calculations with MP2/6-311++G** at SCC-DFTBPR geometries. As discussed in the literature, [135] such a simple correction may not always improve the energetics for semi-empirical methods given the errors in geometry; however, our previous tests [46, 52, 169] indicated that this correction scheme appears useful for SCC-DFTBPR since the method gives fairly reliable structures, even for transition states.

3.3 Results and Discussions

3.3.1 Cluster model binding energies in training set and test set

As the condense phase performance is the main concern, we adopt the cluster type of model to implicitly include the important multibody interactions. Our test indicates that using the pair-wise solute water model as in Ref. [57] fails to produce satisfactory parameters.

The training set (Table 3.2) includes 23 molecules and ions for amino acid sidechains and phosphate species with great chemical properties and various charge states. It is obvious to see that the binding energies of ions are typically one magnitude larger than those of neutral molecules, therefore we deliberately put more weights on ions.

Besides the conventional QM/MM interaction scheme, we optimize two sets of KO parameters with respect to different MM Hubbard parameters. In one set, we simply set MM Hubbard parameters as zero, thus, only consider the "size" of QM atoms; in the other set, we use atomic Hubbard parameters from electronic structure calculations. The three sets of schemes are referred as QM/MM, KO and KO-MM, respectively. As shown in Table 3.2, for the binding energies of the training set, KO-MM gives the best performance, consistent with the fact that a better physical picture is described by including MM "atomic size". Although KO gives only slightly better results than the conventional QM/MM interaction scheme with the set of optimized vdW parameters in this work, it obtains tremendous improvement compared with the results by the parameters in Ref. [57] which gives the Mean Unsigned Error (MUE) as 14.1 kcal/mol. Therefore the current optimization protocol, together with the KO interaction scheme, can significantly improve the computational accuracy for the model cluster in the training set.

To test the transferability of the parameter sets, we also construct a test set with twelve molecules that are not included in training set (see Table 3.3). Similar to the results of training set, KO-MM gives the best performance while the conventional QM/MM produces the largest error. It is very encouraging to see that the performance of KO and KO-MM do not deteriorate compared to that for the training set, indicating good parameter transferability in KO scheme. On the contrary, the MUE increases drastically for the conventional QM/MM scheme, cautioning the fact that although it is possible to obtain specific parameters for given problems, the transferability of those parameters is questionable.

In addition, we test the performance for phosphate hydrolysis reactions by 10 RNA model reactions from QCRNA database established by York group. [170] These includes 16 stable states and 24 transition states. The similar cluster model is constructed with fixed solute

Table 3.2: Error (in kcal/mol) analysis of binding energies for training set

		Unsigned Error[a]		
Solute	E_{SCC}	KO	KO-MM	vdW_{opt}
Propane	-0.6	4.3	3.8	3.4
Isobutene	-1.6	5.4	5.0	4.6
Butane	-1.1	5.4	4.7	4.5
Toluene	-3.5	5.6	4.5	4.5
4-cresol	-10.5	2.6	2.5	2.4
Methanol	-9.2	2.1	1.9	2.5
Ethanol	-9.0	1.5	1.3	2.0
Acetaldehyde	-6.7	1.3	1.1	1.8
Methylacetate	-9.3	1.9	1.6	2.8
Acetic acid	-7.6	1.5	1.2	2.7
Propanic acid	-15.8	4.4	3.7	5.6
Dimethyl ether	-4.2	1.6	1.3	2.0
Methylphosphate	-21.9	7.4	5.3	8.6
Dimethylphosphate	-15.9	2.8	3.3	5.2
Acetate (-1)	-78.4	3.0	2.8	6.8
Propanate (-1)	-88.3	2.5	2.7	2.1
4-cresol (-1)	-83.2	6.5	5.8	7.0
Methoxide (-1)	-94.0	5.3	3.7	7.4
Ethoxide (-1)	-95.5	6.7	4.1	10.0
Hydroxide (-1)	-68.6	8.8	6.9	5.4
Methylphosphate (-1)	-84.1	6.4	5.3	5.5
Dimethylphosphate (-1)	-79.5	3.6	1.8	5.5
Methyl phosphate (-2)	-249.6	7.6	5.6	8.6
Error Analysis[b]				
MUE		4.3	3.3	4.8
MSE		-0.5	-0.8	-0.8

a. The unsigned error is averaged over 10 snapshots for each solute; b. MUE: mean unsigned error; MSE: mean signed error.

Table 3.3: Error (in kcal/mol) analysis of binding energies for test set[a]

		Unsigned Error		
Solute	E_{SCC}	KO	KO-MM	vdW_{opt}
Methane	-0.9	0.9	0.9	0.8
Phenol	-6.3	4.8	4.2	4.7
Propanol	-10.8	4.2	3.7	4.5
Formaldehyde	-3.1	1.2	1.0	1.6
Formic acid	-21.6	2.0	1.6	3.2
Trimethyl phosphate	-21.0	7.2	6.4	12.0
Formate (-1)	-93.6	5.6	1.7	10.8
Benzoate (-1)	-94.1	10.7	5.3	15.1
Propanoate (-1)	-96.5	6.8	3.1	10.9
Dihydrogen phosphate (-1)	-105.4	2.6	2.6	9.0
Methyl phenyl phosphate (-1)	-110.5	9.0	5.8	17.0
Hydrogen phosphate (-2)	-271.5	2.1	5.9	17.1
Error Analysis[b]				
MUE		4.7	3.5	8.9
MSE		-4.2	-2.0	-8.8

a. The unsigned error is averaged over 10 snapshots for each solute; b. MUE: mean unsigned error; MSE: mean signed error.

geometries obtained in gas phase reactions. The results (Table 3.4) indicate that KO-MM gives the best performance while KO is slightly worse. The conventional QM/MM interaction scheme results in quite large errors.

Table 3.4: Energetics Benchmark Calculations for different QM/MM interaction schemes based on 10 phosphate reactions from the QCRNA database[a]

Reactions	States	SCC binding	Errors		
			QM/MM	KO	KO-MM
CH3O...P(O)(O)(OH)(OCH3)	ts12	-255.0	28.0	14.0	2.4
HO...P(O)(O)(OH)(OCH3)	ts12	-252.3	25.8	11.8	4.2
HO...P(O)(OH)(OH)(OCH3)	ts12	-96.7	12.0	3.5	5.8
	min2	-100.5	10.5	3.1	3.6
	ts23	-98.2	10.6	3.1	3.4
HOH...P(O)(O)(OCH3)(OCH3)	min1	-99.3	11.2	3.6	5.2
	ts12	-102.1	12.0	3.4	6.6
	min2	-98.7	12.0	3.9	7.3
	ts23	-100.2	10.8	2.8	6.2
	min3	-109.1	10.9	3.1	2.5
	ts34	-96.5	13.3	5.1	2.2
	min4	-98.4	11.3	3.8	2.0
	ts45	-98.6	14.4	6.4	1.9
	min1	-94.5	11.6	4.0	2.3
HO...P(O)(O)(OCH3)(OCH3)	ts12	-257.6	23.0	7.6	2.1
	min2	-274.0	26.3	11.4	2.3
	ts23	-267.2	24.5	10.4	1.7
CH3O...P(O)(O)(OCH3)(OCH3)	ts12	-244.1	32.1	18.9	2.8
CH3O...P(O)(OH)(OH)(OCH3)	ts12	-102.0	13.1	4.8	1.9

	min2	-93.3	13.2	4.9	3.9
	ts23	-99.6	14.4	6.0	2.6
CH3O...P(O)(OH)(OCH3)(OCH3)	ts12	-94.8	16.6	9.1	3.3
	min2	-98.1	13.2	5.2	6.0
	ts23	-107.4	16.1	7.6	4.0
	min3	-96.4	13.0	5.5	12.1
	ts34	-103.5	14.4	6.5	2.4
CH3O...P(O)(OCH3)(OCH3)(OCH3)	ts12	-97.9	18.9	10.5	2.6
	min2	-106.2	16.5	8.7	3.6
	ts23	-103.5	17.2	9.4	6.4
	min3	-99.4	16.9	9.0	2.6
	ts34	-92.0	17.4	10.0	4.5
	min4	-103.3	14.6	7.1	3.4
	ts45	-102.8	19.5	11.3	3.9
HO...P(O)(OCH3)(OCH3)(OCH3)	ts12	-113.8	19.3	2.7	6.9
	min2	-94.6	15.4	7.2	5.4
	ts23	-96.9	16.5	8.5	6.1
	min3	-101.0	14.6	6.3	5.6
	ts34	-96.6	13.8	6.0	6.7
	min4	-98.3	16.5	8.0	4.9
	ts45	-96.5	17.6	9.4	7.4
Error Analysis[b]					
Overall Performance	MUE		16.2	7.1	4.3
Stable States Performance	MUE		14.2	5.9	3.5
Transition States Performance	MUE		17.6	7.9	4.8

a. The unsigned error is averaged over 10 snapshots for each solute; b. MUE: mean unsigned error; MSE: mean signed error.

3.3.2 PMF for phosphate monoester reactions

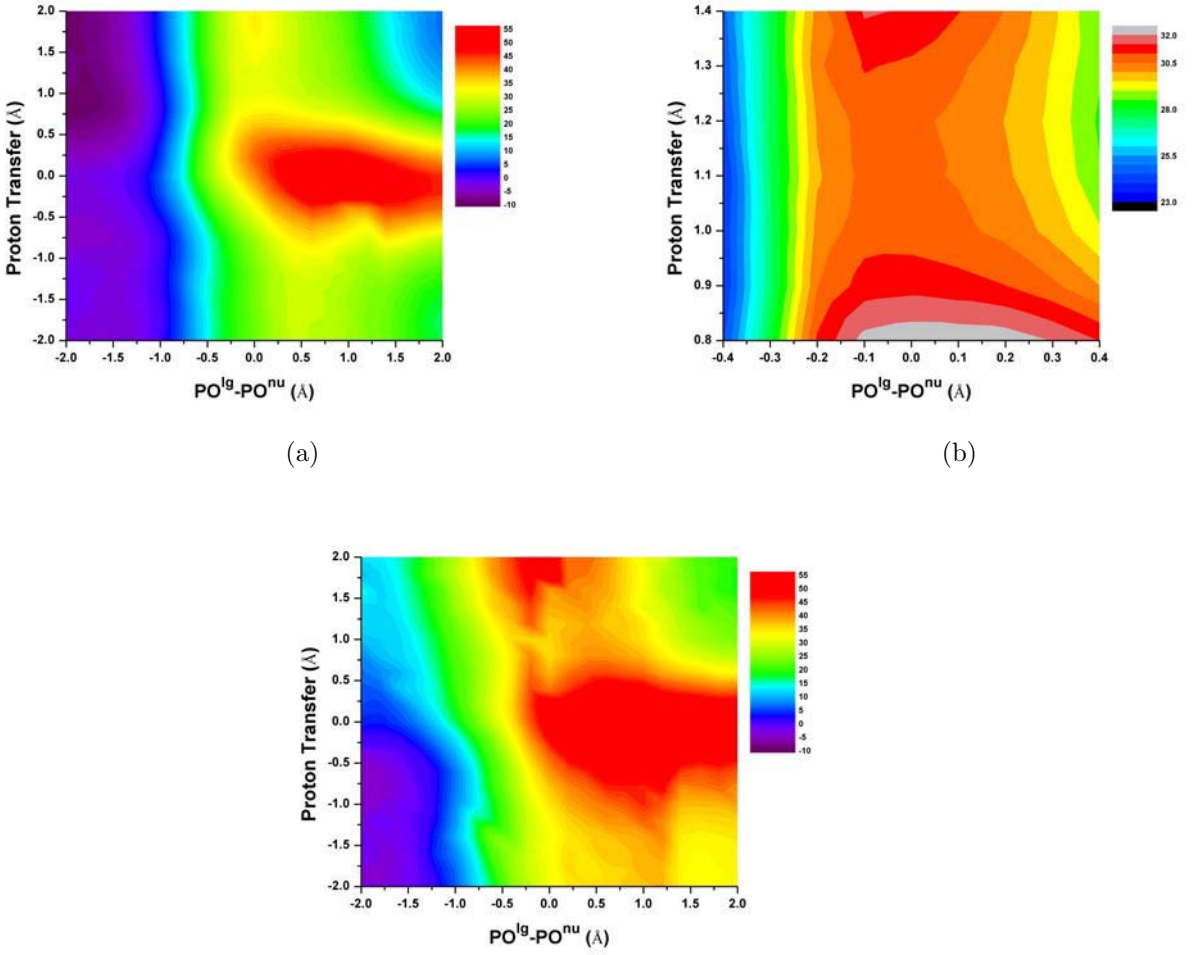
3.3.2.1 MMP^{2-} hydrolysis reaction

Since our goal is to use the KO scheme for condense phase chemical reactions, it is necessary to investigate its performance for more realistic systems other than the cluster type of model. MMP^{2-} is a simple phosphate monoester and its solution reaction has been extensively studied by experimental and computational methods. In aqueous, the nucleophile has been determined as a water molecule and the experimental free energy barrier is 44.3 kcal/mol at 298K calculated by transition state theory. [159] Computationally the barrier has been well reproduced as 47 kcal/mol at 312 K by B3LYP/COSMO model. [5] The calculated transition state structure indicates that the water first transfers a proton to MMP^{2-} to become a hydroxide that further attacks the protonated phosphate monoester. The P-O^{lg} P-O^{nu} bond lengths are 1.8 and 2.0 Å, respectively.

Before studying this reaction by different QM/MM interaction schemes, it is crucial to establish the intrinsic error of the QM method in use, i.e., dissecting the errors in QM/MM simulations from the QM method and from the QM/MM framework. For this purpose, we use a recently developed implicit solvent model that combines SCC-DFTB method with Poisson-Boltzmann (PB) and a set of charge dependent radii. [52] It has been demonstrated that the SCC-DFTB/PB model can describe the aqueous reactions for highly charged species comparable to SM6 method. [89] More importantly, by using the implicit solvent model, we can avoid the potential sampling issue in QM/MM simulations to quantify the inherent errors from the QM method. In the calculated potential energy surface shown in Fig. 3.2(a), the reactant state corresponds to the bottom left corner while the product state corresponds to the upper right corner. The first step involves an exothermic proton transfer reaction from the water to MMP^{2-}, followed by the nucleophilic attacking, which is consistent with the picture in previous studies. Rescanning the TS region by a finer grid size, the reaction barrier

is estimated as 30.5 kcal/mol. By adding entropy effects and zero point energy corrections at 300 K, the free energy barrier is 39.5 kcal/mol which agrees with previous studies. The calculated transition state has the reaction coordinate of PO^{lg}-PO^{nu} as 0.0 Å and the proton transfer coordinate as 1.1 Å. The PO^{lg} and PO^{nu} bond lengths are both 1.95 Å, also consistent with previous theoretical studies. Based on our experience, adding MP2 single point energy corrections can usually improve the accuracy of SCC-DFTB/PB. Indeed, the reaction barrier is further refined to 45.7 kcal/mol and the overall PES landscape (Fig. 3.2(c)) is similar to that of SCC-DFTB/PB. However, it is obvious that SCC-DFTB(PR) systematically underestimates the relative energy compared with infinite separated reactants, especially for the upper left corner which corresponds to the exothermicity of the proton transfer step. It has been noted before that SCC-DFTB(PR) can be problematic for calculating the proton affinity of phosphate species and the inclusion of full third order terms in principle can improve the results. [46, 153] Overall, the current SCC-DFTB(PR) method can describe the MMP^{2-} hydrolysis reaction accurately although the description of the exothermicity of the first proton transfer process is less satisfactory.

We further study this reaction by QM/MM simulations with the conventional QM/MM interaction scheme (QM/MM), KO and KO-MM schemes (see Fig.3.3 and Table 3.5 for details). By using the conventional QM/MM scheme with optimized vdW parameters, the reaction barrier is calculated as 55 kcal/mol, which is about 10 kcal/mol higher than the experimental value. With the KO or KO-MM schemes, the results are improved to be 41 and 40 kcal/mol, respectively. One point worth mentioning is that for the conventional QM/MM scheme, the first proton transfer step is exothermic while for KO and KO-MM it becomes endothermic by a magnitude of 10 kcal/mol. Warshel and coworkers studied this step [4] by MP2/LD method and obtained an endothermic reaction with 9 kcal/mol difference. As noted above, the SCC-DFTB(PR)/PB model (Fig. 3.2) has quite large errors in this region due to the QM method. Since the overall performance of QM/MM interactions relies on the QM method, this error is inherited in all three QM/MM schemes. However, the error cancellations in KO and KO-MM schemes partially compensate for the intrinsic errors in



(a) (b)

(c)

Figure 3.2: Potential energy surface (PES) of MMP^{2-} hydrolysis reaction (kcal/mol). (a) 2D PES of MMP^{2-} hydrolysis reaction by SCC-DFTB(PR)/PB; (b) 2D PES of the TS region with a finer grid size by SCC-DFTB(PR)/PB; (c) 2D PES by adding MP2/6-311++G** single point energy corrections.

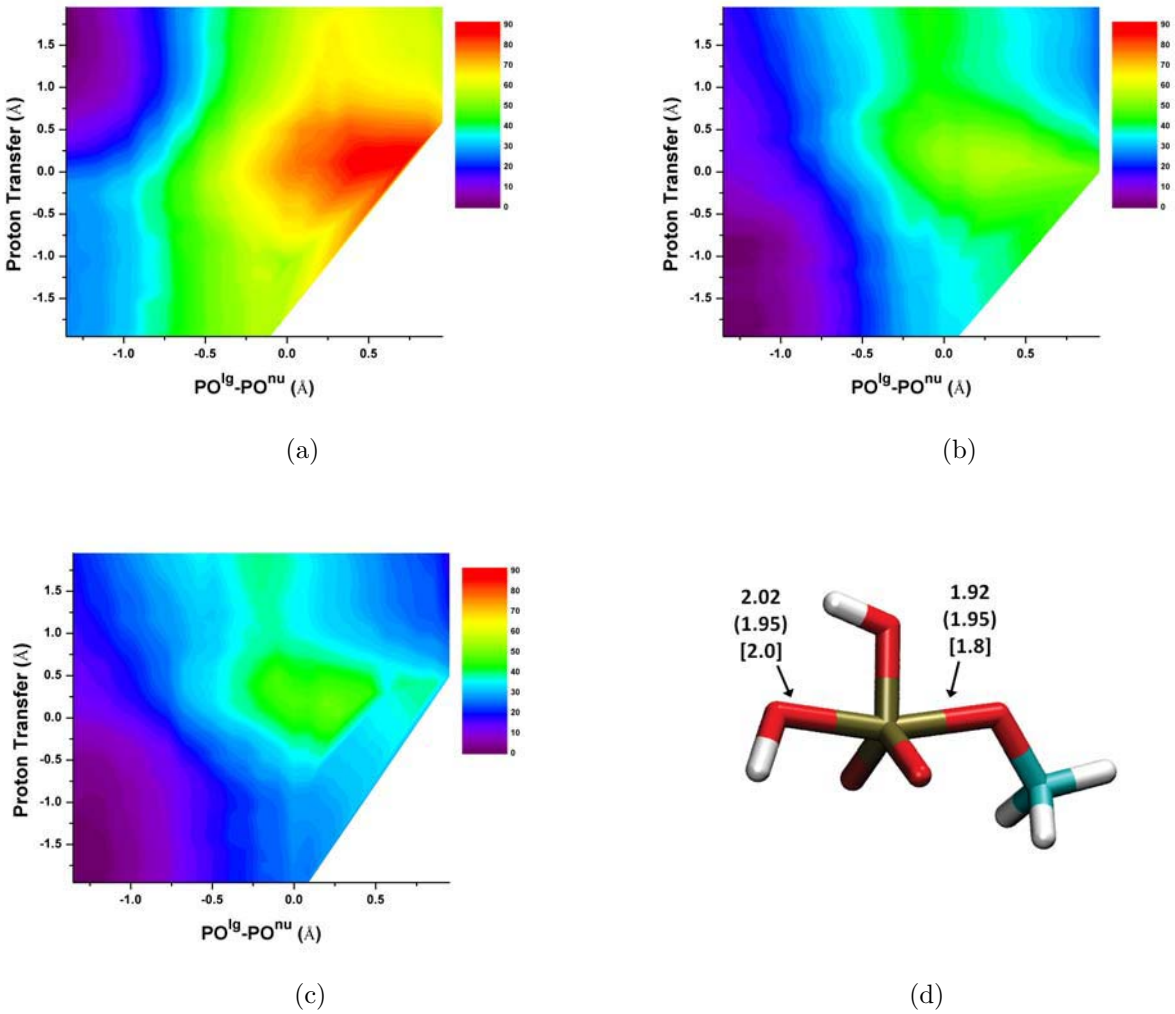


Figure 3.3: 2D PMF of MMP^{2-} hydrolysis reaction by different QM/MM interaction schemes (kcal/mol). (a) Conventional QM/MM scheme with optimized vdW parameters; (b) KO scheme; (c) KO-MM scheme ; (d) The transition state structure. The numbers without parenthesis are calculated by KO-MM, with parenthesis are calculated by SCC-DFTB(PR)/PB, with bracket are taken from Ref. [5].

SCC-DFTBPR method, resulting in a better description. The calculated free energy surface indicates a similar reaction mechanism as in previous studies: the proton transfer takes place first, followed by the nucleophilic attacking. The transition state region calculated by KO scheme is at reaction coordinate PO^{lg}-PO^{nu} slightly less than 0 Å and proton transfer coordinate at 1.2 Å. The averaged PO^{lg} and PO^{nu} bond lengths are 1.94 and 2.04 Å, similar to those in previous studies [5] and the SCC-DFTB(PR)/PB model.

3.3.2.2 $pNPP^{2-}$ hydrolysis reaction

In addition to MMP^{2-}, we also study another phosphate monoester, $pNPP^{2-}$, which has quite different ester group. $pNPP^{2-}$ is an important substrate for phosphatase studies, therefore the aqueous results provide important reference for enzyme studies. Since we do not have KO parameters for nitrogen, the parameters of oxygen are used instead and the effects are expected to be small.

Similar to the MMP^{2-} reaction, we also use SCC-DFTB(PR)/PB method to estimate the inherent error in the QM method. The overall potential energy landscape (Fig. 3.4(a)) is similar to that of MMP^{2-} in which the phosphate nonbridging oxygen first abstracts a proton from the water and then a nucleophilic attacking follows. The free energy barrier is calculated as 29.3 kcal/mol after adding entropic effects and zero point energy corrections and further refined to 27.0 kcal/mol after adding MP2 single point energy corrections. The experimental value is 31.8 kca/mol at 298 K which is in decent agreement with our results. The transition state structure (Fig. 3.4(b)) has the reaction coordinate PO^{lg}-PO^{nu} as -0.3 Å and the proton transfer coordinate as 1.0 Å. The PO^{lg} bond length is 1.95 Å similar to that of MMP^{2-}; however, PO^{nu} bond length increases to 2.26 Å and the overall transition state structure becomes looser than that of MMP^{2-} (described by the sum of PO^{nu} and PO^{lg}). These observations are also consistent with the trend in previous studies that the transition state changes from associative to dissociative upon decrease in the pKa of the leaving group. [5] Therefore, the SCC-DFTBPR method is able to describe this reaction at the satisfactory level.

Table 3.5: Free energy barriers (kcal/mol) of phosphate monoester hydrolysis reactions by different methods[a]

Solute	Method	$\Delta G^{\ddagger}$
MMP^{2-}	Exp[b]	44.3 (298K)
	MP2/LD[c]	43 (312K)
	SCC/PB[d]	39.5/45.7
	QM/MM opt	55
	KO	41
	KO-MM	40
$pNPP^{2-}$	Exp[e]	31.8 (298K)
	SCC/PB[d]	29.3/27.0
	KO	33
	KO-MM	32

a. All results are under 300 K unless noted otherwise; b. Results taken from Ref. [159]; c. Results taken from Ref. [5]; d. The number before slash is SCC-DFTB/PB result with entropic and ZPE corrections; the number after slash is with MP2/6-311++G** single point corrections; e. Results taken from Ref. [171].

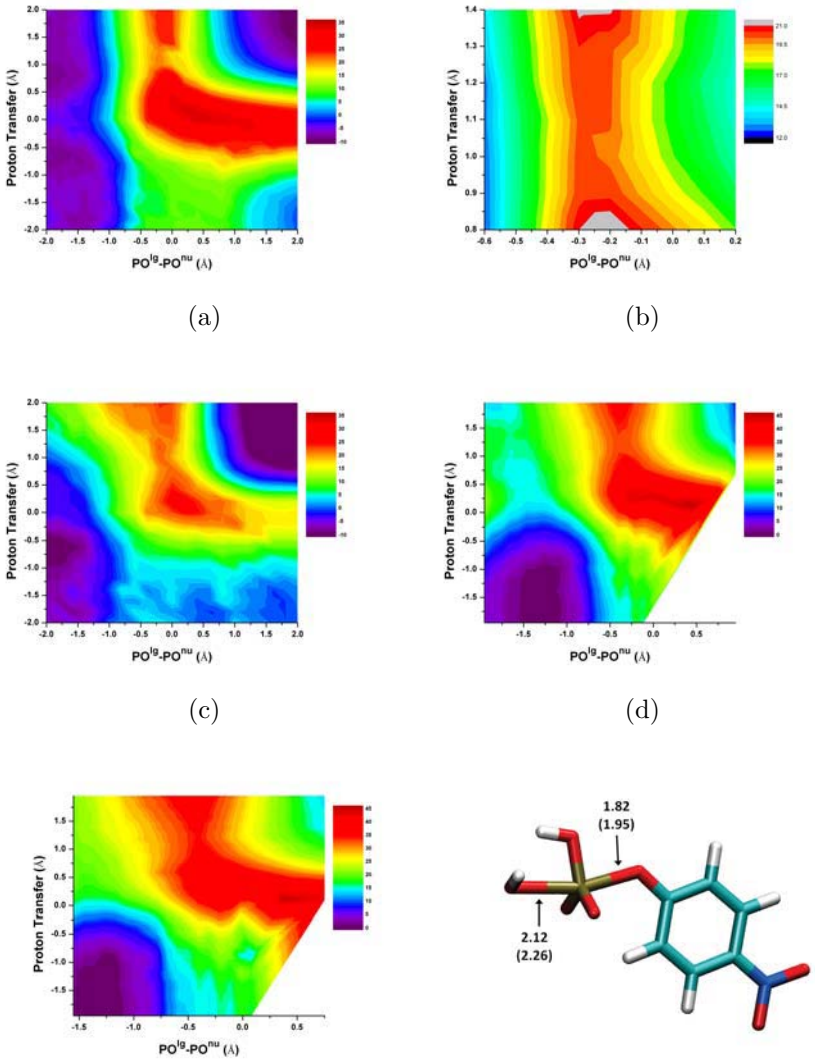


Figure 3.4: 2D potential energy surface (PES) and potential of mean force (PMF) of $pNPP^{2-}$ hydrolysis reaction (kcal/mol) by SCC-DFTB(PR)/PB and QM/MM KO scheme. (a) 2D PES for $pNPP^{2-}$ hydrolysis reaction by SCC-DFTB(PR)/PB; (b) 2D PES for the transition state region with a finer grid size by SCC-DFTB(PR)/PB; (c) 2D PES by adding MP2/6-311++G** single point energy corrections; (d) 2D PMF of $pNPP^{2-}$ hydrolysis reaction by KO scheme; (e) 2D PMF of $pNPP^{2-}$ hydrolysis reaction by KO-MM scheme; (f) The transition state structure. The numbers without parenthesis are by KO-MM, with parenthesis are by SCC-DFTB(PR)/PB.

We also apply the KO and KO-MM schemes to obtaining the 2D PMF for this reaction (see Fig. 3.4(c),(d)). The calculated reaction barrier is 33 kcal/mol which is very close to the experimental results. The transitions state structure is also consistent with the SCC-DFTB(PR)/PB results. The averaged PO^{lg} and PO^{nu} bond lengths are 1.84 and 2.14 Å, respectively, also in decent agreement with the trend calculated by SCC-DFTB(PR)/PB and previous theoretical studies.

3.4 Concluding remarks

QM/MM protocol has been demonstrated as an effective approach of balancing the computational accuracy and cost for condense phase chemical reactions, therefore widely used in studying enzyme catalysis. Usually semi-empirical type of QM methods are used due to the demanding requirements of the problem size and time scale. However, the lack of one electron integral and predetermined vdW parameters significantly undermine its performance for systems that involve large amount of charge redistribution. Although it may be possible to develop specific parameter sets for a given problem, the conventional QM/MM scheme lacks general flexibility and parameter transferability, therefore requires further improvement.

In this study, we develop a state-dependent QM/MM interaction scheme based on the Klopman-Ohno functional form. The major part of the state-dependence is accounted by the damped electrostatic interactions that is correlated to the "atomic size" via Hubbard parameters. With careful parametrization with respect to condense phase properties, the accuracy of QM/MM interactions can be significantly improved for highly charged systems, making it especially useful for studying phosphate hydrolysis in biological systems. The extensive benchmarks for training and test sets and an independent set constructed from QCRNA database demonstrate its good performance for both stable state and transition state, which is crucial for producing reliable results for chemical reactions. The element type dependent parameters significantly simply the algorithm and result in good parameter

transferability. With the KO scheme, we study the aqueous hydrolysis reactions of two phosphate monoesters, MMP^{2-} and $pNPP^{2-}$, and achieve decent agreement with experimental energetic data and previous high level theoretical results.

Besides the general success of the KO scheme, our work also indicates several limitations that need to be better addressed in the future work. The first and foremost is that the quality of the QM method directly affects the overall performance, as demonstrated in the aqueous phosphate hydrolysis studies. Therefore, the further improvement of SCC-DFTB method is imperative which includes the full third order expansion [153] and systematic reparametrization for phosphate hydrolysis. In addition, as the current work focuses on the collective condense phase properties, e.g., the parametrization implicitly takes the multi-body interactions into account, its performance for individual interactions can be compromised. Hence, we caution that the QM/MM boundary still needs to be carefully selected to avoid cutting any important specific interactions. Last but not least, the parametrization of KO scheme is subject to a few factors that can limit its performance. For example, the cluster models are taken from the snapshots produced by conventional QM/MM scheme. Although our tests indicate that the effects are negligible, bias can exist for the cluster configurations. Moreover, the reference is chosen as full SCC-DFTB(PR) method which may also limit the overall accuracy. However, an estimation of the quantitative influence requires extensive benchmarks which we leave for further work.

Chapter 4

QM/MM analysis suggests that Alkaline Phosphatase (AP) and Nucleotide pyrophosphatase/phosphodiesterase slightly tighten the transition state for phosphate diester hydrolysis relative to solution: implication for catalytic promiscuity in the AP superfamily

4.1 Introduction

Although a high-level of catalytic specificity has been regarded as an important hallmark of enzymes, it is increasingly recognized that many enzymes have promiscuous catalytic activities. [7–11] Moreover, it has been proposed that catalytic promiscuity plays an important role in enzyme evolution since it can give an enzyme an evolution "head start", providing a modest rate enhancement that is sufficient as a selective advantage [7, 12, 13, 172, 173]. Therefore, identifying factors that dictate the level of catalytic promiscuity in enzymes can help better understand enzyme evolution and improve design strategies for evolving new catalytic functions.

In this context, members in the Alkaline Phosphatase (AP) superfamily present striking examples of catalytic specificity and promiscuity. [25, 26] They have been demonstrated to catalyze the hydrolysis of a broad range of substrates that differ in charge, size, intrinsic reactivity and transition state (TS) nature [174]. For example, *E. coli* AP catalyzes the hydrolytic reaction of phosphate monoesters for its physiological function but also exhibits promiscuous activity for the hydrolysis of phosphate diesters and sulfate esters of diverse

structural/chemical features. Similarly, although the main function of Nucleotide pyrophosphatase/phosphodiesterase (NPP) is to hydrolyze phosphate diesters, it can also cleave phosphate monoesters and sulfate esters with considerable acceleration over solution reactions. The catalytic proficiencies (defined by the ratio of k_{cat}/K_M and rate of the uncatalyzed reaction in solution, k_w) vary greatly, ranging from $> 10^{27}$ for the cognate activity [28,175] to $\sim 10^6$ for the promiscuous activity [176]. The reaction specificities of AP and NPP (characterized by ratios of k_{cat}/K_M for cognate and promiscuous substrates in the two enzymes) for phosphate mono- and di-esters differ by up to a remarkable level of 10^{15} fold. [27, 28] These significant levels of catalytic specificity and promiscuity are particularly striking in light of the fact that AP and NPP are very similar in their active site features yet have limited sequence identity (8%): as illustrated in Fig.4.2, both AP and NPP feature a highly conserved bi-metallo zinc active site with the same set of metal ligands (three Asp and three His residues). These characteristics make this pair of enzymes ideal for in-depth comparative analyses, i.e., to understand how they combine the high levels of catalytic proficiency and promiscuity by making use of similar active sites.

Extensive work has been carried out to characterize the structure and function of AP and NPP. Crystal structures [27, 177] show that, in spite of the similarities, several differences can be noted between these enzymes. First, the AP active site has additional positively charged motifs, in particular a magnesium ion and Arg166; these are replaced in NPP by charge-neutral residues, Thr205 and Asn111, respectively. The extra positive charges in AP likely help stabilize phosphate monoesters over diesters due to difference in the charge states of these substrates (dianionic vs. monoanionic). Second, the NPP active site is featured with additional hydrophobic residues (e.g., Leu123, Phe91 and nearby residues, see Fig.4.2), which are expected to help bind diesters more tightly than monoesters. Motivated by these observed differences, systematic analyses over the last few years have helped quantitatively account for a significant fraction of the 10^{15} fold differential catalytic specificity in AP and NPP: [27,28] (1) Arg166 in AP interacts favorably with two negatively charged nonbridging

phosphoryl oxygen atoms present in phosphate monoesters but not diesters, giving a preference to monoesters of $\sim 10^4$ fold; (2) the hydrophobic R' binding pocket of NPP provides $\sim 10^4$ fold preference to diester catalysis; (3) the Mg^{2+} site in AP contributes through water-mediated hydrogen-bonding interaction with the transferred phosphoryl group, which bears less negative charge in the case of diesters, to favor the monoester reaction by a $\sim 10^4$ fold.

Despite progress, crucial questions remain to be answered for catalysis in AP and its superfamily members. In particular, a fundamental hypothesis regarding catalytic promiscuity in AP/NPP, which was motivated by experimental linear free energy relation (LFER [178]) and kinetic isotope effect (KIE) data [29–33], is that AP and NPP do not alter the nature of phosphoryl transfer TS relative to solution reactions, instead they recognize and stabilize TSs of different nature for cognate and noncognate substrates. This property has been proposed to assist in the evolutionary optimization of promiscuous activities and challenges the traditional notion that an enzyme active site is evolved to stabilize a single type of TS. Recent QM/MM calculations [1, 2, 179] using the AM1(d)-PhoT method [180] as the QM level, however, do not seem to support this model. Although the calculations found that phosphate monoester hydrolysis in AP proceeds via a loose TS [179], similar to in aqueous solution, the TS for phosphate diester was found to change from synchronous in solution to very loose in both NPP [1] and AP [2]. The latter is in contrast to conclusions from LFER analysis for phosphate diester hydrolysis in AP [32], in which the TS is determined to be synchronous, similar to its solution reference. Nevertheless, citing the previous discussion [6] of ambiguity in using LFER data to infer the structure of TS, the authors of Ref. [2] proposed a picture for the evolution (and catalytic promiscuity) of the AP superfamily in which the nature of the TS (loose) is maintained for different substrates (e.g., mono- and di-esters) [2]; this scenario has been established to explain catalytic promiscuity observed for protein phosphatase-1 [181]. It should be noted, however, that whether the computational method was sufficiently reliable in the recent QM/MM studies is not clear; for example, the Zn^{2+}-Zn^{2+} distance was found to vary greatly during the reaction for both mono- and di-ester substrates in AP and NPP [1, 2, 179], reaching 7.0 Å as compared to the value of

$\sim$4 Å in the crystal structure [27, 182] and other structural characterizations (Lassila and Herschlag, private communications).

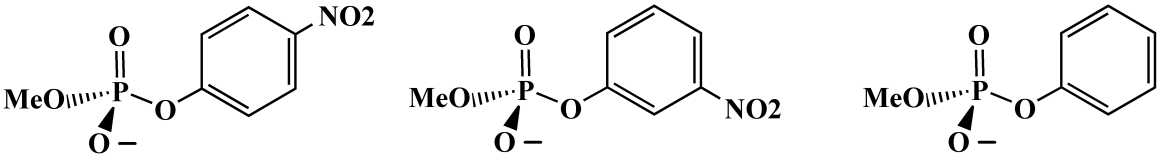


Figure 4.1: Methyl p-nitrophenyl phosphate ($MpNPP^-$) and its two diester analogs studied in this work.

To help clarify the situation, we set out to use combined QM/MM potential to systematically investigate the hydrolysis reaction of various cognate/noncognate substrates of AP/NPP in solution and enzymes. In this paper, we focus on the hydrolysis reactions for the same diester substrate studied in previous QM/MM calculations [1, 179], $MpNPP^-$ (Fig.4.1), and its phosphorothioate analog ($MpNPPS^-$), in solution, two experimentally well-characterized variants of AP (R166S and R166S/E322Y), and the wild type NPP. Since the active sites of AP and NPP are fairly open and readily accessible to solvents (which is what made it possible to carry out LFER studies for these systems), conformational sampling is expected to be crucial. This consideration together with the fairly large size of the bi-metallic zinc catalytic center suggest that an appropriate approach is to use the Self-Consistent-Charge Density-Functional-Tight-Binding (SCC-DFTB) [45] as QM in a QM/MM framework. With respect to computational efficiency, SCC-DFTB is comparable to widely used semi-empirical methods such as AM1 and PM3 [183], i.e., it is 2-3 orders of magnitude faster than popular DFT methods. In terms of accuracy, fairly extensive benchmark calculations have indicated that it is particularly reliable for structural properties, while energetics are generally comparable to AM1 and PM3. [103–105] For phosphoryl transfer reactions, however, a reaction-specific parameterization based on hydrolysis reactions of model phosphate

species, referred to as SCC-DFTBPR, [46] appears to be more effective than standard semi-empirical methods and has been found successful in several applications to solution and enzyme systems. [47–49]

Here we further test the reliability of SCC-DFTBPR for $MpNPP^-$ in different environments (solution, AP and NPP) by comparing results to higher-level QM (QM/MM) calculations as well as available experimental data. A more systematic comparison with LFER and KIE data requires much more extensive calculations and is left as a separate study. Nevertheless, the encouraging benchmark results obtained so far suggest that the SCC-DFTBPR based QM/MM approach can be used to probe the nature and energetics of phosphoryl transfer TS in AP and NPP at a semi-quantitative level. In contrast to recent QM/MM calculations [1, 179], which found a much looser TS in NPP than in solution, the results here support that the nature of the phosphoryl transfer TS for phosphate diesters is not loosened in neither AP nor NPP relative to solution; in fact, the TS becomes slightly tighter in AP and NPP than in solution, due in part to the geometry of the bimetallic zinc motif. Therefore, our study highlights the importance of using a carefully benchmarked QM/MM model to investigate the nature of phosphoryl transfer TS; moreover, these data provide the first explicit computational support of the hypothesis that the nature of TS for the same substrate is similar in the AP family and in solution.

The paper is organized as follows: in Sect.4.2 we summarize computational methods and simulation setup. In Sect.4.3, we first present results for the reference solution reactions, and then analysis of the phosphoryl transfer TS for phosphate diesters in several variants of AP and wild type NPP; we also analyze the effect of thio substitution of the diester substrate, which was used experimentally to probe the orientation of the substrate in the active site. Since the Zn^{2+}-Zn^{2+} distance exhibits rather different behaviors in this and previous QM/MM simulations of AP/NPP [1, 179], we also explicitly analyze the impact of this fundamental geometrical feature of the bimetallic zinc site on the catalysis. Before concluding in Sect.4.4, we summarize the key differences between our and recent QM/MM

studies [1, 2, 179] and also a number of issues that we recommend to examine by future experiments.

4.2 Computational Methods

4.2.1 Diester hydrolysis in solution with the SCC-DFTBPR based implicit solvent model

As an important benchmark and reference, we first study the hydrolysis of $MpNPP^-$ and two of its analogs (Fig.4.1) in solution, which have been thoroughly studied experimentally. To focus on the quality of the QM model rather than other technical details such as QM/MM coupling and sampling, we use the implicit solvent model that we have implemented and parameterized for SCC-DFTBPR. [52] In this model, the solute radii are dependent on the charge distribution, which makes it particularly useful for studying solution reactions that involve highly charged species; our previous benchmark calculations suggest that the method has comparable accuracy as the SM6 model [89], while being much more efficient (due to the use of SCC-DFTB) and having only a small number of parameters.

The potential energy surface (PES) relevant to the hydrolysis reaction is first explored by adiabatic mapping calculations, in which the reaction coordinates are the P-O^{lg} and P-O^{nu} distances; here hydroxide is the nucleophile, and "O^{lg}" and "O^{nu}" indicate the reactive oxygen in the leaving group and nucleophile, respectively. Each point in the 2D PES is obtained by starting the constrained optimization from several different initial structures and taking the lowest energy value. Following the adiabatic mapping calculations, structures along the approximate reaction path are examined carefully to ensure that the change of geometry is continuous along the path; subsequently, the saddle point is fully optimized by conjugated peak refinement (CPR) [133] to obtain more precise TS structure and energy. Finally, frequency calculations are carried out to confirm the nature of the stationary points and to compute the vibrational entropy and zero point energies to obtain approximate activation free energy; although using a harmonic approximation to estimate activation entropy is known to be of limited accuracy, previous studies of phosphate diester hydrolysis found

that activation entropy does not differ much between different diesters [6]. The vibrational frequencies are also used to estimate ^{18}O kinetic isotope effects (KIEs) for $MpNPP^-$ as an additional benchmark of the methodology (see **Supporting Information**).

To correct for intrinsic errors of SCC-DFTBPR energies, we explore corrections based on gas phase single-point energy calculations with both B3LYP/6-311++G** and MP2/6-311++G** at SCC-DFTBPR geometries; the B3LYP level was found to give very similar results to MP2 for simple phosphate hydrolysis reactions [46] (however, see below). As discussed in the literature, [135] such a simple correction may not always improve the energetics for semi-empirical methods given the errors in geometry; however, our previous tests [46, 52] indicated that this correction scheme appears useful for SCC-DFTBPR since the method gives fairly reliable structures, even for transition states.

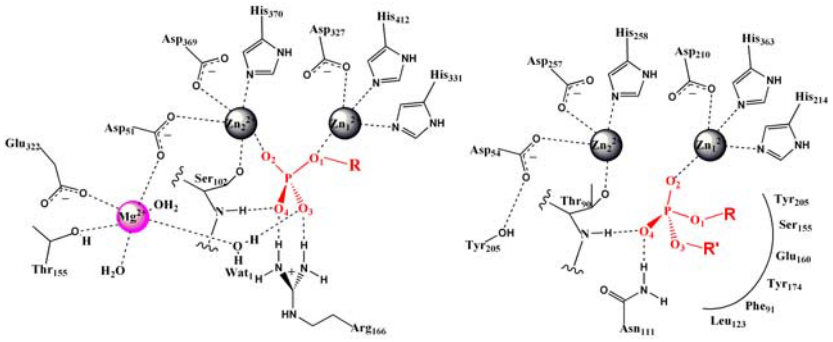
To further facilitate the analysis of sources of errors in both QM and QM/MM calculations, additional analysis for gas-phase/solution proton affinities (PA) for the leaving groups in $MpNPP^-$ and its analogs. QM-only calculations are carried out by Gaussian03 [184]; PCM [71] and SM6 [89] models are employed to describe solvation effects. To test the accuracy of QM/MM coupling, solution PAs are also calculated using SCC-DFTB/MM based free energy perturbation calculations [185].

4.2.2 Enzyme Model Setup

For the hydrolytic reaction catalyzed by the AP superfamily members, a two-step mechanism is usually followed, [22] in which an oxygen nucleophile (e.g., Ser or Thr) first attacks the phosphorus/sulfur, then a water (hydroxide) replaces the leaving group in a step that is essentially the reverse of the first; for some family members of the superfamily, however, the mechanism can be more complex [186]. In this work, to understand the catalytic mechanism of AP with phosphate diesters, we investigate the first step of the hydrolysis reaction of $MpNPP^-$ in an *E. coli* AP variant in which Arg166 is mutated to Ser; this is expected to be the rate-limiting step given the experimental leaving group LFER analysis. Experimentally, this mutant was used to avoid inhibition by P_i, [32] and it is believed that the mutation

does not alter the reaction mechanism of AP since LFERs are similar in the mutant and WT. [30,187] Moreover, the chemical step is fully rate-limiting in this mutant. We also study a double mutant, R166S/E322Y AP, which was constructed in recent experimental studies to analyze the contribution(s) from the Mg^{2+} site in AP. For NPP, we study the wild type enzyme from *Xanthomonas axonopodis pv. citri (Xac)*.

The enzyme models are constructed based on the X-ray structures for the *E. coli* AP mutant R166S with bound inorganic phosphate at 2.05 Å resolution (PDB code 3CMR [182]) and *Xac* NPP with bound Adenosine Mono-Phosphate (AMP) at 2.00 Å resolution (PDB code 2GSU [27]). The enzyme model for R166S/E322Y AP is constructed based on the crystal structure of E322Y AP (PDB code 3DYC [28]) by mutating Arg166 to serine. In each case, starting from the PDB structure, the ligand is first "mutated" to the substrate of interest, $MpNPP^-$ or $MpNPPS^-$; two possible orientations of the substrates are considered for the AP active site, with the -OMe group oriented towards either the magnesium ion or Ser102 backbone amide (see additional discussions in Sect.4.3). Hydrogen atoms are added by the HBUILD module [188] in CHARMM. [189] All basic and acidic amino acids are kept in their physiological protonation states except for Ser102 and Thr90 in AP and NPP, respectively, which are assumed to be the neucleophiles and deprotonated in the reactive complex. Water molecules are added following the standard protocol of superimposing the system with a water droplet of 27 Å radius centered at $Zn1^{2+}$ (see Fig.4.2 for atomic labels) and removing water molecules within 2.8 Å from any atoms resolved in the crystal structure. [161] Protein atoms in the MM region are described by the all-atom CHARMM force field for proteins [190] and water molecules are described with the TIP3P model. [162] The QM region includes groups most relevant to the reaction: the two zinc ions and their 6 ligands (Asp51, Asp369, His370, Asp327, His412, His331), Ser102 and $MpNPP^-$ for R166S AP; for NPP, this includes two zinc ions and their 6 ligands (Asp54, Asp257, His258, Asp210, His363, His214), Thr90 and $MpNPP^-$. Only side chains of protein residues are included in the QM region and link atoms are added between C_α and C_β atoms. A larger QM region also has been tested for R166S AP which further incorporates the entire magnesium site,



(a) (b)

Figure 4.2: The active sites of Alkaline Phosphatase (AP) and Nucleotide PyrophosPhatase/ phosphodiesterase (NPP) are generally similar, with a few distinct differences. (a) *E. coli* AP active site. (b) *Xac* NPP active site. The cognate substrates for AP and NPP are phosphate monoesters and diesters, respectively. The labeling scheme of substrate atoms is used throughout the paper. We propose that diesters and monoesters have different binding modes in the active site (see Sect.4.3.2 for discussions).

including Mg^{2+}, sidechains of Thr155, Glu322 and three ligand water molecules. Comparison of optimized structures using different QM regions indicates fairly similar optimized structures (see **Supporting Information** for details), thus the smaller QM region is used for the majority of the calculations. The treatment of the QM/MM frontier follows the DIV scheme in CHARMM; previous benchmark calculations have shown that this scheme generally gives reliable results for structure and energetics in QM/MM calculations provided that the MM charge is small near the QM/MM boundary. [191] Since the Mg^{2+} near the QM region in AP is treated as a point charge (otherwise the QM region will become substantially larger), to avoid over-polarization of nearby QM groups, a NOE potential is added to the C-O bonds in Asp51, which is coordinated to both Mg^{2+} and Zn^{2+}. The NOE potential takes the form:

$$\begin{aligned} E_R &= 0.0 && R_{min} < R < R_{max} \\ &= 0.5 \cdot K_{max} \cdot (R - R_{max})^2 && R_{max} < R \end{aligned} \tag{4.1}$$

in which R_{min} and R_{max} set the interval between which the restraining potential is zero; they are taken to be 0 and 1.28 Å, respectively. K_{max} is set to be $10^4 kcal/(mol \cdot \mathring{A}^2)$.

Due to the fairly large size of the QM region (more than 80 atoms) and extensive sampling required for the open active site of AP and NPP, the SCC-DFTBPR method [46] is used for PMF calculations. Extensive benchmark calculations and applications indicate that it is comparable to the best semi-empirical method available in the literature for phosphate chemistry. [180, 192]

The generalized solvent boundary potential (GSBP) [124, 163] is used to treat long range electrostatic interactions in geometry optimizations and MD simulations. The system is partitioned into a 27-Å spherical inner region centered at the Zn1 atom, with the rest in the outer region. Newtonian equations-of-motion are solved for the MD region (within 23 Å), and Langevin equations-of-motion are solved for the buffer regions (23-27 Å) with a temperature bath of 300 K; protein atoms in the buffer region are harmonically constrained with force constants determined from the crystallographic B-factors. [193] All bonds involving hydrogen

are constrained using the SHAKE algorithm, [166] and the time step is set to 1 fs. All water molecules in the inner region are subject to a weak GEO type of restraining potential to keep them inside the inner sphere with the MMFP module of CHARMM. The static field due to outer-region atoms, ϕ_s^{io}, is evaluated with the linear Poisson-Boltzmann (PB) equation using a focusing scheme with a coarse cubic grid of 1.2 Å spacing, and a fine grid of 0.4 Å spacing. The reaction field matrix **M** is evaluated using 400 spherical harmonics. In the PB calculations, the protein dielectric constant of $\epsilon_p = 1$, the water dielectric constant of $\epsilon_w = 80$, and 0.0 M salt concentration are used; the value of ϵ_p is not expected to make a large difference in this particular case because the active site is already very solvent accessible and the inner/outer boundary is far from the site of interest. The optimized radii of Nina et al. [194, 195] based on experimental solvation free energies of small molecules as well as the calculated interaction energy with explicit water molecules are adopted to define the solvent-solute dielectric boundary. To be consistent with the GSBP protocol, the extended electrostatic model [164] is used to treat the electrostatic interactions among inner region atoms in which interactions beyond 12 Å are treated with multipolar expansions, including the dipolar and quadrupolar terms.

4.2.3 Benchmark enzyme calculations based on minimizations and reaction path calculations

To further test the applicability of SCC-DFTBPR/MM to AP and NPP, geometry optimization for the reactant (Michaelis) complex is compared to results from B3LYP [196–198]/MM calculations. The basis set used in the B3LYP/MM calculations is 6-31G* [199], and the calculations are carried out with the QChem [200] program interfaced with CHARMM (c36a2 version). [201] Due to the rather large size of the QM region and the high cost of *ab initio* QM/MM calculations, atoms beyond 7 Å away from Zn1 are fixed to their crystal positions in these minimizations (note that these are not fixed in the potential of mean force simulations, see below. Also, test calculations at the SCC-DFTBPR level show that fixing atoms beyond 7 Å from Zn1 in minimizations do not lead to much difference as compared to

a fully flexible inner-region calculation within the GSBP framework). The convergence criteria for geometry optimization are that the root-mean-square (RMS) force on mobile atoms is smaller than $0.30 kcal/(mol \cdot \mathring{A})$ and the maximum force smaller than $0.45 kcal/(mol \cdot \mathring{A})$.

The Minimum Energy Path (MEP) calculations are carried out by one-dimensional adiabatic mapping at both SCC-DFTBPR and B3LYP/6-31G* levels; the reaction coordinate is the antisymmetric stretch involving the breaking and forming P-O bonds (PO^{lg}-PO^{nu}), and the step size for the adiabatic mapping is 0.2 Å. At the SCC-DFTBPR level, the transition state is further refined using CPR.

4.2.4 1D and 2D Potential of mean force (PMF) simulations

To study the free energy profile of enzyme reactions, PMF simulations have been carried out for R166S AP, R166S/E322Y AP and NPP with $MpNPP^-$ and $MpNPPS^-$ as the substrates. After the initial minimizations starting from the relevant crystal structure, the enzyme system is slowly heated to 300 K and equilibrated for 100 ps. The reaction coordinate is defined as PO^{lg}-PO^{nu}. The umbrella sampling approach [167] is used to constrain the system along the reaction coordinate by using a force constant of 150 kcal/mol·$\mathring{A}^{-2}$. In total, more than 51 windows are used for each PMF and 100 ps simulations are performed for each window. The first 50 ps trajectories are discarded and only the last 50 ps are used for data analysis. Convergence of the PMF is monitored by examining the overlap of reaction coordinate distributions sampled in different windows and by evaluating the effect of leaving out segments of trajectories. The probability distributions are combined together by the weighted histogram analysis method (WHAM) [168] to obtain the PMF along the reaction coordinate. The averaged key structural properties for each window are calculated and summarized in Table 4.4.

In a separate set of PMF calculations, the Zn^{2+}-Zn^{2+} distance is constrained to be 3.6, 4.1 and 4.6 Å, respectively, by a strong constraint with a force constant of 2,000 kcal/mol·$\mathring{A}^{-2}$, to investigate the impact of this fundamental variable of the bimetallic site on catalysis in

AP and NPP. For reference, the Zn^{2+}-Zn^{2+} distance found in the various crystal structures for AP and NPP is close to be 4.1 Å.

To verify that the 1D PMFs capture the nature of the phosphoryl transfer transition state, we also carry out 2D PMF calculations for the α orientation of $MpNPP^-$. The reaction coordinates are defined as the P-O^{lg} and P-O^{nu} distances, and the range of each distance is similar to that in the 1D PMF calculations. In total, 272 windows are used, and a force constant of 200 kcal/mol·$\mathring{A}^{-2}$ is used for all windows; for each window, 100 ps simulations are carried out and only the last 50 ps are used for the subsequent WHAM analysis.

4.3 Results and Discussion

4.3.1 $MpNPP^-$ hydrolysis in solution

The hydrolysis of $MpNPP^-$ has been studied extensively by experiments and computations. Experimental studies [32] determined the activation free energy of 25.7 kcal/mol at 42°C with hydroxide as the nucleophile, and the mechanism is established as concerted with a synchronous TS based on LFER analysis. Several computational work also studied the same reaction by employing various levels of theory. By using B3LYP/6-31+G* and the PCM model, Rosta and coworkers obtained a fairly loose TS with P-O^{lg} and P-O^{nu} as 1.86 and 2.49 Å, respectively. [6] By using B3LYP/6-311+G** with the COSMO continuum model on PCM-minimized geometries and a careful treatment of solute configurational entropy, they obtained an activation free energy barrier of 24.4 kcal/mol. In the more recent QM/MM simulations using explicit solvent (TIP3P) and AM1(d)-PhoT [180] as QM, López-Canut et al. obtained a free energy barrier of 20.5 kcal/mol; the transition state was featured with the P-O^{lg} and P-O^{nu} distances of 1.81 and 2.23 Å, respectively, somewhat more compact compared with the PCM result.

With our SCC-DFTB(PR)/PB method and charge dependent atomic radii [52], the adiabatic map for $MpNPP^-$ hydrolysis (Fig. 4.3a) is qualitatively consistent with previous studies and indicates a synchronous TS with an energy barrier of around 30 kcal/mol. After adding higher level (B3LYP or MP2) single point energy corrections, the general landscape

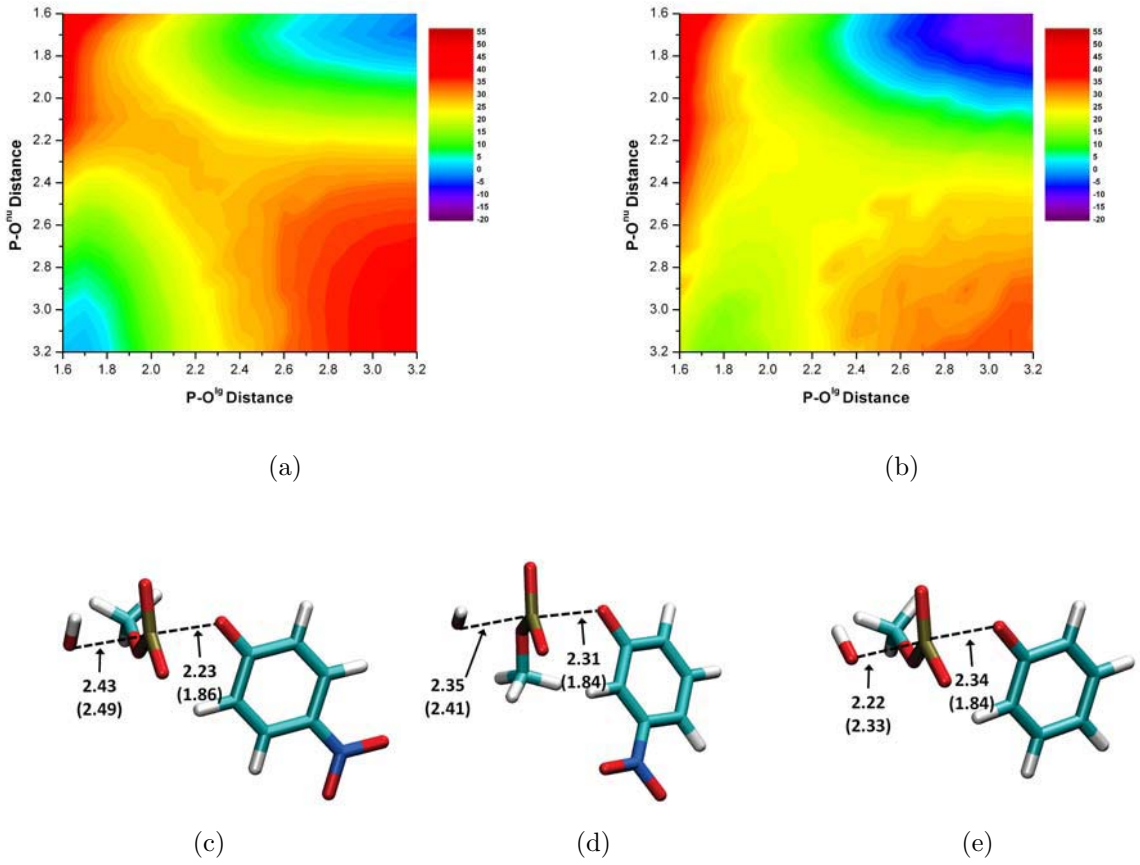


Figure 4.3: Aqueous hydrolysis of phosphate diesters with hydroxide as the nucleophile. Key distances are labeled in Å and energies are in kcal/mol. (a) Adiabatic mapping results for MpNPP$^-$ by SCC-DFTBPR/PB. (b) Adiabatic mapping results for MpNPP$^-$ after including single point gas phase correction at the MP2/6-311++G** level. (c-e) Hydrolysis transition state optimized with Conjugate Peak Refinement (CPR) calculations for MpNPP$^-$, MmNPP$^-$ and MPP$^-$. Numbers without parentheses are obtained by SCC-DFTBPR/PB; those with parentheses are taken from Ref. [6]. As shown in the **Supporting Information**, including the MP2 correction tends to slightly tightens the transition state, especially along P-O^{lg}.

Table 4.1: Energetics for diester hydrolysis reactions in solution from experiments and calculations

Diester	pK_a^a	ΔG_{exp}^b	ΔG_{lit}^c	ΔG_{lit}^d	ΔG_{calc}^e
$MpNPP^-$	7.14	25.7	24.4	20.5	29.3/21.3/24.4
$MmNPP^-$	8.35	26.3	27.3		33.3/28.1/27.2
MPP^-	9.95	28.6	29.9		39.8/30.5/30.6

a. Leaving group pKa. b. Experimental result taken from Ref. [32] c. Calculation result taken from Ref. [6] d. Calculation result taken from Ref. [1] e. For each entry, the numbers are: SCC-DFTBPR/PB, SCC-DFTBPR/PB result including gas-phase B3LYP/6-311++G** single point energy correction, and SCC-DFTBPR/PB result including gas-phase MP2/6-311++G** single point energy correction.

of the adiabatic map does not change (Fig. 4.3b); the synchronous TS is still preferred with the barrier lowered to around 25 kcal/mol when MP2 corrections are used. After CPR refinement, the fully optimized TS (Fig. 4.3c) has a P-O^{lg} of 2.23 Å and P-O^{nu} of 2.43 Å; i.e., the P-O^{lg} distance is longer compared with previous theoretical results, while P-O^{nu} is similar. We note, however, the PES is rather flat near the transition state. The free energy barrier by including ZPE and solute configurational entropy is 29.3 kcal/mol at the SCC-DFTBPR level, which is decreased to 24.4 kcal/mol by including gas phase single point energy correction at the MP2/6-311++G** level (Table 4.1); the latter value compares favorably with experimental value.

In addition, we study the hydrolysis of two other related diesters (Fig. 4.1), methyl 3-nitrophenyl phosphate ($MmNPP^-$) and methyl phenyl phosphate(MPP^-) with the approach. Experimentally, the trend is that the hydrolysis barrier increases as the pK_a of the leaving group increases (Table 4.1) [32]. This trend has been reproduced by a previous theoretical study [6] in which the nature of transition states is found to be synchronous and becomes looser as the pK_a of the leaving group decreases; P-O^{lg} ranged from 1.84 to 1.86 Å and P-O^{nu} ranged from 2.33 to 2.49 Å. With our SCC-DFTBPR/PB method, this trend is

also qualitatively reproduced, regardless of whether the gas-phase correction at higher level (B3LYP or MP2) is included. At a quantitative level, however, the SCC-DFTBPR/PB barriers are too high and the effects of the substitution are overestimated (see Table 4.1); for example, the barrier difference between MpNPP$^-$ and MPP$^-$ is only 2.9 kcal/mol according to experiment, but 10.5 kcal/mol at the SCC-DFTBPR/PB level. The discrepancy remains fairly large even with B3LYP corrections, while including the MP2 gas-phase corrections significantly improves the agreement with experimental value. The nature of the TS also becomes somewhat tighter (especially P-O^{lg}, by $\sim$ 0.2 Å) when MP2 correction is included (see **Supporting Information**).

To better understand the quantitative differences between SCC-DFTBPR, B3LYP and MP2 results, we examine the relative PAs of the leaving groups in the three phosphate diesters in both gas-phase and solution; gas-phase PAs reflect the intrinsic accuracy of the QM method, while solution PA calculations also examine the accuracy of either the implicit solvent model or QM/MM interactions in explicit solvent simulations [202]. As shown in Table 4.2, all DFT methods, which include both SCC-DFTB(PR) and B3LYP, have errors much larger than "chemical accuracy" (1 kcal/mol) for the relative gas-phase PAs, especially concerning the effect of introducing the nitro group; by comparison, MP2 does a much better job. The errors in the relative solution PAs follow the same trend as the relative gas-phase PAs, suggesting that errors in the gas-phase PAs are the major source of error; this is confirmed by the observation that computed relative solvation free energies for the leaving groups considered here are in good agreement with experimental values using both SCC-DFTB(PR) and B3LYP based implicit solvent models, with the exception of B3LYP and UAKS radii (see **Supporting Information**).

As additional benchmark for the nature of the TS, ^{18}O KIE calculations for MpNPP$^-$ in solution are carried out with the SCC-DFTBPR transition state and harmonic vibrational frequencies. As shown in Table S2, the trends are in qualitative agreement with experimental results [205,206] and previous AM1(d)-PhoT and B3LYP results. [207] On the other hand, we note that SCC-DFTBPR overestimates the magnitude of the KIEs, especially for the effects

Table 4.2: Relative proton affinities (in kcal/mol) for leaving groups in the studied diesters[a]

Diester	Exp[b]	SCC-DFTBPR[c]	SCC-2nd[c]	B3LYP[d]	MP2[e]
$MpNPP^-$	0	0	0	0	0
$MmNPP^-$	6.6 (1.7)	8.7/9.7 (4.7)	7.6/8.3 (3.3) [4.9]	9.1/11.1 (1.1)	5.5/7.9
MPP^-	21.4 (3.8)	28.3/30.0 (10.0)	26.7/27.8 (6.8) [10.1]	25.3/28.2 (2.2)	21.2/24.1

a. Since only relative proton affinities (PAs) are of interest, no zero-point energy or thermal corrections has been included. The numbers without parenthesis are gas-phase PAs calculated at gas-phase/solution optimized structures; those with parenthesis are solution PAs. Numbers with bracket are obtained by explicit solvent QM/MM free energy perturbation. b. Experimental values for gas-phase PA are taken from ref [203]. The solution PAs are converted based on experimental pK_a differences at 298K [204]. c. The solution geometries are optimized by SCC-PB [52] at the corresponding level; "SCC-2nd" indicates the standard second-order SCC-DFTB [45]. d. Gas-phase geometries are optimized with B3LYP/6-311++G(d,p); solution geometries are optimized with PCM/UAKS at the same level of theory. e. Gas-phase geometries are optimized with B3LYP/6-311++G(d,p); solution geometries are optimized with PCM/UAKS at the same level of theory. Single point energies are calculated with MP2/6-311++G(d,p).

associated with the leaving group oxygen and the non-bridging oxygen. This is consistent with the trend discussed above that SCC-DFTBPR predicts a solution TS that is looser as compared to previous theoretical calculations, with most notably a weaker and longer P-O^{lg} bond.

In short, the benchmark calculations for the phosphate diesters and their leaving groups suggest that SCC-DFTBPR can provide fairly reliable structural properties of these species and a semi-quantitative description of energetics and the nature of hydrolysis transition state, especially for relative trends associated with different substituents on the leaving group.

4.3.2 First step of $MpNPP^-$ hydrolysis in R166S AP

Based on the crystal structure of the AP R166S mutant complexed with inorganic phosphate, the phosphate ligand is "mutated" to $MpNPP^-$ by adding necessary functional groups to phosphate oxygen. The leaving 4-nitrophenyl group is added to O1 due to the geometrical requirement of the in-line attack from Ser102 (see Fig. 4.2a). The methyl group can be added to O3 or O4, which correspond to two different substrate orientations (denoted as α and β orientations, respectively, following the notation of Ref. [28]). Recent experimental studies [28] using a double mutant AP with the Mg^{2+} site removed and phosphorothioate diesters suggested that the α orientation is preferred over the β orientation. As discussed below (see Sect.4.3.3), however, the interpretation of those elegant experiments may not be as clearcut as presented. Moreover, even if the α orientation is indeed dominant, it is not clear if the discrimination comes from binding or the chemical step. Therefore, it is informative to study both orientations.

The comparison of optimized structures by B3LYP/MM and SCC-DFTBPR/MM shows good agreement between the two levels (Fig.4.4a). The O_{Ser102}-P distance increases from 3.1 Å in the crystal structure to 3.4 (3.3) /3.8 (3.9)Å in B3LYP/MM (SCC-DFTBPR/MM) optimized structure for the α/β orientation, leading to a stable reactant complex. The O2 of the substrate coordinates to one of the zinc ions and O1 with the phenyl group is solvated by water molecules. Interestingly, the slight shift of the substrate position also

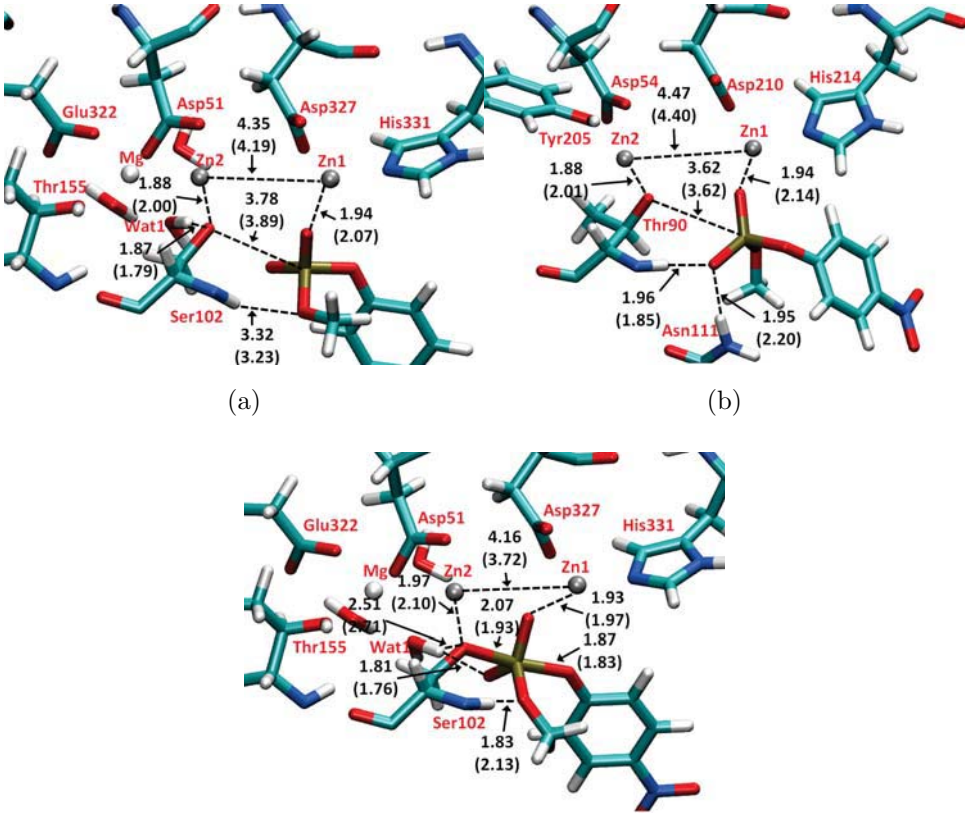


Figure 4.4: Benchmark calculations for $MpNPP^-$ in enzymes. Key distances are labeled in Å. Numbers without parentheses are obtained with B3LYP/6-31G*/MM optimization; those with parentheses are obtained by SCC-DFTBPR/MM optimization. (a) In R166S AP with the substrate methyl group pointing toward Ser102 backbone (the β orientation). (b) In NPP with the substrate methyl group pointing toward the hydrophobic pocket. (c) Comparison of transition state obtained by adiabatic mapping for the β orientation in R166S AP. In (a,c), Asp369, His370 and His412 are omitted for clarity, while in (b), Asp257, His258, His363 are omitted for clarity.

increases the distance between a Mg^{2+}-bound water (Wat1) and substrate O3 such that a hydrogen bond is formed between Wat1 and the Zn^{2+}-activated Ser102, instead of with an inorganic phosphate oxygen as in the crystal structure (this holds also in MD simulations at the SCC-DFTBPR/MM level; also see discussion below for interactions in the TS). O4 and the nearby Ser102 backbone amide forms the only direct hydrogen bond between the substrate and the enzyme, which is shorter (1.9 Å vs. 3.3 Å) with the substrate in the α orientation than in the β orientation. If this is the only major difference between those two orientations, the binding affinity of $MpNPP^-$ to the enzyme is likely stronger with the α orientation than with the β orientation, although a more quantitative estimate remains to be carried out with free energy simulations, which we defer to a future study. Many other hydrogen-bonding distances (e.g., between Wat1 and Ser102, $MpNPP^-$ and Ser102 backbone amide) and distances involving the zinc ions (e.g., between Ser102/$MpNPP^-$ and the zinc ions) are similar at the two levels of theory (see Fig.4.4). The Zn^{2+}-Zn^{2+} distance is generally shorter at the SCC-DFTBPR/MM level (by ∼0.1-0.2 Å) while the distances between Zn and its ligand oxygen are generally longer. Overall, however, the agreement between optimized structures at the two levels of theory is excellent, supporting the use of SCC-DFTBPR/MM.

In addition to the structural similarity in optimized reactants, the MEP results (for β orientation) from adiabatic mapping also show good agreement (16.9 vs. 15.7 kcal/mol) between SCC-DFTBPR/MM and B3LYP/6-31G*/MM calculations, which further supports the use of SCC-DFTBPR/MM; the adiabatic mapping and CPR calculations at the SCC-DFTBPR/MM level give similar transition states (see **Supporting Information**), although the barrier height is slightly lower with fully (CPR) optimized saddle point (see Table 4.3). As shown in Fig.4.4c, the main differences between SCC-DFTBPR/MM and B3LYP/MM transition states include a shorter Zn^{2+}-Zn^{2+} distance at the former level, a shorter P-O_{Ser102} bond length and hydrogen-bond distances for the interaction between the substrate and nearby groups (e.g., Ser102 backbone amide and Mg^{2+}-bound Wat1). We note that, in those MEP calculations with SCC-DFTBPR, the Zn^{2+}-Zn^{2+} distance appears somewhat

shorter in the TS for the β orientation than the α orientation (see **Supporting Information**), although the difference is much smaller in the PMF calculations (see below), again highlighting the importance of sampling protein fluctuations. Nevertheless, there is room to further improve the SCC-DFTBPR method, which may require including complete third-order terms in the SCC-DFTB expansion [208] and a more systematic refitting of the P-O repulsive potential [209].

For both substrate orientations, the PMF peaks at the reaction coordinate (PO^{lg}-PO^{nu}) slightly less than 0 Å and then drops in the product region, corresponding to an exothermic reaction (Fig. 4.5a, 4.7a). The free energy barriers are 23.4 and 19.6 kcal/mol for the α and β orientations, respectively (see Table 4.3). It is worth mentioning that these barriers correspond to the free energy difference between the TS and the Michaelis complex, i.e., k_{cat}, while experimentally reported values are k_{cat}/K_M, which prevents a direct comparison between calculation and experiment. Nevertheless, the measured k_{cat}/K_M for R166S AP with $MpNPP^-$ corresponds to a free energy barrier of 18.0 kcal/mol, which gives the lower bound for the free energy barrier for the chemical step. Therefore, our calculated barriers for the chemical step are qualitatively consistent with the measured k_{cat}/K_M value. The calculations also suggest that the α orientation has a higher barrier for the chemical step. Therefore, for the α orientation to be at least as competitive as the β orientation in terms of k_{cat}/K_M, the corresponding binding free energy should be at least 3.8 kcal/mol stronger, which is qualitatively consistent with the above observation of a stronger hydrogen bonding interaction between the enzyme and $MpNPP^-$ in the α orientation. Although further binding free energy calculations need to be carried out, the results suggest that the model [28] in which the α orientation is the only productive binding mode seems oversimplified (see Sect.4.3.3 below for additional discussions).

The key structural properties of the active site are averaged over the trajectory of each window and plotted as functions of the reaction coordinate (Fig. 4.5b, 4.7b). The changes of P-O^{lg} and P-O^{nu} clearly show that the concerted pathway with a synchronous TS is operative for both substrate orientations and similar to the reaction in aqueous solution, supporting

Table 4.3: Barriers and experimental rates for the first step of $MpNPP^-$ hydrolysis in AP variants and wild type NPP

System[a]	Substrate	k_{cat}/K_M	Exp[b]	α/R_p [c]		β/S_p [d]	
R166S AP	$MpNPP^-$	0.48	18.0	23.4^f	(13.1^g)	19.6^f	(12.1^g/15.7^h)
	$MpNPPS^-$	1.1×10^{-3}	21.6	25.5^f	(15.2^g)	33.4^f	(20.7^g)
R166S/E322Y AP	$MpNPP^-$	0.24	18.4	22.6^f		$>30^f$	
	$MpNPPS^-$	3.0×10^{-3}	21.0	19.7^f		$>40^f$	
NPP	$MpNPP^-$	2.3×10^2	14.3			$20.2^{e,f}$	
R166S AP cons-3.6	$MpNPP^-$			19.0			
R166S AP cons-4.1	$MpNPP^-$			22.7			
R166S AP cons-4.6	$MpNPP^-$			>29			
NPP cons-3.6	$MpNPP^-$			17.0			
NPP cons-4.1	$MpNPP^-$			19.2			
NPP cons-4.6	$MpNPP^-$			24.9			

a. In the "cons" simulations, the Zn^{2+}-Zn^{2+} distance is constrained to be a specific value; b. Free energy barrier (kcal/mol) calculated by transition state theory at 300 K based on experimental k_{cat}/K_M value; c. For $MpNPP^-$, the substrate methyl group points toward the Mg^{2+} site; for $MpNPPS^-$, it's the R_p enantiomer. d. For $MpNPP^-$, the substrate methyl group points toward the Ser102 backbone; for $MpNPPS^-$, it's the S_p enantiomer. e. the substrate methyl group points toward the hydrophobic pocket of NPP; f. PMF barrier with SCC-DFTBPR/MM; g. Barrier from CPR calculations with SCC-DFTBPR/MM; h. adiabatic mapping barrier with B3LYP/6-31G*/MM.

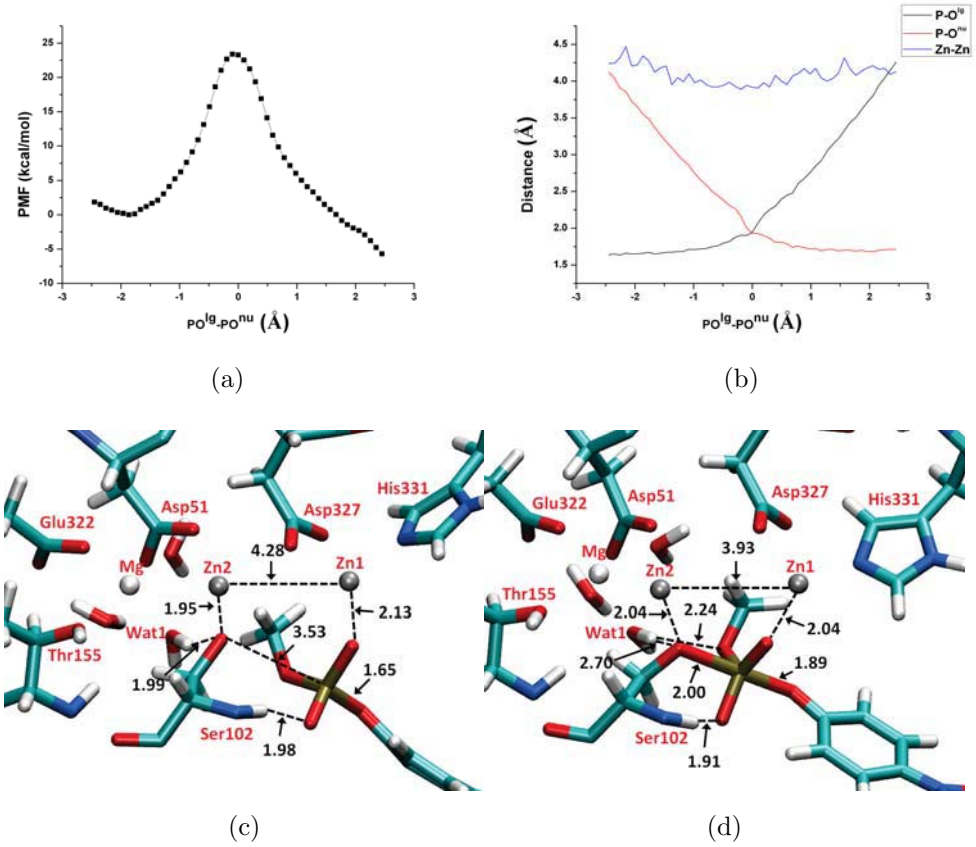
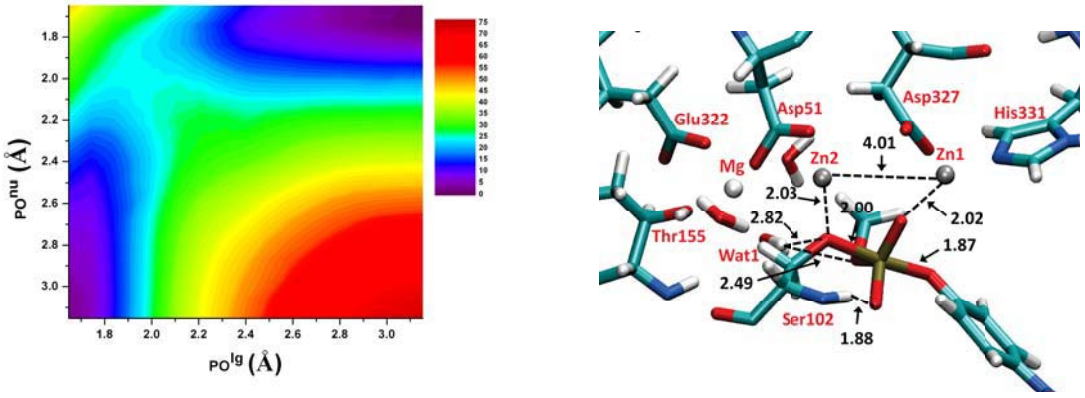


Figure 4.5: Potential of Mean Force (PMF) calculation results for MpNPP$^-$ hydrolysis in R166S AP with the substrate methyl group pointing toward the Mg^{2+} site (the α orientation). Key distances are labeled in Å and energies are in kcal/mol. (a) PMF along the reaction coordinate (the difference between P-O^{lg} and P-O^{nu}); (b) changes of average key distances along the reaction coordinate; (c) A snapshot for the reactant state, with average key distances labeled. (d) A snapshot for the TS, with average key distances labeled. In (c-d), Asp369, His370 and His412 are omitted for clarity.



(a) (b)

Figure 4.6: 2D Potential of Mean Force (PMF) calculation results for MpNPP$^-$ hydrolysis in R166S AP with the substrate methyl group pointing toward the Mg^{2+} site (the α orientation). Key distances are labeled in Å and energies are in kcal/mol. (a) The 2D PMF along the reaction coordinates; (b) A snapshot for the TS, with average key distances labeled. Asp369, His370 and His412 are omitted for clarity. Note that the 2D PMF results are consistent with the 1D PMF results shown in Fig.4.5.

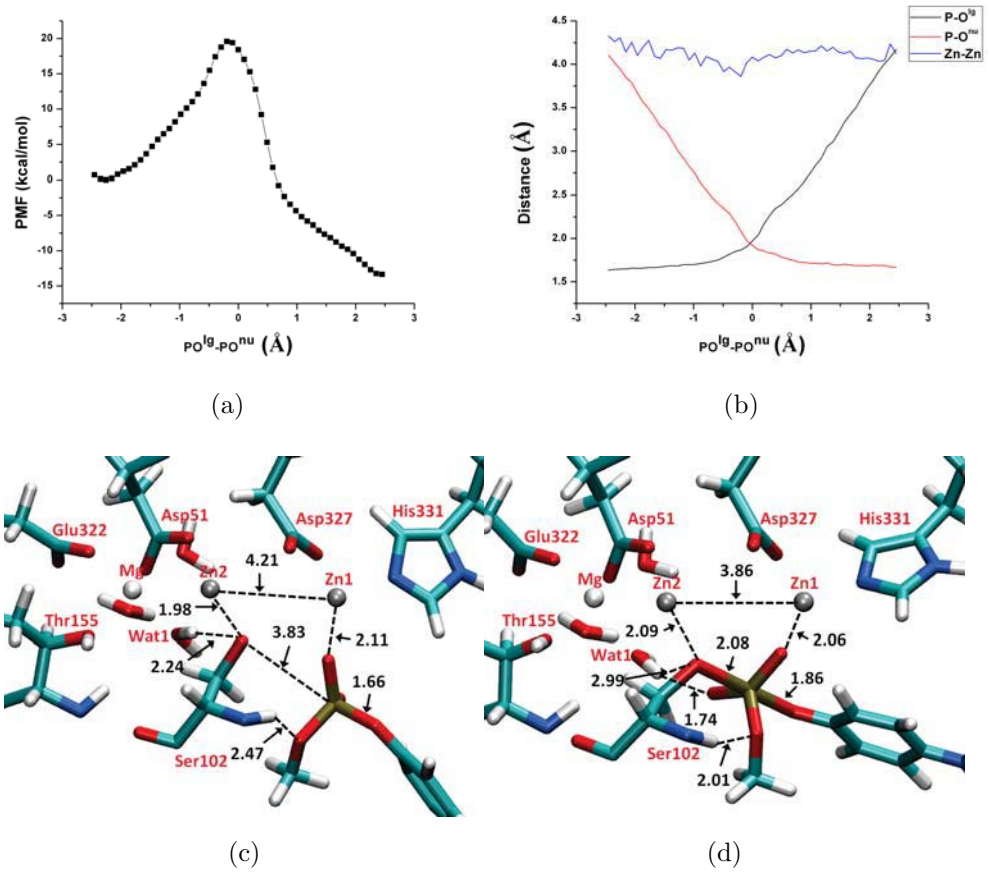


Figure 4.7: Potential of Mean Force (PMF) calculation results for MpNPP$^-$ hydrolysis in R166S AP with the substrate methyl group pointing toward Ser102 backbone (the β orientation). All other format details follow Fig.4.5.

that AP does not significantly alter the nature of hydrolysis TS for phosphate diesters; this is in qualitative agreement with LFER data for R166S AP and a series of substituted methyl phenyl phosphate diesters [32]. The nature of the calculated TS is not sensitive to the definition of the 1D reaction coordinate, since the 1D (Fig.4.5) and 2D (Fig.4.6) PMF calculations show very consistent results in terms of both structures and energetics. In terms of the tightness coordinate (TC=PO^{lg}+PO^{nu}), it decreases from 4.66 Å in solution to 3.89 (α) and 3.94 Å (β) in R166S AP (Table 4.4); this decrease is likely due to the bi-metallic zinc motif in AP, since both $MpNPP^-$ and Ser102 are coordinated with Zn^{2+} ions, which are separated by $\sim$ 4 Å throughout the reaction. We note that the degree of tightening is likely not as large as the values for TC imply since our calculations appear to overestimate the value of TC for solution reactions as compared to previous calculations [1, 6].

We note that recent QM/MM simulation of WT AP with both mono- and di-esters [2,179] showed rather large structural changes compared to the crystal structure; for example, the Zn^{2+}-Zn^{2+} distance increases up to 7 Å, while no such major distortion is observed here (also see discussions below in Sect.4.3.5). In the TS for the β orientation (Fig.4.7d), Wat1 of the Mg^{2+} breaks the hydrogen bond with Ser102 and forms a new one with O4 of $MpNPP^-$, which is presumably due to the larger reduction of charge on the Ser102-O than the substrate O4; the average Mulliken charges for Ser102 O and O4 are -0.79 and -0.92, respectively, in the reactant, -0.56 and -0.89, respectively, in the TS. This change of hydrogen-bonding interactions likely helps lower the reaction barrier compared with the α orientation in which Wat1 interacts loosely with both Ser102-O and O3 in the TS (see Fig.4.5d); the average Mulliken charges for Ser102 O and O3 are -0.79 and -0.32, respectively, in the reactant, -0.50 and -0.40, respectively, in the TS.

Another interesting observation is that the leaving group does not form a direct interaction with any zinc ion in either the reactant state or TS; rather, it is "solvated" by water molecules accessible to the fairly open active site. This binding mode (especially in the TS) is in contrast to that observed for vanadate, a widely used transition state analog for

Table 4.4: Calculated key structural properties for the first step of $MpNPP^-$ hydrolysis in AP variants and wild type NPP

	solution		R166S AP				NPP	
			α^a		β^b		HP^c	
	reactant	TS	reactant	TS	reactant	TS	reactant	TS
P-O^{lg}	1.67	2.23	1.65±0.03	1.89±0.07	1.66±0.03	1.86±0.06	1.63±0.03	1.83±0.06
P-O^{nu}	∞	2.43	3.53±0.07	2.00±0.09	3.83±0.07	2.08±0.08	3.51±0.06	2.03±0.09
RC^d	$-\infty$	-0.20	-1.88±0.06	-0.11±0.07	-2.17±0.06	-0.22±0.08	-1.88±0.06	-0.20±0.07
TC^e	∞	4.66	5.18±0.09	3.89±0.14	5.49±0.08	3.94±0.13	5.14±0.08	3.86±0.14
		(4.05)	(5.85)	(5.00)			(6.29)	(5.66)
Zn^{2+}-Zn^{2+}			4.28±0.22	3.93±0.18	4.21±0.22	3.86±0.16	4.38±0.21	3.92±0.17
Cons-3.6^f								
TC^e			4.65±0.09	3.88±0.13			4.65±0.09	3.85±0.11
Cons-4.1^f								
TC^e			5.31±0.09	3.97±0.18			5.08±0.08	3.97±0.16
Cons-4.6^f								
TC^e			5.83±0.10	4.08±0.25			5.34±0.08	4.09±0.19

a. The substrate methyl group points toward the Mg^{2+} site (the α orientation); b. the substrate methyl group points toward the Ser102 backbone (the β orientation); c. the substrate methyl group points toward the hydrophobic pocket; d. The Reaction coordinate (RC) is defined as the difference between P-O^{lg} and P-O^{nu}; e. The Tightness coordinate (TC) is defined as the sum of P-O^{lg} and P-O^{nu}; in parentheses are values from previous QM/MM simulations [1]. f. Zn^{2+}-Zn^{2+} distance constrained at 3.6, 4.1 and 4.6 Å respectively. The RMS fluctuations are 0.01 Å.

phosphoryl transfers, in the crystal structures for AP-vanadate and NPP-vanadate complexes [27, 210]; these structures suggest a binding mode in which one non-bridging oxygen and the leaving oxygen interact directly with one of the zinc ions. Benchmark calculations suggest that our QM/MM protocol is able to reproduce the binding mode of vanadate in the active site of both AP and NPP, and that SCC-DFTBPR describes the interaction between the di-metallic zinc motif and phosphate diesters in good agreement with B3LYP (see **Supporting Information**). Moreover, MD simulations starting from the vanadate binding mode with the reaction coordinate constrained to be zero converge to the same binding mode shown in Figs.4.5-4.7. Collectively, these results suggest that the binding mode observed in the current work is unlikely an artifact of the computational methodology and indeed energetically favorable for systems studied here. We note that the TS for diesters in AP (and NPP, see below) is rather tight in nature, thus the leaving group oxygen doesn't bear any significant formal charge. Therefore, the binding mode captured in the vanadate structures better reflects the situation for monoesters, which feature a much looser TS in which the leaving group is substantially more charged. Our preliminary calculations for monoesters in AP indeed find tighter interactions between the zinc ion and the leaving group (Hou and Cui, work in progress).

4.3.3 Additional analysis of substrate orientation: activity in the double mutant (R166S/E322Y) and thio effects in R166S AP

To further clarify the issue of substrate orientation in AP, we carry out simulation studies to analyze two sets of experiments that were designed to answer the same question. In the first set of experiments, Zalatan and coworkers constructed mutants with the Mg^{2+} site removed (E322Y, E322A, R166S/E322Y) and measured the catalytic activities for phosphate monoester, phosphate diester and sulphate monoester in these mutants [28]. Based on the observed large detrimental effects of Mg^{2+} removal on phosphate monoester and sulphate monoester hydrolysis but negligible effect on phosphate diester hydrolysis, they concluded

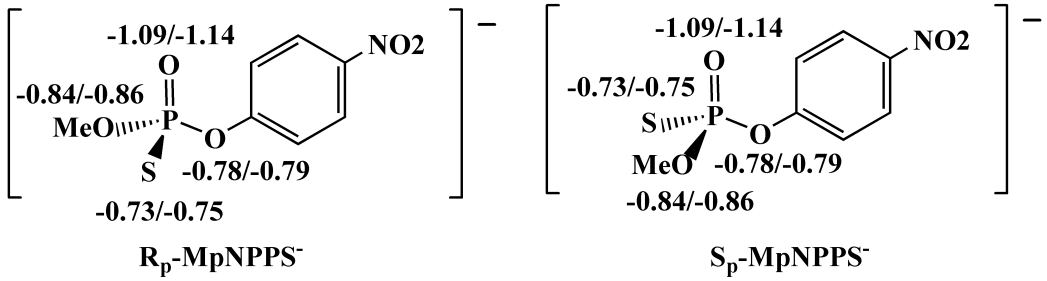
that the α orientation is preferred over the β orientation (also see the discussion on thio effects below). However, these observations alone do not rule out the possibility that the two orientations are in fact similar in activity in the WT (and R166S) enzyme, which is what we have observed in this study (see Table 4.3). In a mutant where one pathway is significantly perturbed, the other can still provide an alternative route, which may explain the only 2 fold decrease in k_{cat}/K_M for R166S/E322Y AP as compared to R166S AP (both with $MpNPP^-$ as the substrate). To support this, we explicitly carry out calculations for the double mutants.

As mentioned in **Computational Methods**, the R166S/E322Y double mutant simulations are prepared based on the crystal structure of E322Y AP [28], in which the hydroxyl group of Tyr322 occupies the region corresponding to the Mg^{2+} site in WT AP and forms a hydrogen bond with Asp51; the bimetallic zinc site is largely unaffected. Based on the comparison of results for the α and β orientations in R166S AP (Figs.4.5-4.7), we expect that mutating away the Mg^{2+} site, which turns off interactions between the Mg^{2+}-bound water and substrate oxygen, will result in a large detrimental effect on the β orientation but a much smaller effect on the α orientation. This is exactly what we observe from the double mutant calculations; as shown in Table 4.3, the reaction barrier is slightly decreased to 22.6 kcal/mol for the α orientation but is increased to be over 30 kcal/mol for the β orientation. These results directly support the important role of the Mg^{2+} site in reducing the barrier for the β orientation, and by inference, for the hydrolysis of phosphate monoesters, as observed experimentally [28]. Analysis of structures from PMF calculations also indicates that a water molecule penetrates into the double mutant active site for the β orientation to further stabilize the nucleophile in the reactant state, thus also contributing to the significant increase of the barrier; no such water penetration is observed for the α orientation since the hydrophobic -OMe group in the substrate helps block additional water from the active site. The double mutant calculations also explicitly support that the α orientation is not affected much by the Mg^{2+} site, in agreement with the experimental observation [28] that the activity of phosphate diesters remains largely unperturbed in the Mg^{2+}-site mutants. In short, our

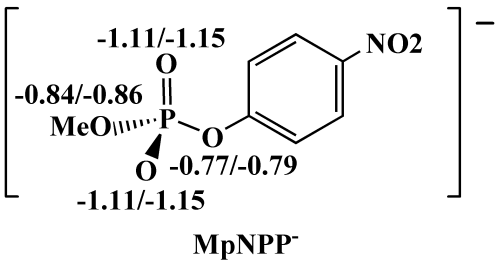
model that both α and β orientations are productive binding modes in R166S AP while only α is reactive in the Mg^{2+}-site mutants is consistent with available experimental data.

In the second set of experimental studies, Zalatan and coworkers analyzed the reactivities of phosphorothioate diesters in several variants of AP [28]. Take $MpNPPS^-$ as an example (Fig.4.8), the key observation was that R166S AP reacts with the R_p enantiomer at least 10^2 times faster than with the S_p enantiomer, suggesting that the binding mode with the R' (Me) group toward the Mg^{2+} site and the sulfur toward the Ser102 backbone amide is dominant; a relevant piece of information here is that previous experiments in the same group established that the sulfur is not placed between the Zn^{2+} ions [176]. We note, however, these discussions rely on the assumption that phosphorothioate diesters behave, in terms of hydrogen bonding with active site groups, similarly to phosphate diesters, which may not be as clearcut as commonly believed. For example, NBO charge analysis in both gas phase and solution (see Fig. 4.8) shows that in $MpNPPS^-$ the oxygen bonded with the methyl group is in fact more negatively charged than the sulfur; i.e., it is not obvious that the R_p enantiomer of $MpNPPS^-$ reflects the α-orientation of $MpNPP^-$. Therefore, although the experimental observation that removal of the Mg^{2+} site (R166S/E322Y AP vs. R166S) has a similar impact on the hydrolysis of $MpNPPS^-$ and $MpNPP^-$ suggests that interactions from the Mg^{2+} site are similar in the phosphorothioate and phosphate esters, it is not clear if the thio effects can unambiguously infer the binding mode of phosphate diester.

To help better understand the thio effects, we calculate the PMF for $MpNPPS^-$ hydrolysis in both R166S AP and R166S/E322Y AP. The results support that in R166S AP, the R_p enantiomer is indeed favored over S_p enantiomer by 7.9 kcal/mol for the chemical step (see Table 4.3), in qualitative agreement with experimental findings. The experimental thio effect (the ratio of rate constant for phosphate ester substrate over that for the phosphorothioate analog) corresponds to a free energy barrier difference of 3.6 kcal/mol, while our calculated value is 2.1-5.9 kcal/mol, depending on whether the α or β orientation for $MpNPP^-$ is used as reference; note again, however, the experimental value is based on k_{cat}/K_M while our values are based on the chemical step only. One point worth mentioning is that due to the weaker



(a)



(b)

Figure 4.8: NBO charge analysis for MpNPPS$^-$ and MpNPP$^-$ in gas phase and solution. Geometries are optimized in gas phase by B3LYP/6-311++G(d,p). Solvation effects are added by PCM with UAKS radii. Numbers before/after slash are gas-phase/solution NBO charges. (a) Enantiomers of MpNPPS$^-$; (b) MpNPP$^-$.

substrate binding following thio substitution, the Solvent Accessible Surface Area (SASA) of the sulfur in MpNPPS$^-$ is much larger than that for the corresponding oxygen in MpNPP$^-$, especially for the S_p enantiomer (see Table 2 in **Supporting Information**); this does not occur at the transition state. The larger SASA provides extra solvent stabilization for the reactant state and probably accounts partially for the much larger thio effects calculated for the S_p enantiomer (13.8 kcal/mol higher in barrier relative to the β orientation of MpNPP$^-$). The PMF profiles for MpNPPS$^-$ also peak at where the value of the reaction coordinate is slightly less than 0 while the position of the reactant state is decreased from ∼2 Å in MpNPP$^-$ to ∼-2.5 to -3 Å, reflecting the larger substrate size and weaker binding interactions with the active site than MpNPP$^-$ (see **Supporting Information**); the transition state is still synchronous in nature and slightly tighter than MpNPP$^-$.

The effect of removing the Mg^{2+} site (in R166S/E322Y AP) on the hydrolysis of MpNPPS$^-$ is expected to be small for both enantiomers since the interactions between Mg^{2+}-water and either the -OMe or -S$^-$ are fairly weak. Indeed, for the R_p enantiomer, the barrier actually decreases in the R166S/E322Y mutant relative to the R166S AP to 19.7 kcal/mol; this value is slightly lower than the experimental k_{cat}/K_M value of 21.0 kcal/mol. For the S_p enantiomer, surprisingly, the barrier increases to be over 40 kcal/mol. A closer examination of the simulation snapshots shows that for the reactant state with the S_p enantiomer a solvent water penetrates into the active site and forms a hydrogen bond with the nucleophile (Ser102 oxygen), as illustrated by the comparison of integrated water distribution near the nucleophilic oxygen in Ser102 in the reactant state (see **Supporting Information**). As mentioned above for the β orientation of MpNPP$^-$ in R166S/E322Y AP, which also features a significantly increased barrier (Table 4.3), a similar water penetration is observed as well. The water penetration to Ser102 only happens for these two cases and is reproducible in simulations in which the penetrated water is first deleted and then the system further equilibrated. We note that in R166S AP the active site is already rather open to solvent molecules, thus additional water penetration into the active site is not unexpected when the Mg^{2+} site is removed; the β orientation and the S_p enantiomer are particularly susceptible to

water penetration since they lack the bulky methyl group near Ser102. Nevertheless, water penetration in AP mutants remains an interesting issue that deserves in-depth analysis from future experimental and computational studies.

Considering the results for $MpNPP^-$ and those for the R_p enantiomer for $MpNPPS^-$ in the two AP variants, our calculations qualitatively reproduced key experimental observations concerning the effects of Mg^{2+} site removal and thio substitution, further supporting the argument in the last subsection that experimental data so far can not be used to unambiguously determine the orientation of diester substrates in the AP active site. In the broader context, as we mentioned above, the charge distributions for $MpNPP^-$ and $MpNPPS^-$ bear some nontrivial differences (Fig.4.8); in addition, the possibilities of water penetrating into the active site for certain orientation of the (thio-substituted) substrate and that different substrate orientations are dominant in different variants of AP (R116S vs. R116/E332Y) further complicate interpretation of the observed thio effects.

4.3.4 First step of $MpNPP^-$ hydrolysis reaction in NPP

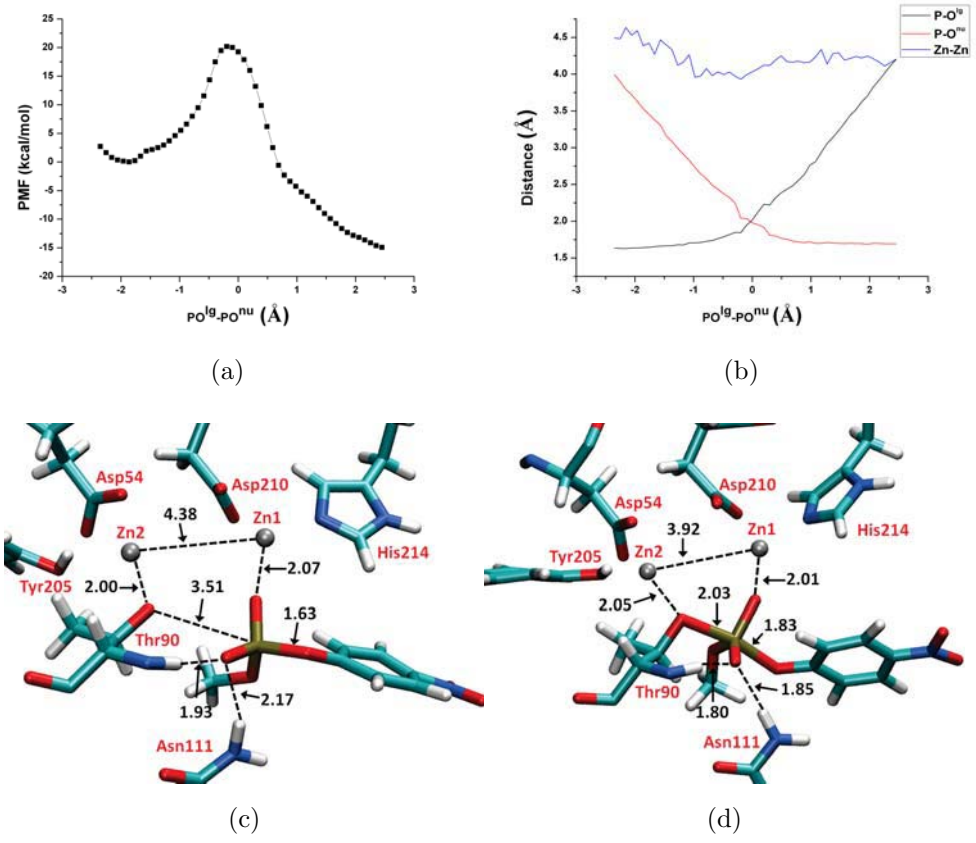
The hydrophobic groove in NPP has been suggested to contribute at least 10^4-fold to the catalysis of phosphate diester reactions [27] by favorable interactions with the extra R' group in diesters (Fig.4.2b). Therefore, only one orientation is studied here for $MpNPP^-$ in which the methyl group points toward the hydrophobic pocket. Similar to the comparisons made above for AP, SCC-DFTBPR/MM minimizations for $MpNPP^-$-NPP also give similar reactant complex structure to B3LYP/MM calculations (Fig.4.4b). The O_{Thr90}-P distance increases from 3.2 Å in the crystal, which contains AMP as the inhibitor, to 3.6 (3.6) Å at the SCC-DFTBPR/MM (B3LYP/MM) level. The substrate O2 coordinates with Zn1 while O4 forms hydrogen bonds with Asn111 and the backbone amide of Thr90. The optimized Zn^{2+}-Zn^{2+} distance is 4.47 (4.40) at the B3LYP/MM (SCC-DFTBPR/MM) level. The two hydrogen bonds formed between O4-Asn111 and O4-Thr90-backbone-amide are also in decent agreement at different levels of theory (Fig.4.4b).

The PMF calculation shows that the free energy profile corresponds to an exothermic process with the barrier located at PO^{lg}-PO^{nu}∼-0.20 Å and a barrier height of 20.2 kcal/mol. The measured k_{cat}/K_M corresponds to 14.3 kcal/mol at 300 K [27] and sets the lower limit for the chemical step barrier. Compared with AP, the calculated barrier for NPP is close to the β orientation but lower than that in the α orientation. Since $MpNPP^-$ is the cognate substrate of NPP and therefore expected to bind tighter to NPP than to AP, the calculated barrier implies a higher k_{cat}/K_M value for NPP than for AP, which is consistent with experimental observations. In other words, the calculations explicitly support that although diesters are cognate substrates for NPP and promiscuous substrate for AP, the chemical step in NPP is not much accelerated over (R166S) AP.

The changes of P-O^{lg} and P-O^{nu} (Fig. 4.9b) show that the concerted pathway is also operative for NPP, with a TS similar to that in aqueous solution and AP. For example, the tightness coordinate is 3.86 Å (Table 4.4), as compared to the values of 3.89(α)/3.94(β) and 4.66 Å, respectively, in R166S AP and solution, respectively. Considering that the tightness coordinate for solution TS seems overestimated by our method compared to previous work [1,6], our calculations support the idea motivated by LFER data [32] that, instead of altering transition state structure, NPP and AP catalyze phosphoryl transfer reactions by recognizing and stabilizing transition states similar to those in aqueous solution.

4.3.5 Comparison to recent QM/MM simulations [1, 2]

As just stated, our calculations find that the transition states for diester hydrolysis in AP and NPP are similar and slightly tighter than that in solution. This is in direct contrast to the recent QM/MM studies [1, 2] which found that the TS in AP/NPP is much looser in nature with the tightness coordinate of 5.66 (NPP)/5.00 (AP) vs. a value of 4.05 in solution. Several pieces of evidence suggest that those calculations are less reliable than our SCC-DFTBPR/MM calculations. First, as noted above for AP, their calculations led to large structural distortions in the bi-metallic zinc motif relative to the crystal structure, while no such distortions occur in our calculations; the same trends hold for NPP calculations, and the



(a) (b)

(c) (d)

Figure 4.9: Potential of Mean Force (PMF) calculation results for MpNPP$^-$ hydrolysis in NPP with the substrate methyl group pointing toward the hydrophobic core. Other format details follow Fig.4.5. In (c-d), Asp257, His258, His363 are omitted for clarity.

loose TS found in previous work [1, 2] might be a result of the substantially elongated Zn^{2+}-Zn^{2+} distance. Another example concerns the hydrogen-bond between Asn111 and $MpNPP^{-}$ in NPP, which is observed in the crystal structure (with AMP and vanadate) and throughout our simulations; by contrast, this interaction was broken in the recent QM/MM simulations. [1] Second, our calculated barrier heights are consistently higher than the barriers that correspond to experimentally measured k_{cat}/K_M values, while this is not the case for recent QM/MM calculations for both NPP [1] and AP [2]. For example, for NPP with the same substrate, their best estimate of the barrier [1] was ∼11 kcal/mol, which is even lower than the barrier estimated based on experimental k_{cat}/K_M, casting further doubt on the quantitative nature of their result.

What could be the origin for the differences between our and the recent QM/MM calculations [1, 2]? Since the AM1(d)-PhoT approach seems to give fairly reasonable results for the solution reaction, both in terms of energetics and KIEs [207], we suspect that the main cause is the use of AM1 for zinc in Refs. [1, 2]. Although AM1/PM3 has been used successfully to describe a number of metalloenzymes that catalyze phosphoryl transfers [211, 212], combining AM1 for zinc and AM1(d)-PhoT for phosphoryl transfers in zinc enzymes has not been carefully tested. In this regard, although it is possible that the crystal structure with an inhibitor (e.g., inorganic phosphate) doesn't capture all structural features (e.g., variation in the zinc-zinc distance) of the transition state, the fact that the active site undergoes little change with a transition state analogue (vanadate, see discussions in Sect.4.3.2 and also **Supporting Information**) suggests that large variations in the zinc-zinc distance seen in the recent QM/MM simulations [1, 2] (including in the Michaelis complexes!) are unlikely realistic. Since the large increase in the zinc-zinc distance occurs in their calculations of several AP and NPP variants [1, 2], the effect likely cancels out for relative trends; this explains why mutation effects were adequately captured in Ref. [2]. Finally, we noted that the QM/MM boundary in Refs. [1, 2] cuts across fairly polar covalent bonds yet the simple link-atom scheme was used. As emphasized by several researchers [59, 64, 191, 213],

extra care needs to be exercised when the QM/MM boundary involves polar bonds. This technical detail likely also contributes to the uncertainty of the results in Refs. [1, 2].

4.3.6 Why is the nature of TS for phosphate diesters in AP and NPP similar to that in solution?

As another way to evaluate the conflicting findings in the current and previous QM/MM studies, we further dissect whether our observation that the nature of TS for phosphate diesters in AP and NPP is similar to that in solution is consistent with other known experimental facts. We do so with the scheme outlined in Fig.4.10, which is qualitatively similar to that used by Herschlag and co-workers [32, 33, 178].

For aryl phosphate diester hydrolysis in solution, the measured reaction barrier, $\Delta G^{\ddagger(aq)}$ is ∼26 kcal/mol (e.g., see Table 4.1 for $MpNPP^-$); the corresponding barrier in the enzyme, $\Delta G^{\ddagger(E)}$, is, for R166S AP, 18.0 kcal/mol (see k_{cat}/K_M in Table 4.3). Therefore, the enzyme binds to the TS, which is shown here to be of rather similar synchronous nature in solution and AP/NPP (Table 4.4), by about 8 kcal/mol ($\Delta\Delta G^{b}_{syn^{\ddagger}}$).

For the enzyme to shift the nature of TS from synchronous to loose, the driving force needs to be large enough to overcome the binding energy for the synchronous TS ($\Delta\Delta G^{b}_{syn^{\ddagger}}$) plus the energy gap between these two kinds of structures in solution ($\Delta\Delta G^{\ddagger(aq)}_{syn/loose}$). The latter, although not measurable directly with experiments, can be estimated based on calculations; our calculations shown in Fig.4.3 give a value ∼ 8 kcal/mol (the "loose" structure is taken to have a tightness coordinate of ∼5.7Å as found for the TS in NPP in previous QM/MM calculations [1]). In other words, the enzyme needs to bind to the loose TS by more than 16 kcal/mol ($\Delta\Delta G^{b}_{syn^{\ddagger}}$+$\Delta\Delta G^{\ddagger(aq)}_{syn/loose}$=8 + 8) to make the loose TS more favorable in the enzyme than a synchronous one.

Considering what we know about the activity of AP toward phosphate monoesters, however, we argue that such a strong binding is unlikely for phosphate diesters. For a phosphate monoester related to the diesters studied here, such as $pNPP^{2-}$ (*p*-nitrophenyl phosphate), the solution barrier is about 32 kcal/mol [178] and the barrier in R166S AP is 10.6

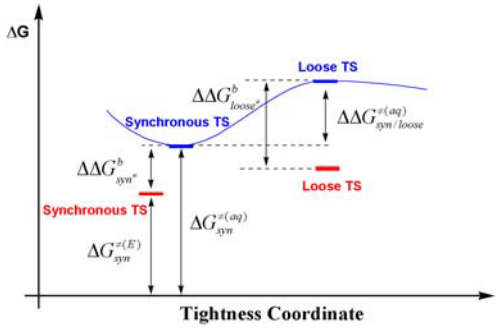


Figure 4.10: A scheme that illustrates how relative energetics of synchronous and loose transition states in the enzyme (in red) compare to those in solution (in blue). $\Delta G^{\ddagger(aq/E)}_{syn}$ gives the free energy barrier (relative to infinitely separated substrate and nucleophile) in solution/enzyme; $\Delta\Delta G^{b}_{syn/loose^{\ddagger}}$ gives the binding free energy of a syn/loose TS structure to the enzyme; $\Delta\Delta G^{\ddagger(aq)}_{syn/loose}$ is the free energy difference between the synchronous and loose transition state structures in solution. For the enzyme to shift the nature of TS from synchronous to loose, $\Delta\Delta G^{b}_{loose^{\ddagger}}$ needs to be larger than $\Delta\Delta G^{b}_{syn^{\ddagger}} + \Delta\Delta G^{\ddagger(aq)}_{syn/loose}$, which we argue is unlikely for AP and diesters (see text for discussions).

kcal/mol [160]. Since LFER data indicate that the nature of TS is loose in both AP and solution [30, 160], these results suggest that R166S AP binds a loose TS for monoesters by ∼21 kcal/mol. Since diesters feature less charge and are promiscuous substrates of (R166S) AP, we expect that the binding energy of a loose TS for diesters to R166S AP is substantially lower than 21 kcal/mol. Therefore, we don't expect that R166S AP is able to shift the TS for diesters to be much looser in nature, in agreement with our QM/MM calculations.

For NPP, the diesters are cognate substrates, thus it is conceivable that the active site has been evolved to optimize the catalysis of diester hydrolysis, which might involve modifying the nature of TS relative to solution. However, we note that at least for the diester substrate studied here, the calculated chemical step barrier in NPP (20.2 kcal/mol) is not significantly lower than that in solution (25.7 kcal/mol). Therefore, it is not unreasonable that the nature of TS in NPP is not significantly changed relative to solution and a significant component of rate enhancement over solution (k_{cat}/K_M relative to k_w) is due to substrate binding.

4.3.7 The effects of Zn^{2+}-Zn^{2+} distance on reaction energetics

Since the bimetallic zinc site is a prevalent catalytic motif [214–218], it is of interest to establish what features (e.g., structural vs. electrostatic) are important to the catalytic proficiency. In this work, motivated in part by the fact that the Zn^{2+}-Zn^{2+} distances behave rather differently in the current and previous QM/MM calculations [1, 179] of AP and NPP, we examine the effect of this fundamental structural feature on the catalysis in AP and NPP.

Specifically, we design two sets of simulations for R166S AP and NPP with $MpNPP^-$ (α orientation only) to study the effects of Zn^{2+}-Zn^{2+} distance fluctuation and variation. In the first set, we constrain the Zn^{2+}-Zn^{2+} distance close to the value in crystal structures of AP and NPP (4.1 Å). Due to this constraint, the Zn^{2+}-Zn^{2+} distance fluctuations in PMF simulations are significantly damped (root-mean-square-fluctuation, rmsf, of ∼0.01 Å) as compared to unconstrained simulations (rmsf∼0.2 Å). As shown in Fig.4.11a-b, the PMFs for AP and NPP do not exhibit much change, especially in the barrier height, due to the constraint; the nature of the TS (see **Supporting Information**) also does not change.

These observations suggest that fluctuation of the Zn^{2+}-Zn^{2+} distance *during* the reaction is not critical to the reaction barrier or the nature of the TS. The significant overlaps of the PMFs also indirectly support the reproducibility and convergence of our PMF simulations.

In the second set of simulations, the Zn^{2+}-Zn^{2+} distances are constrained at 3.6 and 4.6 Å, respectively, which represent the two extreme values observed in unconstrained simulations. The changes of PMF are similar in AP and NPP (see Fig.4.11c-d): the reactant state position is shifted to a less negative value and the barrier height is reduced as the Zn^{2+}-Zn^{2+} distance is decreased from 4.6 to 3.6 Å. In terms of the nature of TS, there is no qualitative change as reflected by the essentially invariant peak position of the PMFs. At the quantitative level, there are variations as reflected by the average tightness coordinate (see Table 4.4) and other structural parameters (see **Supporting Information**); for example, the tightness coordinate (TC) increases, as expected, as the Zn^{2+}-Zn^{2+} distance increases.

In short, these calculations explicitly demonstrate that the Zn^{2+}-Zn^{2+} distance of the bi-metallic zinc site plays an important role in tuning the catalysis. This is not unexpected since the distance between the zinc ions influences the electrostatic properties in the active site, which are crucial to the phosphoryl transfers [219]. Nevertheless, the results clearly underlines the importance of reproducing geometrical properties of a bimetallic site for a meaningful analysis of its catalytic properties.

4.3.8 Issues worthwhile investigating with future experiments

The key objective of this work is to characterize the nature of the hydrolysis TS of phosphate diesters in AP and NPP so as to evaluate different hypotheses [2, 32] regarding the catalytic promiscuity in these enzymes. To evaluate the main findings of this work, we propose that the following experimental studies are worthwhile.

First, unlike the recent QM/MM simulations [1, 2], our calculations suggest that the nature of diester hydrolysis TS is largely similar in AP, NPP and solution. Along this line, LFER and KIE studies for diester hydrolysis in NPP will be highly informative.

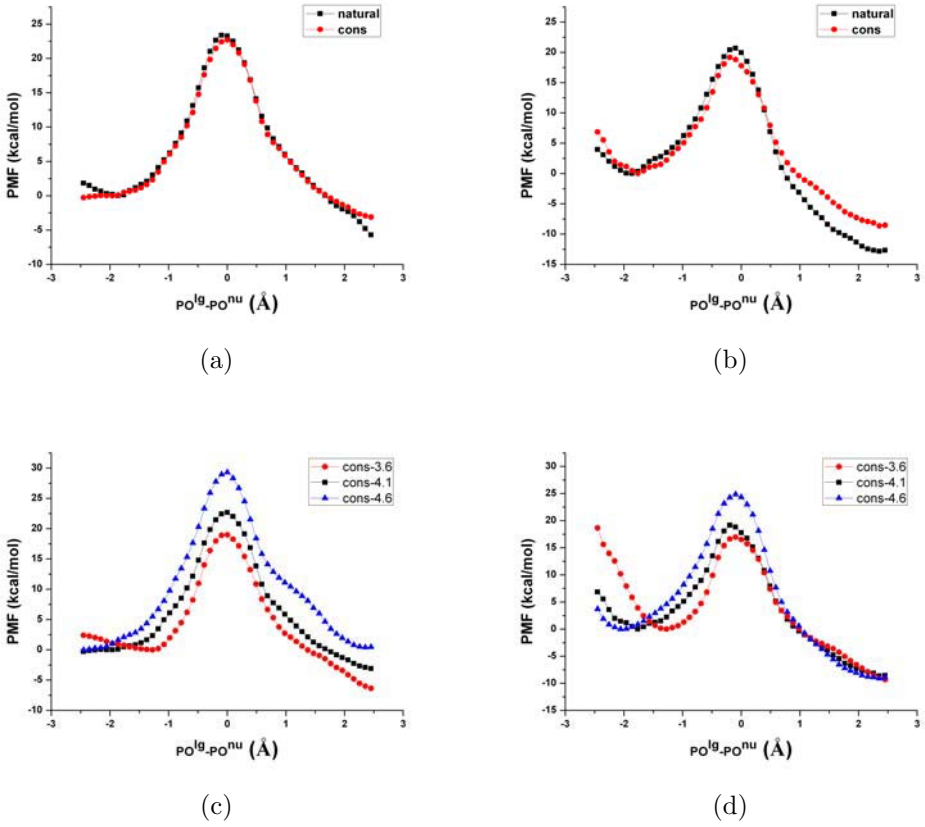


Figure 4.11: Potential of Mean Force (PMF, in kcal/mol, along the reaction coordinate defined as the difference between P-O^{lg} and P-O^{nu}) comparisons for $MpNPP^-$ hydrolysis in R166S AP and NPP with the Zn^{2+}-Zn^{2+} distance constrained at different values. (a) Between unconstrained and constrained (4.1 Å) simulations for R166S AP. (b) Between unconstrained and constrained (4.1 Å) simulations for NPP. (c) Between constrained simulations at 3.6, 4.1 and 4.6 Å for R166S AP. (d) Between constrained simulations at 3.6, 4.1 and 4.6 Å for NPP. For structural information, see Table 4.4 and **Supporting Information**.

Second, our calculations find that the leaving group does not interact strongly with the zinc ion in either the reactant or TS for diester hydrolysis in AP/NPP. This is rather unexpected considering the crystal structures of vanadate-AP/NPP complexes [27, 210] and therefore worth further investigations. Note that this result does not suggest that the second zinc ion does not play an essential role in catalysis since it coordinates with the bridging oxygen. The leaving group is stabilized by active site solvent molecules, suggesting that the dependence of the catalytic rate with respect to the substitution of the leaving group, including both chemical and isotope substitutions, should be close to that in solution.

We have extensively discussed in Sect.4.3.3 the issue of different binding orientations of the substrate and relation of current calculations with available experimental studies. As highlighted by the NBO charges shown in Fig.4.8, one complicating factor for the interpretation of thiol experiments is that phosphorothioate and phosphate esters have rather different charge distributions; in $MpNPPS^-$ the oxygen bonded with the methyl group is in fact more negatively charged than the sulfur. Therefore, it will be valuable to explore the reactivity of thiol substrates with both well-defined stereochemistry and relative charge distributions closer to the phosphate esters, such as by replacing both O3 and O4 in MpNPP by sulfur.

Finally, as alluded to in Sect.4.3.7, the metal-metal distance appears to be essential to both the barrier height and the nature of the transition state; the barrier is higher and the nature of TS looser with a longer metal-metal distance. This can be tested by substituting the zinc ions with other metals of different size (e.g., replacing Zn^{2+} by Co^{2+}) and performing kinetic and LFER analyses. Along this line, analysis of TS nature with synthetic analogs of di-metallic zinc motifs [220, 221] will be highly informative, since the structure of such artificial catalysts can be better controlled and monitored.

4.4 Concluding remarks

In this work, we have studied the hydrolysis of $MpNPP^-$ and its several analogs in solution, two experimentally well characterized variants of AP (R166S and R166S/E332Y) and wild type NPP using SCC-DFTBPR/MM simulations. The main goal is to investigate

whether the nature of phosphoryl transfer transition state for the same substrate is significantly different in these enzymes and in solution, a question that has a direct implication to the remarkable catalytic promiscuity exhibited by members of the AP superfamily.

Overall, our results are consistent with available experimental observations. In solution, we show that our calculations are able to capture trends in the hydrolysis barrier for $MpNPP^-$ and its two analogs, and the reaction proceeds through a synchronous TS, consistent with expectation based on experimental LFER data. For several variants of AP, our simulations support that the nature of TS is not perturbed significantly relative to solution, with a small degree of tightening due presumably to the interaction with the bi-metallic zinc motif; these are also in qualitative agreement with LFER data for AP variants. Therefore, our calculations support the picture that although the native function of AP is to catalyze the hydrolysis of phosphate monoesters through a loose transition state, its active site accommodates the tighter transition state for diesters. Such active site "plasticity" has been proposed to be related to the catalytic promiscuity of AP. Our analysis (Fig.4.10) of the free energy surfaces for phosphate diester hydrolysis in solution and AP/NPP supports a simple interpretation of such functional plasticity: the binding of AP to diester substrates is simply not strong enough, presumably due to their lower charge compared to monoesters, to significantly shift the nature of the transition state from synchronous to loose.

For NPP, for which diesters are cognate substrates, our calculations also do not support any major change in the nature of TS relative to solution, in contrast to the recent QM/MM calculations [1]. Since both structural features of the active site and energetics from our calculations are in better agreement with available experimental data, we expect that the nature of TS from the current work is more realistic. The lack of any significant change in the nature of TS relative to solution is not unexpected considering that the calculated chemical step barrier in NPP is not significantly lower than that in solution. Nevertheless, we hope the current work will stimulate additional LFER studies for NPP to further confirm the nature of TS.

The calculations also reveal several features and underlying complexity of AP catalysis not thoroughly recognized by previous work. For example, concerning the orientation of a diester substrate in the AP active site, our calculations for two variants of AP and two diester substrates collectively indicate that the experimental data alone can't be used to unambiguously show that a single orientation (α) is the only reactive binding mode. In fact, we find that the β orientation with the substrate methyl group pointing toward the Ser102 backbone amide has a reaction barrier lower than that for the α orientation, which has the substrate methyl group pointing toward the Mg^{2+} site; however, it is possible that the binding free energy for the α orientation is larger to make the overall free energy profile at least comparable to the β orientation. We also argue that the thio-substitution experiments are not always straightforward to interpret, because there is nontrivial differences in the charge distributions for phosphorothioate and phosphate esters; the possibilities of water penetrating into the active site for certain orientation of the (thio-substituted) substrate and that different substrate orientations dominate in different variants of AP (R116S vs. R116/E332Y) further compromise the clarity of interpretation of thio effects. Finally, we discuss results supporting that the Zn^{2+}-Zn^{2+} distance plays a significant role in modulating the energetics, especially barrier, of phosphoryl transfer in AP and NPP. This result can be probed experimentally by metal substitution and underlines the importance of reproducing geometrical properties of a bimetallic site for a meaningful computational analysis of its catalytic properties.

For a more thorough understanding of catalytic promiscuity and functional evolution of the AP superfamily, it is crucial to carry out similar systematic benchmark and analyses for monoester hydrolysis in solution and AP family enzymes. Since phosphate monoesters bear higher charges than diesters, and that SCC-DFTBPR has been developed based mainly on diesters, it is likely that further improvements of the computational methodology are needed [208, 209]. Once validated for AP and NPP, the computational approach coupled with other methodological advances [222–225] is potentially applicable in the prediction and rational design of catalytic promiscuity in other enzyme families.

Chapter 5

QM/MM studies of Linear Free Energy Relationship of a series of phosphate diesters in solution and Alkaline Phosphatase superfamily

5.1 Introduction

With the increasing recognition that many enzymes have promiscuous catalytic activities besides their high catalytic proficiencies for cognate substrates, enzyme promiscuity has become an interesting subject and attracts more and more studies. Understanding the principles that control enzyme promiscuity and their significance in physiological and evolutional functions can benefit our understanding of enzyme catalysis and provide invaluable directions for related engineering work. [7–13, 172, 173] In this context, members in the Alkaline Phosphatase (AP) superfamily present striking examples of catalytic specificity and promiscuity. [25, 26] They have been demonstrated to catalyze the hydrolysis of a broad range of substrates that differ in charge, size, intrinsic reactivity and transition state (TS) nature. [174] For example, *E. coli* AP mainly catalyzes the hydrolytic reaction of phosphate monoesters, presumably to harvest phosphate for nucleic acids and metabolites, but also exhibits promiscuous activities for the hydrolysis of phosphate diesters and sulfate esters. The catalytic proficiencies (defined as $k_{cat}/K_M/k_w$) ranges from $> 10^{27}$ for the cognate activities [28, 175] to $\sim 10^6$ for the promiscuous activities. [176] Similarly, although the main function of Nucleotide pyrophosphatase/phosphodiesterase (NPP) is to hydrolyze phosphate diesters, it can also cleave phosphate monoesters and sulfate esters with considerable acceleration over solution reactions. The reaction specificities of AP and NPP for phosphate mono-

and di-esters differ by up to a remarkable level of 10^{15} fold! [27,28] This is particularly striking in light of the fact that AP an NPP have very similar active sites. These interesting features make this pair of enzymes ideal for in-depth comparative analyses.

Extensive experimental work through the past years has gleaned precious understanding of AP and NPP catalysis. Crystal structures [27, 177] demonstrate many similarities in their active sites: a bimetallo zinc site with the same six ligands (3 Asp, 3 His) exists in both enzymes; an arginine/asparagine residue in AP/NPP is positioned to provide favorable interactions with the substrate; a serine/thereonine alkoxide displaces the leaving group in the first step of the reaction to produce a covalent enzyme-phosphate intermediate (Fig.5.1). Besides these similarities, several differences also exist: in AP there are extra positive charged motifs including a third Mg^{2+} ion that coordinates with one of the Zn^{2+} ligand (Asp51) and lies within hydrogen bond distance of the substrate and the serine alkoxide; in NPP, a hydrophobic pocket close to the active site provides extra stabilization for its cognate substrate. With respect to their functions, it has been proposed that AP and NPP stabilize TSs closely related to solution analogs as this would require the least amount of stabilization. Indeed, linear free energy relation (LFER) [178] and kinetic isotope effect (KIE) data [29–33] strongly suggest that AP and NPP catalyze phosphate monoester and diester reactions via loose and synchronous TSs analogous to solution reactions.

However, recent QM/MM calculations [1,2,179] do not support this proposal. The calculations found that phosphate monoester hydrolysis in AP proceeds via a two-step mechanism, fundamentally different from the one-step mechanism with a loose TS in solution; although similar to aqueous reactions, a one-step mechanism is also adopted for phosphate diesters, the TS was found to change from synchronous in solution to very loose in AP and NPP. [1,2] Therefore, the computational work reached completely distinct conclusions from experimental studies of AP and NPP catalysis. [32] Nevertheless, it should be noted that whether the computational method was sufficiently reliable in previous QM/MM studies is not clear; for example, the Zn^{2+}-Zn^{2+} distance was found to vary greatly during the reaction, reaching 7.0 Å as compared to the value of ∼4 Å in the crystal structure.

To better address the conflicts between previous experimental and theoretical studies, in our recent paper, [58] we studied the hydrolysis of a phosphate diester, $MpNPP^-$, in solution, two experimentally well-characterized variants of AP (R166S AP, R166S/E322Y AP) and wild type NPP by carefully benchmarked QM/MM calculations. The good agreements for structural and energetic properties in solution and enzyme with available experimental data support the use of our enzyme model for a semi-quantitative analysis of the catalytic mechanisms in AP and NPP. The calculations suggest that the hydrolysis reactions of phosphate diesters catalyzed by AP and NPP feature similar synchronous transition states that are slightly tighter compared to in solution. Therefore, it supports the proposal based on previous experimental observations that enzymes in the AP superfamily catalyze cognate and promiscuous substrates via similar transition states to those in solution; it does not support the finding of previous QM/MM study, which suggested that the same diester substrate goes through a very loose transition state in AP and NPP, a result likely biased by the large structural distortion of the bimetallic zinc site in their simulations.

As a following work, we further analyze two similar aryl phosphate diesters, $MmNPP^-$ and MPP^- (see Fig. 5.2) in R166S AP and NPP which have been systematically studied by LFER. Together with the previous work of $MpNPP^-$, these efforts serve as a more stringent benchmark of our enzyme models. Although we successfully reproduce the correct trend of experimental measured reaction energetics for these similar substrates, the substitution effects of the leaving group are over exaggerated, mainly due to the semi-empirical feature of the QM method we use as indicated by the model benchmark analysis. Therefore, we further explore the possible approaches of adding corrections by a one-step free energy perturbation (FEP) by high level *ab initio* QM methods to improve the quantitative agreement with experimental data.

The paper is organized as follows: in Sect.5.2 we summarize computational methods and simulation setup. In Sect.5.3, we first present results for enzyme PMF calculations, and then analysis of the errors in those calculations; we also explore the FEP corrections for

the intrinsic error of the semi-empirical QM method we use. Finally, we summarize a few conclusions in Sect.5.4.

5.2 Computational Methods

5.2.1 Enzyme Model Setup

The construction of enzyme model is similar to our previous study [58] so we only summarize several key points briefly. The enzyme models are constructed based on the X-ray structures for the *E. coli* AP mutant R166S with bound inorganic phosphate at 2.05 Å resolution (PDB code 3CMR [182]) and *Xac* NPP with bound Adenosine Mono-Phosphate (AMP) at 2.00 Å resolution (PDB code 2GSU [27]). Starting from the PDB structure, the ligand is first "mutated" to the α orientation of substrate of interest with the -OMe group oriented towards the magnesium ion. Hydrogen atoms are added by the HBUILD module [188] in CHARMM. [189] All basic and acidic amino acids are kept in their physiological protonation states except for Ser102 and Thr90 in AP and NPP, respectively, which are assumed to be the nucleophiles and deprotonated in the reactive complex. Water molecules are added following the standard protocol of superimposing the system with a water droplet of 27 Å radius centered at $Zn1^{2+}$ (see Fig.5.1 for atomic labels) and removing water molecules within 2.8 Å from any atoms resolved in the crystal structure. [161] Protein atoms in the MM region are described by the all-atom CHARMM force field for proteins [190] and water molecules are described with the TIP3P model. [162] The QM region includes the two zinc ions and their 6 ligands (Asp51, Asp369, His370, Asp327, His412, His331), Ser102 and the substrate for R166S AP; for NPP, this includes two zinc ions and their 6 ligands (Asp54, Asp257, His258, Asp210, His363, His214), Thr90 and the substrate. Only side chains of protein residues are included in the QM region and link atoms are added between C_α and C_β atoms. The treatment of the QM/MM frontier follows the DIV scheme in CHARMM. [191] A similar NOE potential is added to the C-O bonds in Asp51 in AP as before.

The SCC-DFTBPR method [46] is used for PMF calculations with the generalized solvent boundary potential (GSBP) [124, 163] setup. The system is partitioned into a 27-Å

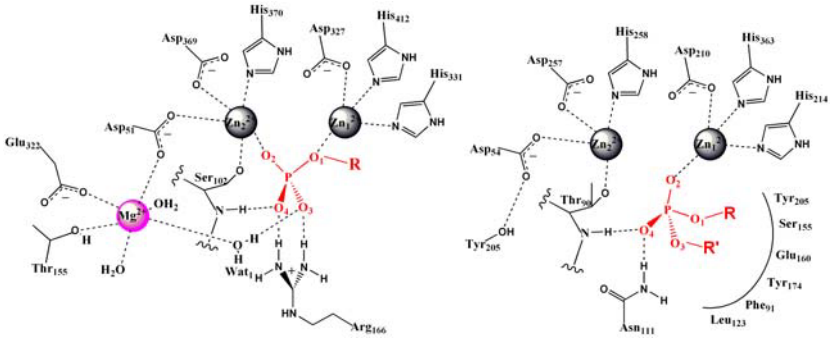


Figure 5.1: The active sites of Alkaline Phosphatase (AP) and Nucleotide PyrophosPhatase/ phosphodiesterase (NPP) are generally similar, with a few distinct differences. (a) *E. coli* AP active site. (b) *Xac* NPP active site. The cognate substrates for AP and NPP are phosphate monoesters and diesters, respectively. The labeling scheme of substrate atoms is used throughout the paper.

spherical inner region centered at the Zn1 atom, with the rest in the outer region. Newtonian equations-of-motion (EOM) are solved for the MD region (within 23 Å), and Langevin EOM are solved for the buffer regions (23-27 Å) with a temperature bath of 300 K; protein atoms in the buffer region are harmonically constrained with force constants determined from the crystallographic B-factors. [193] All bonds involving hydrogen are constrained using the SHAKE algorithm, [166] and the time step is set to 1 fs. All water molecules in the inner region are subject to a weak GEO type of restraining potential to keep them inside the inner sphere with the MMFP module of CHARMM. The static field due to outer-region atoms, ϕ_s^{io}, is evaluated with the linear Poisson-Boltzmann (PB) equation using a focusing scheme with a coarse cubic grid of 1.2 Å spacing, and a fine grid of 0.4 Å spacing. The reaction field matrix $\mathbf{M}$ is evaluated using 400 spherical harmonics. In the PB calculations, the protein dielectric constant of $\epsilon_p = 1$, the water dielectric constant of $\epsilon_w = 80$, and 0.0 M salt concentration are used. The optimized radii of Nina et al. [194, 195] based on experimental solvation free energies of small molecules as well as the calculated interaction energy with explicit water molecules are adopted to define the solvent-solute dielectric boundary. To be consistent with the GSBP protocol, the extended electrostatic model [164] is used to treat the electrostatic interactions among inner region atoms in which interactions beyond 12 Å are treated with multipolar expansions, including the dipolar and quadrupolar terms.

5.2.2 Potential of mean force (PMF) simulations

To study the free energy profiles of enzyme reactions, PMF simulations have been carried out for R166S AP and NPP with MmNPP$^-$ and MPP$^-$ as the substrates. After the initial minimizations starting from the relevant initial structure, the enzyme system is slowly heated to 300 K and equilibrated for 100 ps. The reaction coordinate is defined as PO^{lg}-PO^{nu}. The umbrella sampling approach [167] is used to constrain the system along the reaction coordinate by using a force constant of 150 kcal/mol·$\mathring{A}^{-2}$. In total, more than 51 windows are used for each PMF and 100 ps simulations are performed for each window. The first 50 ps trajectories are discarded and only the last 50 ps are used for data analysis. Convergence of

the PMF is monitored by examining the overlap of reaction coordinate distributions sampled in different windows and by evaluating the effect of leaving out segments of trajectories. The probability distributions are combined together by the weighted histogram analysis method (WHAM) [168] to obtain the PMF along the reaction coordinate. The averaged key structural properties for each window are calculated and summarized in Table 5.3.

5.2.3 Active site model benchmark calculations

To analyze the source of errors in PMF calculations, an active site model is constructed including all atoms in the QM region in QM/MM enzyme model with the valence saturated by hydrogen atoms. The β carbon and the link hydrogen atoms are fixed at their positions in crystal structure during geometry optimization. The reactant (Michaelis) complex and transition state are located for MpNPP$^-$, MmNPP$^-$ and MPP$^-$ under SCC-DFTBPR and B3LYP [196–198] with 6-31G* [199] basis set. Then single point energy calculations are carried out by B3LYP, M06 [226] and MP2 methods at 6-311++G** level. The calculations are carried out with the CHARMM [189] and Gaussian09 [227] software packages respectively.

5.2.4 M06/MM correction

As demonstrated by model benchmarks, there are systematic errors in SCC-DFTBPR for the energetics of phosphate diester reactions in AP, so it is necessary to include high level QM method corrections. M06 functional is used with 6-31+G** basis set which appears to give the best balance between accuracy and computational cost. The comparison with a large basis set (6-311++G**) used in model calculations indicates negligible differences. In addition, in the SCC-DFTB/MM method, the electrostatic interaction between QM and MM atoms is calculated based on point charge interactions

$$\hat{H}_{elec}^{SCC/MM} = \sum_{I \in MM} \sum_{J \in QM} \frac{q_I \Delta q_J}{r_{IJ}} \tag{5.1}$$

where Δq_J is the Mulliken charge on QM atom J. The more rigorous QM/MM coupling treatment includes the contribution from the MM point charges in the one-electron

integrals, which is done in M06/MM calculations. Therefore, correcting SCC-DFTB/MM results based on M06/MM calculations improves not only the QM level, but the QM/MM interactions as well. The correction is done on the basis of a straightforward one-step free energy perturbation calculation

$$\Delta G_{M06-SCC} = -kTln < e^{-\beta(U_{M06/MM}-U_{SCC/MM})} >_{SCC/MM} \tag{5.2}$$

at both end states ($\lambda = 0.0$ or 1.0). The difference between the perturbative correction at the two end states gives the M06/MM correction to the reaction free energy. Since only a small number of snapshots from SCC-DFTB/MM trajectories are used, a second-order cumulant expansion is used to improve the numerical stability of the perturbation calculation

$$\Delta G_{M06-SCC} =< U_{M06/MM} - U_{SCC/MM} >_{SCC/MM} - \frac{\beta}{2}[< (U_{M06/MM} - U_{SCC/MM})^2 >_{SCC/MM} - < U_{M06/MM} - U_{SCC/MM} >^2] \tag{5.3}$$

As discussed extensively in the literature, [228] such one-step perturbation is effective only if the configuration space at the two levels overlaps significantly; this is assumed to be the case considering the previous observations [105, 229] that SCC-DFTB often gives reliable geometries and energetics compared to DFT.

5.3 Results and Discussions

5.3.1 PMF for the first step of a series of phosphate diester reactions in R166S AP and NPP

AP and NPP feature a highly conserved bi-metallo zinc active site with essentially the same set of metal ligands. They catalyze the hydrolytic reactions of various phosphates via a two-step mechanism: an oxygen nucleophile first attacks the phosphorus, then a water (hydroxide) replaces the leaving group in a step that is essentially the reverse of the first. In a previous study, we explored the PMF of a particular phosphate diester, $MpNPP^-$, in AP and NPP and demonstrated that synchronous TSs are favored similar to those in

solution reaction, consistent with previous experimental observations. As a following work, we calculate the PMF of two similar phosphate diesters: MmNPP$^-$ and MPP$^-$ (see Fig. 5.2) in R166S AP and NPP by similar ways. The corresponding aqueous reactions have been studied in our previous work. For simplicity, we only studied the α orientation of different substrates in AP; for NPP, it is proposed that the extra methyl group points toward a hydrophobic groove nearby due to favorable hydrophobic interactions. [27]

Experimentally, the reaction barriers calculated from k_{cat}/K_m by transition state theory increase from 18.0 to 20.9 with the increase of leaving group pKa. This correlation (LFER) has been used to support the hypothesis that AP catalyzes phosphate diester reactions via a synchronous TS similar to aqueous reactions. [32] It is worth mentioning that the measured reaction barrier composes of two parts: the substrate binding that corresponds to K_m and the chemical step that corresponds to k_{cat}. A complete comparison with computation requires not only the PMF of the reaction pathway, but the binding free energy as well. A quantitative estimation of the latter requires much more efforts, due to the difficulty of obtaining converged results. Therefore, we only focus on the calculation of the chemical step (k_{cat}) in this work, for which the experimental data set the lower limit. Nevertheless, considering the high degree of similarities of the substrates structures and chemical properties, the differences of binding free energies are likely to be small, so the general trend of the reaction barriers for the chemical step (k_{cat}) is likely to resemble the LFER (k_{cat}/K_m).

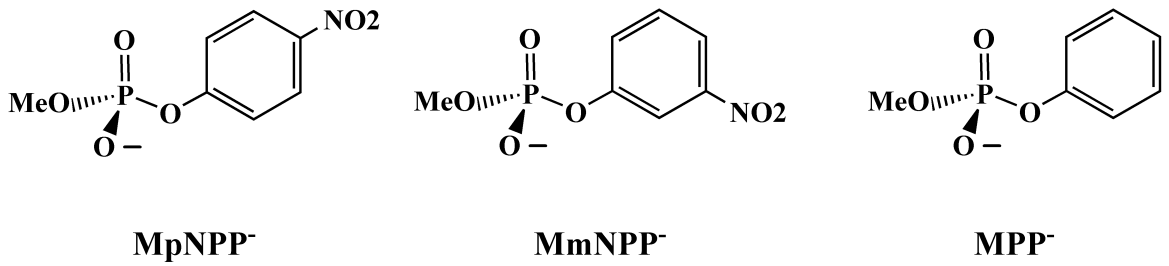


Figure 5.2: Methyl p-nitrophenyl phosphate (MpNPP$^-$) and its two diester analogs studied in this work.

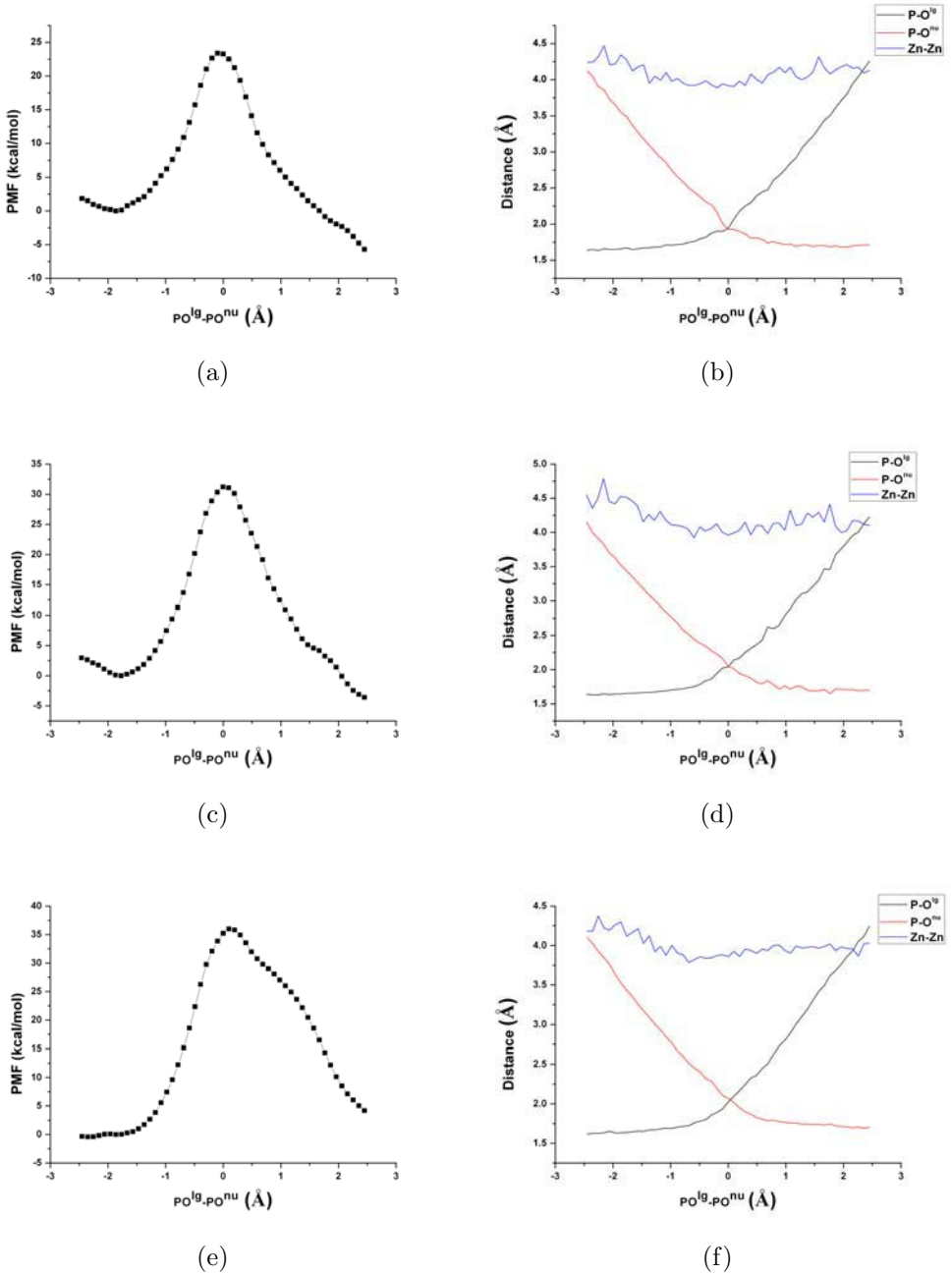


Figure 5.3: Potential of Mean Force (PMF) calculation results for $MpNPP^-$, $MmNPP^-$ and MPP^- hydrolysis in R166S AP. Energies are in kcal/mol. (a) $MpNPP^-$ PMF along the reactant coordinate (the difference between P-O^{lg} and P-O^{nu}); (b) $MpNPP^-$ changes of average key distances along the reaction coordinate; (c) $MmNPP^-$ PMF along the reactant coordinate; (d) $MmNPP^-$ changes of average key distances along the reaction coordinate; (e) MPP^- PMF along the reactant coordinate; (f) MPP^- changes of average key distances along the reaction coordinate.

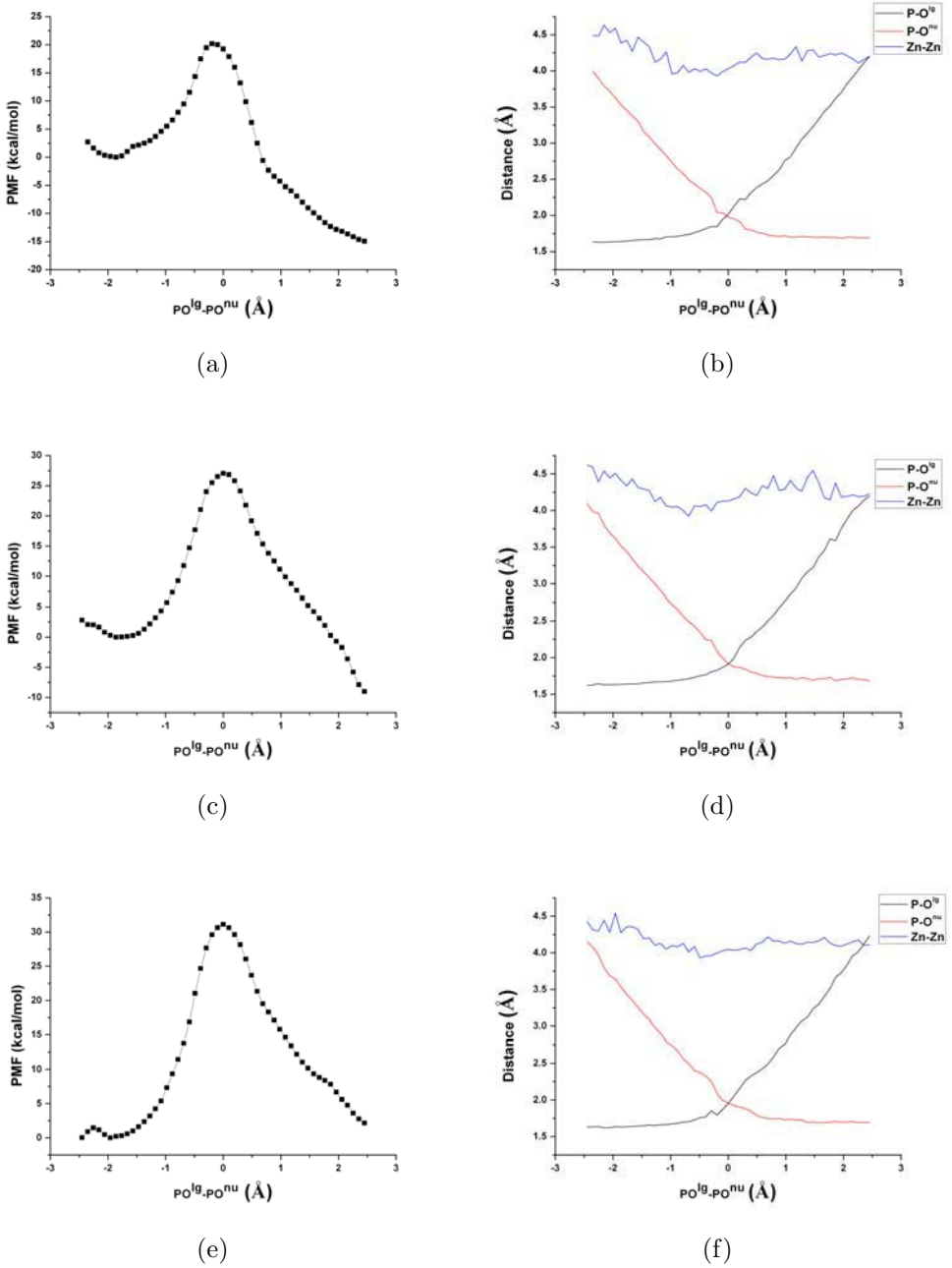


Figure 5.4: Potential of Mean Force (PMF) calculation results for $MpNPP^-$, $MmNPP^-$ and MPP^- hydrolysis in NPP. Energies are in kcal/mol. (a) $MpNPP^-$ PMF along the reactant coordinate (the difference between P-O^{lg} and P-O^{nu}); (b) $MpNPP^-$ changes of average key distances along the reaction coordinate; (c) $MmNPP^-$ PMF along the reactant coordinate; (d) $MmNPP^-$ changes of average key distances along the reaction coordinate; (e) MPP^- PMF along the reactant coordinate; (f) MPP^- changes of average key distances along the reaction coordinate.

Table 5.1: Diester hydrolysis reaction in R166S AP and NPP from experiments and calculations

	Substrate	Exp[a]	SCC/MM[b]	M06/MM[c]
R166S AP	$MpNPP^-$	18.0	23.4	24.4±7.9
	$MmNPP^-$	18.4	31.1	29.4±6.8
	MPP^-	20.9	36.0	29.7±9.6
NPP	$MpNPP^-$	14.3	20.2	
	$MmNPP^-$		27.0	
	MPP^-		31.1	

a. Free energy barrier (kcal/mol) calculated by transition state theory at 300 K based on experimental k_{cat}/K_M value; b. Only α orientation is considered for R166S AP; the calculated results correspond to the k_{cat} in experiment; c. SCC/MM results after M06/MM corrections.

The calculated PMFs of the first step in AP and NPP catalysis for these aryl phosphate diesters (Fig. 5.3,5.4) are similar to $MpNPP^-$. The reaction mechanism is a one-step reaction with a TS peaking at the reaction coordinate (PO^{lg}-PO^{nu}) around 0 Å. The reaction barriers increase (see Table 5.1) with the increase of leaving group pKa, consistent with the trend in LFER. The calculated result for each substrate is higher than the experimental result (k_{cat}/K_M) and the difference, in principle, corresponds to the contribution from binding free energies. Although the LFER of phosphate diesters in NPP has been not measured before, according to the estimation from our calculations, it should be similar to that of AP.

From a more quantitatively point of view, it is obvious that the nitro group substitution effects of different substrates are overestimated (Table 5.1). For example, the difference of k_{cat}/K_m between $MpNPP^-$ and MPP^- in R166S AP is only 2.9 kcal/mol, but the computational result gives 12.6 kcal/mol. A similar overestimation has also been observed for aqueous reactions in our previous work and the analysis suggests that it is due to the intrinsic accuracy of the SCC-DFTBPR method. [58] To evaluate the contribution from the

QM method, we construct an active site model by taking the QM region out of the enzyme model. The TSs of different substrates are rigorously calculated by B3LPY/6-31G* and SCC-DFTBPR. The structures from B3LYP calculations are further used for high level energy corrections, including B3LYP, M06 and MP2 at the 6-311++G** basis set level (Fig. 5.5). The MP2 method has been shown to give good estimation of gas phase proton affinity compared with experimental values while DFT type of methods tend to produce quite large errors. The M06 functional is parametrized including both transition metals and nonmetals and has been demonstrated by extensive benchmarks for it excellent performance for application in organometallic and inorganometallic chemistry. In the calculated reactant and TS geometries (Fig.5.5), several important features, such as the P-O bond distances and Zn^{2+}-Zn^{2+} distance, are closely reproduced by SCC-DFTB compared to B3LYP, indicating a good agreement of structural properties. For the reaction energetics, indeed, the substitution effects are once again over estimated by SCC-DFTB and B3LYP (Table 5.2). For example, the difference of reaction barriers between $MpNPP^-$ and MPP is 5.4 kcal/mol by MP2 but SCC-DFTB and B3LYP give 15.7 and 9.2 kcal/mol, respectively. Alternatively, the M06 functional achieves better agreement with MP2, thus could be a good candidate to correct the intrinsic errors of SCC-DFTBPR. Overall, the analysis indicates that the large errors of substitution effects in PMF calculations are likely stemming from the semi-empirical feature of the QM method we use. With the use of DFTB3 by adding the full third order expansion [230] and a systematic reparametrization for phosphate hydrolysis that is underway, improvements are expected.

In Fig.5.3,5.4, several key structural properties in enzyme active site are averaged over the trajectory of each window and plotted as functions of the reaction coordinate. The averaged Zn^{2+}-Zn^{2+} distance slightly decreases from reactant state to TS, then increases again in the product region. The changes of P-O^{lg} and P-O^{nu} bond lengths of all substrates in AP and NPP demonstrate similar features, indicating a concerted pathway with a synchronous TS that is similar to aqueous reaction. Comparing different TSs, the P-O^{lg} bond length increases while P-O^{nu} bond length decreases from $MpNPP^-$ to MPP^- in AP and NPP (see

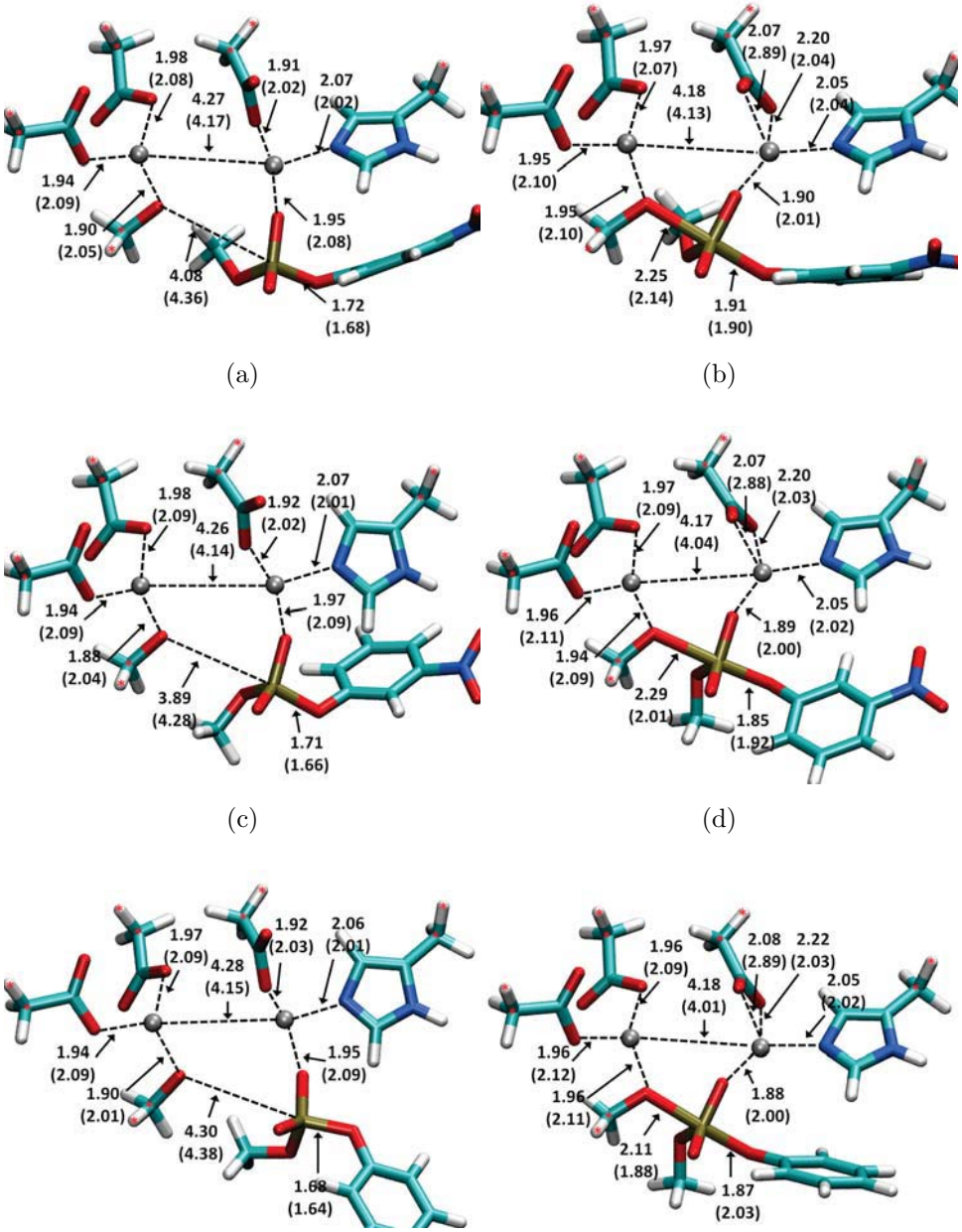


Figure 5.5: AP active site model with MpNPP$^-$, MmNPP$^-$ and MPP$^-$. Geometries are optimized in gas phase by B3LYP/6-31G*. (a) MpNPP$^-$ reactant state; (b) MpNPP$^-$ TS; (c) MmNPP$^-$ reactant state; (d) MmNPP$^-$ TS; (e) MPP$^-$ reactant state; (f) MPP$^-$ TS.

Table 5.2: MEP results for diester hydrolysis reaction in enzymes by a cluster model

Substrate	SCC	B3LYP	M06	MP2
$MpNPP^-$	7.8	11.2	4.2	4.8
$MmNPP^-$	14.1	11.2	7.7	10.1
MPP^-	23.5	20.4	13.0	10.2

a. Basis set is 6-311++G**.

Table 5.3), which are consistent with the trends in our previous solution calculations, [58] probably due to the better water stabilization for leaving group with the increase of its basicity. This underscores the potential role that solvent water may play due to the open active site feature in AP and NPP as we cautioned in our previous work. The TS tightness coordinate ($TC=PO^{lg}+PO^{nu}$) increases from 3.89 in AP and 3.86 in NPP for $MpNPP^-$ to 4.11 and 3.91 for MPP^- respectively, slightly decreased compared with solution reactions due to the constraints from bimetallo zinc motif. This is consistent with our previous conclusion that the nature of TS for phosphate diester is slightly tightened from solution to enzyme.

As shown in the averaged structures (Fig.5.6), the reactant states for $MmNPP^-$ and MPP^- in R166S AP are similar to $MpNPP^-$ in which one magnesium ligand water (Wat1) forms a hydrogen bond with the deprotonated Ser102 oxygen; one nonbond oxygen binds with Zn1 while the other forms a hydrogen bond with Ser102 backbone amide. In TSs, the hydrogen bond with Wat1 is almost broken and a new hydrogen bond is formed between Wat1 and the bridging oxygen which has been suggested to help lower the reaction barrier. Similarly, in the reactant state in NPP (Fig.5.7), one nonbond oxygen also binds with Zn1 while the other forms hydrogen bonds with Asn111 and Thr90 backbone amide. In TS, the substrate binding mode does not change with respect to different substrates; this is on contrary to previous theoretical studies [1] in which the authors find similar TS binding mode for $MpNPP^-$ as our results, however, for MPP^-, the leaving group oxygen also binds with Zn1 instead of being solvated by water as in our model. It would be much more informative

Table 5.3: Key structural properties of the transition states for the first step of phosphate diester hydrolysis in AP and NPP

	Substrate	RC[a]	TC[b]	P-O^{lg}	P-O^{nu}	Zn^{2+}-Zn^{2+}
R166S AP	$MpNPP^-$	-0.11±0.07	3.89±0.14	1.92±0.08	2.00±0.11	3.93±0.18
	$MmNPP^-$	-0.01±0.08	4.09±0.23	2.07±0.13	2.07±0.13	3.96±0.20
	MPP^-	0.08±0.07	4.11±0.18	2.10±0.10	1.98±0.11	3.93±0.18
NPP	$MpNPP^-$	-0.20±0.07	3.86±0.14	1.88±0.08	2.06±0.10	3.92±0.17
	$MmNPP^-$	-0.01±0.07	3.83±0.10	1.91±0.08	1.93±0.07	4.13±0.15
	MPP^-	0.00±0.08	3.91±0.16	1.98±0.11	1.96±0.09	4.05±0.18

a. The Reaction coordinate (RC) is defined as the difference between P-O^{lg} and P-O^{nu}; b. The Tightness coordinate (TC) is defined as the sum of P-O^{lg} and P-O^{nu}.

to find out the possible reason for this change if the corresponding energetic properties had been reported. On the other hand, as we discussed before, due to the synchronous nature of TS for phosphate diester, the extent of P-O^{lg} bond breaking and charge accumulation on leaving group oxygen are much less compared with monoester that goes through a loose TS, therefore the interactions between leaving group oxygen and Zn1 is likely to be much weaker and does not favor the bi-dentate coordination of Zn1.

5.3.2 Corrections of PMF by high level *ab initio* QM methods

The semi-empirical feature of the SCC-DFTBPR method allows reasonable amount of samplings of the relatively large QM region in the enzyme model at the cost of compromised accuracy, as demonstrated by the overestimation of the substitution effects and the cluster model analysis. However, it is encouraging to see that the structural properties calculated by SCC-DFTBPR are more reasonable, therefore allows a post correction scheme based on the sampled conformations. We explore the corrections by M06 functional that gives a good balance between accuracy and efficiency, as indicated in our model cluster benchmarks and abundant benchmarks in previous work. [226] A one-step free energy perturbation scheme is used to evaluate the difference of the energy surface by M06 functional and SCC-DFTBPR. The underlying assumption is that the sampled conformational space by SCC-DFTBPR is a reasonable estimation of that by M06, which is likely the case based on our model analysis. Due to the large number of snapshots in the calculation, a modest basis set (6-31+G**) is used. Our test of this basis set on the cluster model shows negligible difference compared to the larger basis set (6-311++G**).

Indeed, the overestimation of substitution effects is significantly reduced by M06/MM corrections (Table 5.1). The difference of reaction barriers between $MpNPP^-$ and MPP^- is reduced from 12.6 kcal/mol to 5.3 kcal/mol, much more consistent with the experimental results. Therefore, the one-step FEP correction can be useful to quantitatively improve the results by our enzyme models. However, the standard deviations of these corrections are typically 4-6 kcal/mol with 300 snapshots included. A scrutiny of the convergence of the

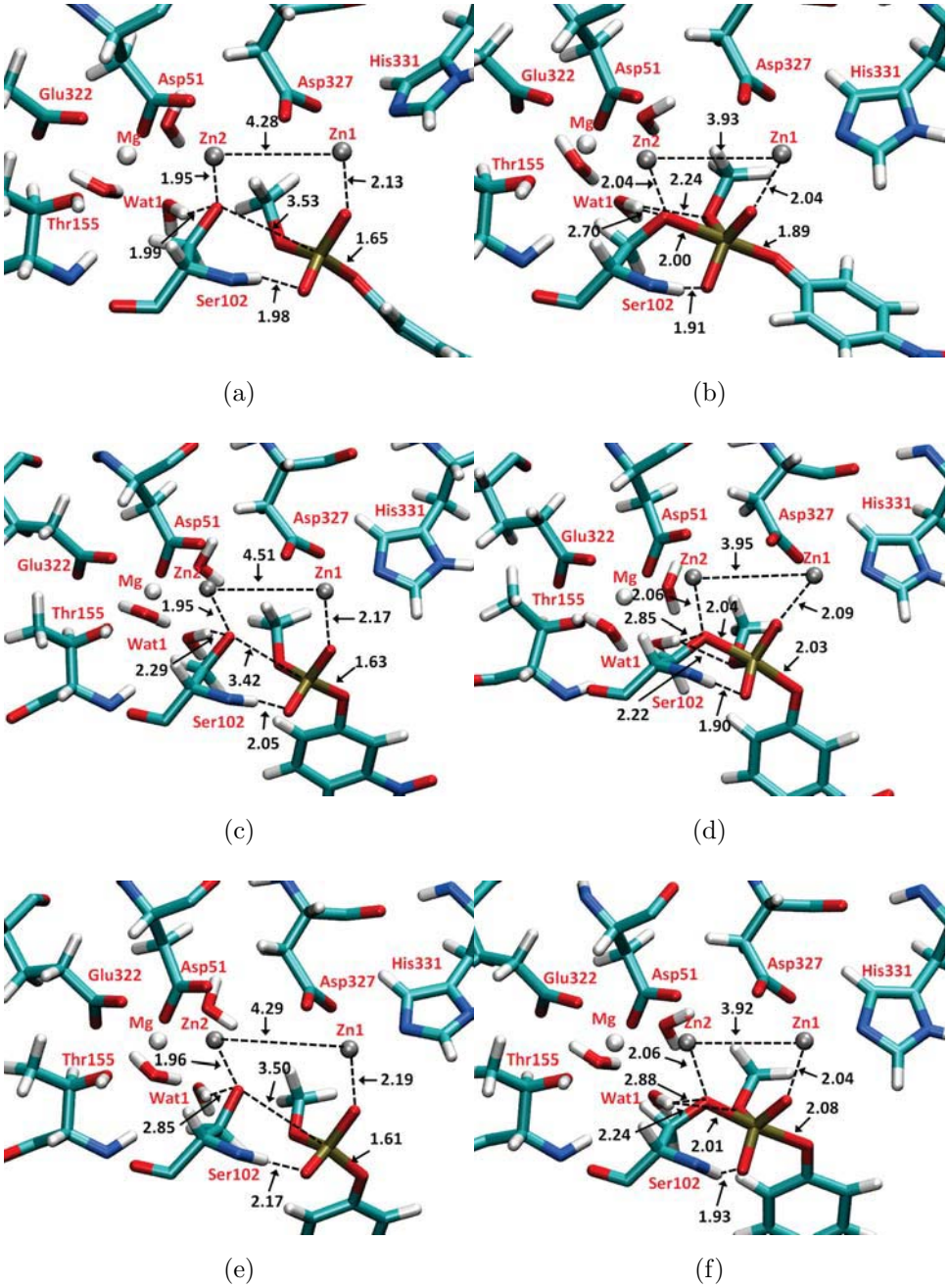


Figure 5.6: Snapshots of $MpNPP^-$, $MmNPP^-$ and MPP^- hydrolysis in R166S AP with average key distances labeled in Å. Asp369, His370 and His412 are omitted for clarity. (a) $MpNPP^-$ reactant state; (b) $MpNPP^-$ TS; (c) $MmNPP^-$ reactant state; (d) $MmNPP^-$ TS; (e) MPP^- reactant state; (f) MPP^- TS.

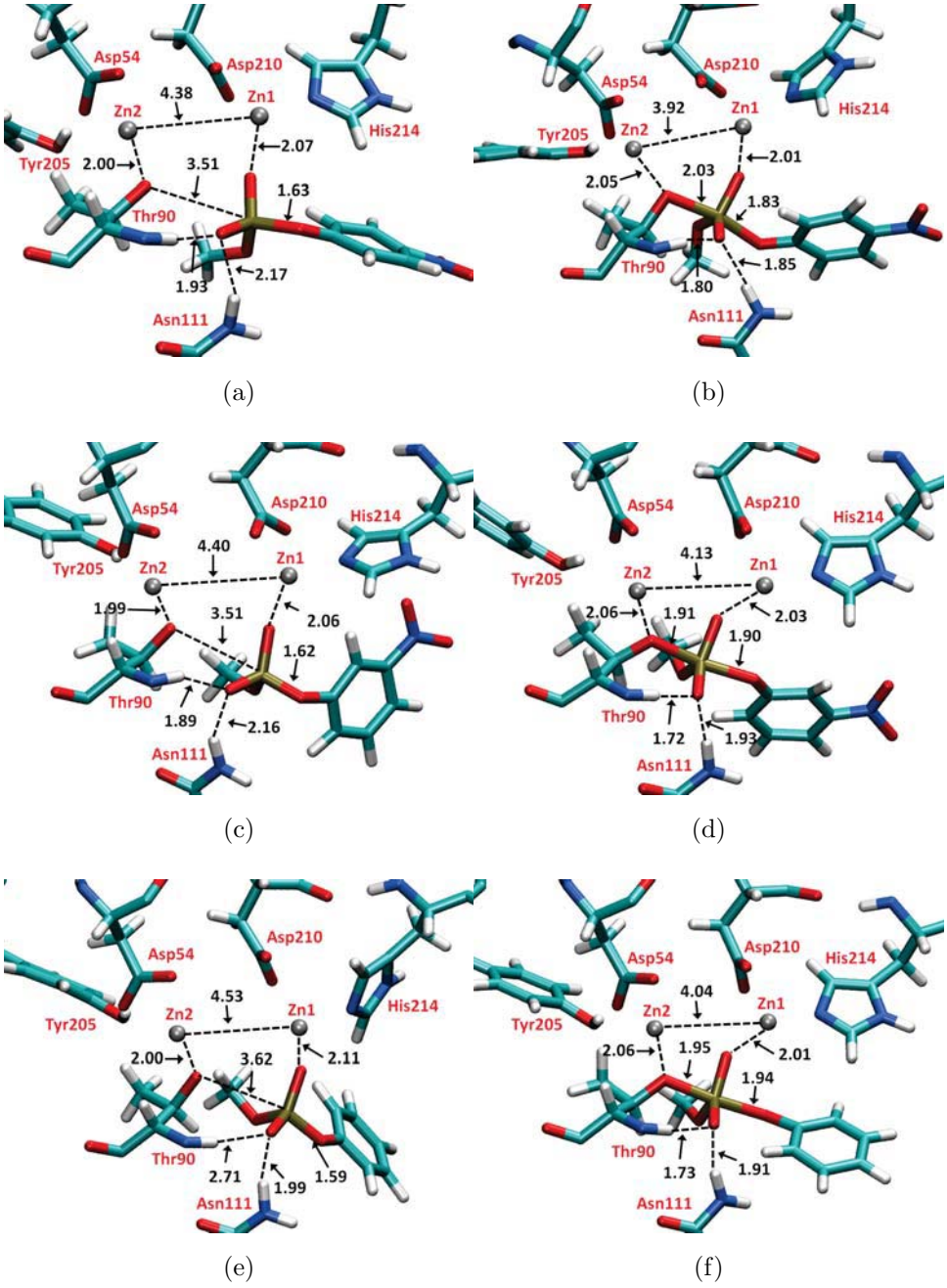


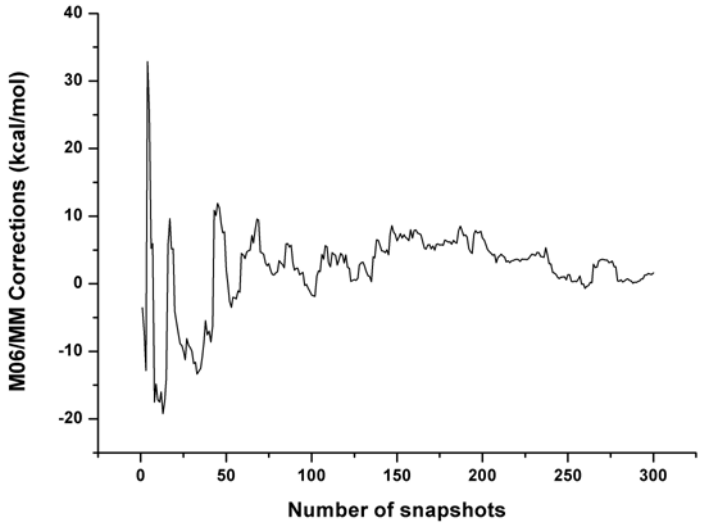
Figure 5.7: Snapshots of MpNPP$^-$, MmNPP$^-$ and MPP$^-$ hydrolysis in NPP with average key distances labeled in Å. Asp257, His258 and His363 are omitted for clarity. (a) MpNPP$^-$ reactant state; (b) MpNPP$^-$ TS; (c) MmNPP$^-$ reactant state; (d) MmNPP$^-$ TS; (e) MPP$^-$ reactant state; (f) MPP$^-$ TS.

corrections with respect to the number of snapshots (see Fig.5.8 for an example) indicate a large fluctuation within 100 snapshots. Even by including 300 snapshots, the fluctuation is still not small. Therefore, we caution a careful interpretation of the corrections and and a stringent check of the convergence before any quantitative conclusions can be reached.

5.4 Concluding remarks

The unique features of AP/NPP active site make it especially challenging for obtaining any meaningful results by computational studies: the open active site requires reliable treatment of solvent water perturbation; the bi-metallo zinc motif and the extra magnesium ion in AP require a good description of the electronic structures; several extra charged motifs and the negatively charged substrate require accurate account of charge-charge interactions. Therefore, it is highly desired to carry out careful and extensive benchmarks with respect to crucial experimental data to understand the creditability and limitation of the enzyme model and computation method. Under this context, we did more systematic studies of the reaction energetics of a series of phosphate diesters in solution, AP and NPP in this study. By using our enzyme model and SCC/MM method, we successfully reproduced the correct trend of the experimental LFER. By including M06/MM energy corrections for the intrinsic error in SCC method, a semiquantitative agreement can be achieved, highlighting the reliability of our model and the simulation protocol. We reported the first LFER of phosphate diesters in NPP that features the similar trend as R166S AP, enabling the comparison with future experimental data.

Our studies for all three phosphate diesters indicate that it is general for AP and NPP to utilize the similar synchronous TS in solution to catalyze the hydrolysis reactions for phosphate diesters. These results agree with the experimental results and the proposal that AP and NPP can recognize different substrates and catalyze them via similar TS to solution reactions, a hypothesis that has implications for enzyme promiscuity.



(a)

Figure 5.8: Convergence of M06/MM one-step free energy perturbation corrections with respect to the number of snapshots for $MpNPP^{-}$.

Chapter 6

QM/MM Studies of Phosphate Monoester Hydrolysis Reactions in Alkaline Phosphatase Superfamily

6.1 Introduction

Alkaline Phosphatase (AP) superfamily contains a set of evolutionarily related enzymes that are structurally related to AP. [20, 21] They catalyze the hydrolytic reactions of various phosphates and sulfates with distinct structures and charge states via a two-step mechanism: an oxygen nucleophile (e.g., Ser or Thr) first attacks the phosphorus/sulfur, then a water (hydroxide) replaces the leaving group in a step that is essentially the reverse of the first. In particular, as one of the most powerful enzymes, AP catalyzes phosphate monoester hydrolysis up to amazingly 10^{27} faster than solution reactions while maintains a lower activity (around 10^{11}) of phosphate diester hydrolysis. Conversely, another family member Nucleotide Pyrophosphatase/Phosphodiesterase (NPP) mainly catalyzes phosphate diesters with 10^{16} times speedup , while maintains a lower reactivity (around 10^{10}) of phosphate monoesters. In addition, AP and NPP have very similar active site structures (Fig.6.1), e.g., the bi-metallo zinc site and the identical six ligands, the deprotonated Ser/Thr nucleophile, making this pair of enzymes ideal for comprehensive studies to understand the structural and functional correlations.

Extensive experimental and computational studies have been carried out to address these interesting questions. [1, 2, 25–28, 30–33, 58, 179, 182] Experimental studies by Linear Free Energy Relationship (LFER), Kinetic Isotope Effects (KIE) and spectroscopy have gleaned insightful recognitions of the reaction mechanisms and the important structural factors that

contribute to AP/NPP catalysis. Among these understandings, a crucial proposal is that members in AP superfamily catalyze cognate and promiscuous substrates by similar natures of TS to their solution reactions. [25, 27] In other words, AP/NPP catalyzes phosphate mono- and di-ester hydrolysis via a loose and a synchronous TS, respectively, indicating that although the active site of AP family is evolutionarily optimized for stabilizing the TS of one type of substrate, it can in fact recognize different types of transition states and perform noticeable stabilization. From the strategic level, it depicts how evolution shapes enzymes that share the same ancestor into a functionally related enzyme family.

However, these results have been challenged by theoretical studies, [34] criticizing the inability of experimental approaches in exploring TS which is a transient species and the ambiguities on data interpretations. For AP superfamily, Tunon and coworkers have studied the phosphate monoester reactions in AP [179] and diester reactions in AP and NPP [1, 2] by QM/MM simulations. It is quite surprising that their studies display completely different pictures of AP catalysis from experimental views: AP catalyzes monoester hydrolysis via a two-step mechanism with a stable intermediate state, thus fundamentally different from the one-step mechanism in solution; for phosphate diester reactions in AP and NPP, the similar loose TSs are fostered, also different from the synchronous TS in solution. In other words, AP/NPP changes the nature of TS of phosphate mono- and di-esters hydrolysis. It is worth pointing out that several important structural features in the active site change drastically in those simulations, e.g., the zinc-zinc distance increases from 4 Å in crystal structures to up to 7 Å! Nevertheless, the significant discrepancy in previous studies highlights the important and controversial feature of this problem and the necessity of further studies.

In our recent work, [58] we carried out systematic theoretical studies of phosphate diester hydrolysis reactions in AP and NPP based on the enzyme models constructed from crystal structures. The calculation is under QM/MM scheme [53] that takes the enzyme matrix effects at modest cost and a semi-empirical QM method that is specifically parametrized for phosphate reactions. [46] By careful kinetics and structural benchmarks with respect to available experimental data, we established the semi-quantitative nature of our methods. Our

studies of a series phosphate diester reactions suggest that neither AP nor NPP significantly change the nature of TS for diester reactions; instead, they employ similar synchronous mechanisms to solution reactions, consistent with previous experimental results. The possible reason is that AP/NPP lacks enough driving force to significantly shift the nature of TS for phosphate diesters, therefore a more "economical" way is to utilize the similar TS as in solution. Compared to previous theoretical studies, our models are able to produce more systematic and consistent energetic and structural data with experimental results.

In this work, we explore the enzyme reactions of the other category of substrates, phosphate monoesters, to obtain a complete view of AP catalysis. The corresponding solution reactions have been studied in our previous work by an implicit solvent model [52] and QM/MM scheme for which good agreement with experimental data and high level QM calculations are reached. Due to the large charge redistribution in phosphate monoester hydrolysis, a novel state-dependent QM/MM interaction scheme (Klopman-Ohno scheme) with significantly improved accuracy is used. We study a particular phosphate monoester, $pNPP^{2-}$ (Fig.6.1(c)) and obtain good agreement with experimental data. Our results suggest that similar loose TSs are adopted in AP and NPP for monoester hydrolysis, qualitatively different from diester reactions. Hence, these results, together with our previous studies, render us the complete view of AP superfamily catalysis and support the previous experimental proposal.

The paper is organized as follows: in Sect.6.2 we summarize computational methods and simulation setup. In Sect.6.3, we first briefly review the reference solution reactions, and then analysis of the phosphoryl transfer TS for phosphate monoesters in AP and NPP. Before concluding in Sect.6.4, we summarize our results for AP catalysis and discuss the controversies and possible reasons.

6.2 Computational Methods

6.2.1 Enzyme Model Setup

The enzyme models used in this work are similar to those in previous studies. [58] Therefore, we only summarize some key features briefly. We investigate the first step of the hydrolysis reaction of $pNPP^{2-}$ in an *E. coli* AP variant in which Arg166 is mutated to Ser and wild type NPP (Fig. 6.1). It worth mentioning that the chemical steps are fully rate-limiting in these enzymes.

The enzyme models are constructed based on the X-ray structures for the *E. coli* AP mutant R166S with bound inorganic phosphate at 2.05 Å resolution (PDB code 3CMR [182]) and *Xac* NPP with bound Adenosine Mono-Phosphate (AMP) at 2.00 Å resolution (PDB code 2GSU [27]). In each case, starting from the PDB structure, the ligand is first "mutated" to $pNPP^{2-}$. Hydrogen atoms are added by the HBUILD module [188] in CHARMM. [189] All basic and acidic amino acids are kept in their physiological protonation states except for Ser102 and Thr90 in AP and NPP, respectively, which are assumed to be the neucleophiles and deprotonated in the reactive complex. Water molecules are added following the standard protocol of superimposing the system with a water droplet of 27 Å radius centered at $Zn1^{2+}$ (see Fig.6.1 for atomic labels) and removing water molecules within 2.8 Å from any atoms resolved in the crystal structure. [161] Protein atoms in the MM region are described by the all-atom CHARMM force field for proteins [190] and water molecules are described with the TIP3P model. [162] The QM region includes groups most relevant to the reaction: the two zinc ions and their 6 ligands (Asp51, Asp369, His370, Asp327, His412, His331), Ser102 and $MpNPP^{-}$ for R166S AP; for NPP, this includes two zinc ions and their 6 ligands (Asp54, Asp257, His258, Asp210, His363, His214), Thr90 and $MpNPP^{-}$. Only side chains of protein residues are included in the QM region and link atoms are added between C_α and C_β atoms. The treatment of the QM/MM frontier follows the DIV scheme in CHARMM. [191] A NOE potential is added to the C-O bonds in Asp51, which is coordinated to both Mg^{2+} and Zn^{2+} to avoid over polarization.

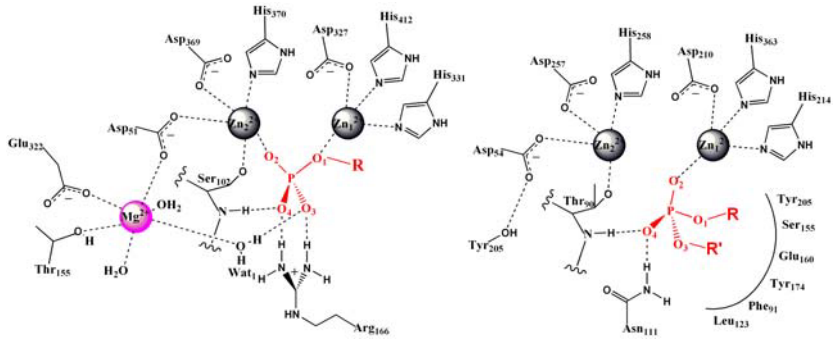


Figure 6.1: The active sites of Alkaline Phosphatase (AP) and Nucleotide PyrophosPhatase/ phosphodiesterase (NPP) are generally similar, with a few distinct differences. (a) *E. coli* AP active site. (b) *Xac* NPP active site. The cognate substrates for AP and NPP are phosphate monoesters and diesters, respectively. (c) The phosphate monoester ($pNPP^{2-}$) studied in this work.

Due to the fairly large size of the QM region (more than 80 atoms) and extensive sampling required for the open active site of AP and NPP, the SCC-DFTBPR method [46] is used for PMF calculations. Extensive benchmark calculations and applications indicate that it is comparable to the best semi-empirical method available in the literature for phosphate chemistry. [180, 192]

The generalized solvent boundary potential (GSBP) [124, 163] is used to treat long range electrostatic interactions in geometry optimizations and MD simulations. The system is partitioned into a 27-Å spherical inner region centered at the Zn1 atom, with the rest in the outer region. Newtonian equations-of-motion are solved for the MD region (within 23 Å), and Langevin equations-of-motion are solved for the buffer regions (23-27 Å) with a temperature bath of 300 K; protein atoms in the buffer region are harmonically constrained with force constants determined from the crystallographic B-factors. [193] All bonds involving hydrogen are constrained using the SHAKE algorithm, [166] and the time step is set to 1 fs. All water molecules in the inner region are subject to a weak GEO type of restraining potential to keep them inside the inner sphere with the MMFP module of CHARMM. The static field due to outer-region atoms, ϕ_s^{io}, is evaluated with the linear Poisson-Boltzmann (PB) equation using a focusing scheme with a coarse cubic grid of 1.2 Å spacing, and a fine grid of 0.4 Å spacing. The reaction field matrix $\mathbf{M}$ is evaluated using 400 spherical harmonics. In the PB calculations, the protein dielectric constant of $\epsilon_p = 1$, the water dielectric constant of $\epsilon_w = 80$, and 0.0 M salt concentration are used; the value of ϵ_p is not expected to make a large difference in this particular case because the active site is already very solvent accessible and the inner/outer boundary is far from the site of interest. The optimized radii of Nina et al. [194, 195] based on experimental solvation free energies of small molecules as well as the calculated interaction energy with explicit water molecules are adopted to define the solvent-solute dielectric boundary. To be consistent with the GSBP protocol, the extended electrostatic model [164] is used to treat the electrostatic interactions among inner region atoms in which interactions beyond 12 Å are treated with multipolar expansions, including the dipolar and quadrupolar terms.

6.2.2 Benchmark enzyme calculations based on minimizations and reaction path calculations

To test the applicability of SCC-DFTBPR/MM to AP and NPP, geometry optimization for the reactant (Michaelis) complex is compared to results from B3LYP [196–198]/MM calculations. The basis set used in the B3LYP/MM calculations is 6-31G* [199], and the calculations are carried out with the QChem [200] program interfaced with CHARMM (c36a2 version). [201] Due to the rather large size of the QM region and the high cost of *ab initio* QM/MM calculations, atoms beyond 7 Å away from Zn1 are fixed to their crystal positions in these minimizations. The convergence criteria for geometry optimization are that the root-mean-square (RMS) force on mobile atoms is smaller than $0.30 kcal/(mol \cdot \mathring{A})$ and the maximum force smaller than $0.45 kcal/(mol \cdot \mathring{A})$.

6.2.3 State-dependent QM/MM interaction scheme and 1D Potential of mean force (PMF) simulations

Due to the large amount of charge redistribution in phosphate monoester hydrolysis and the relative open active site of AP and NPP, conventional QM/MM interaction scheme can result in quite large errors, as demonstrated in our previous studies. [231] Hence, a state-dependent QM/MM interaction scheme (Klopman-Ohno scheme) has been developed by modifying the electrostatic interactions,

$$H_{elec,KO}^{QM/MM} = \sum_{\alpha I} \frac{\Delta q_\alpha Q_I}{\sqrt{R_{\alpha I}^2 + a_\alpha (\frac{1}{U_\alpha(\Delta q_\alpha)} + \frac{1}{U_I})^2 e^{-b_\alpha R_{\alpha I}}}} \tag{6.1}$$

where $U_\alpha(\Delta q_\alpha)$ takes a linear relationship with atomic Mulliken charge via $U_\alpha(\Delta q_\alpha) = U_\alpha^0 + \Delta q_\alpha U_\alpha^d$ and U_α^d is Hubbard derivative. The conventional 6-12 potential for vdW interactions is untouched. With systematic reparametrization, aiming at condense phase chemical reactions, the Klopman-Ohno (KO) scheme can result in large improvement for QM/MM interactions for highly charge systems and have been successfully used to study aqueous reactions for phosphate monoesters.

To study the free energy profile of enzyme reactions, PMF simulations have been carried out for R166S AP and NPP with $pNPP^{2-}$ as the substrates. After the initial minimizations starting from the relevant crystal structure, the enzyme system is slowly heated to 300 K and equilibrated for 100 ps. The reaction coordinate is defined as PO^{lg}-PO^{nu}. The umbrella sampling approach [167] is used to constrain the system along the reaction coordinate by using a force constant of 150 kcal/mol$\cdot\mathring{A}^{-2}$. In total, more than 51 windows are used for each PMF and 100 ps simulations are performed for each window. The first 50 ps trajectories are discarded and only the last 50 ps are used for data analysis. Convergence of the PMF is monitored by examining the overlap of reaction coordinate distributions sampled in different windows and by evaluating the effect of leaving out segments of trajectories. The probability distributions are combined together by the weighted histogram analysis method (WHAM) [168] to obtain the PMF along the reaction coordinate. The averaged key structural properties for each window are calculated and summarized in Table 6.2.

6.2.4 M06/MM free energy perturbation corrections

As indicated in our benchmarks, SCC-DFTBPR/MM underestimates the reaction barriers of $pNPP^{2-}$ in enzymes, therefore it is necessary to include high level QM method corrections. In our previous work, M06 functional with 6-31+G** basis set gives the best balance between accuracy and computational cost. The correction is done on the basis of a straightforward one-step free energy perturbation calculation

$$\Delta G_{M06-SCC} = -kTln < e^{-\beta(U_{M06/MM}-U_{SCC/MM})} >_{SCC/MM} \tag{6.2}$$

at both end states ($\lambda = 0.0$ or 1.0). The difference between the perturbative correction at the two end states gives the M06/MM correction to the reaction free energy. Since only a small number of snapshots from SCC-DFTB/MM trajectories are used, a second-order cumulant expansion is used to improve the numerical stability of the perturbation calculation

$$\Delta G_{M06-SCC} =< U_{M06/MM} - U_{SCC/MM} >_{SCC/MM} -$$
$$\frac{\beta}{2}[< (U_{M06/MM} - U_{SCC/MM})^2 >_{SCC/MM} - < U_{M06/MM} - U_{SCC/MM} >^2] \quad (6.3)$$

As discussed extensively in the literature, [228] such one-step perturbation is effective only if the configuration space at the two levels overlaps significantly; this is assumed to be the case considering the previous observation [105, 229] that SCC-DFTB often gives reliable geometries and energetics compared to DFT.

6.3 Results and Discussion

6.3.1 First step of $pNPP^{2-}$ hydrolysis in R166S AP

The wt AP catalyzes $pNPP^{2-}$ hydrolysis so efficient that the chemical step is no longer the rate-limiting step. Therefore, a mutant is typically used in experiment in which the Arg166 is mutated to a Serine group to study the enzyme catalysis. This mutation is believed not affecting the catalysis mechanisms. [30] $pNPP^{2-}$ is a phosphate monoester that has been widely studied in solution and AP/NPP with abundant experimental data available. In addition, a similar phosphate diester, $MpNPP^-$, has been systematically studied in our previous work, hence making $pNPP^{2-}$ the perfect choice for phosphate monoester studies.

$pNPP^{2-}$ is the cognate substrate of AP. The experimental measured reaction barrier including the binding process (k_{cat}/K_m) equals to 12.1 kcal/mol (Table 6.1) calculated by transition state theory at 300K. The reaction barrier for the chemical step (k_{cat}) has also been measured as 18.0 kcal/mol. [182] Therefore, the estimation of binding free energy is 5.9 kcal/mol. Compared with the similar diester $MpNPP^-$ reaction which has measured as 18.0 kcal/mol, R166S AP favors $pNPP^{2-}$ by 5.9 kcal/mol. Since AP active site features several extra positive charged motifs, e.g., the magnesium site, it is likely that the binding free energies of phosphate monoesters are larger than those of diesters. Therefore, the energy difference in actual chemical steps should be less than 5.9 kcal/mol.

The comparison of optimized structures by B3LYP/MM and SCC-DFTBPR/MM shows good agreement between the two levels (Fig.6.2). The O_{Ser102}-P distances in the optimized structures are 3.3 (3.2) Å in B3LYP/MM (SCC-DFTBPR/MM), very close to the 3.1 Å in crystal structure, leading to a stable reactant complex. The O2 of the substrate coordinates to one of the zinc ions and O1 with the phenyl group is solvated by water molecules. O4 and the nearby Ser102 backbone amide forms a hydrogen bond. It is very interesting to see that in the B3LYP/MM optimized structure, the Wat1 forms a much stronger hydrogen bond with O3 of $pNPP^{2-}$ than Ser102. It is the opposite situation for a phosphate diester, $MpNPP^{-}$, in our previous studies, in which Wat1 only forms a hydrogen bond with Ser102 in the reactant state. This change reflects the difference of substrate charge states: the phosphate monoester is more negatively charged, therefore the hydrogen bond with Wat1 is favored due to the stronger electrostatic interactions. The results of phosphate monoester are also closer to the crystal structure in which a phosphate (PO_4^{3-}) is used as the inhibitor(see Fig.6.2c for the comparison). However, this change of hydrogen bond interactions is not captured by SCC-DFTBPR/MM with the KO scheme. Many other hydrogen-bonding distances and distances involving the zinc ions are similar at the two levels of theory. The Zn^{2+}-Zn^{2+} distance is generally larger at the SCC-DFTBPR/MM level. Overall, the agreement between optimized structures at the two levels of theory is excellent, supporting the use of SCC-DFTBPR/MM with the KO scheme. The minimum energy path (MEP) results from adiabatic mapping indicate that the SCC-DFTBPR/MM underestimate the reaction barrier compared with B3LPY/MM (12.2 vs 6.2 kcal/mol). As shown in Figure 6.2, the main differences of the TSs with SCC-DFTBPR/MM compared with B3LYP/MM includes a tighter P-O^{lg}, the weaker interactions between oxygen and zinc and a weaker hydrogen bond with Wat1.

We calculate the PMF with respect to the anti-symmetric stretch of PO^{lg} and PO^{nu} bonds which has shown to be a good reaction coordinate (RC) in our previous studies. The PMF profile (Fig. 6.3) indicates a single step exothermic reaction with the barrier peaking at RC less than 0 Å, therefore fundamentally different from the two-step mechanism in previous theoretical studies. The reaction barrier height is 13.5 kcal/mol, lower than the

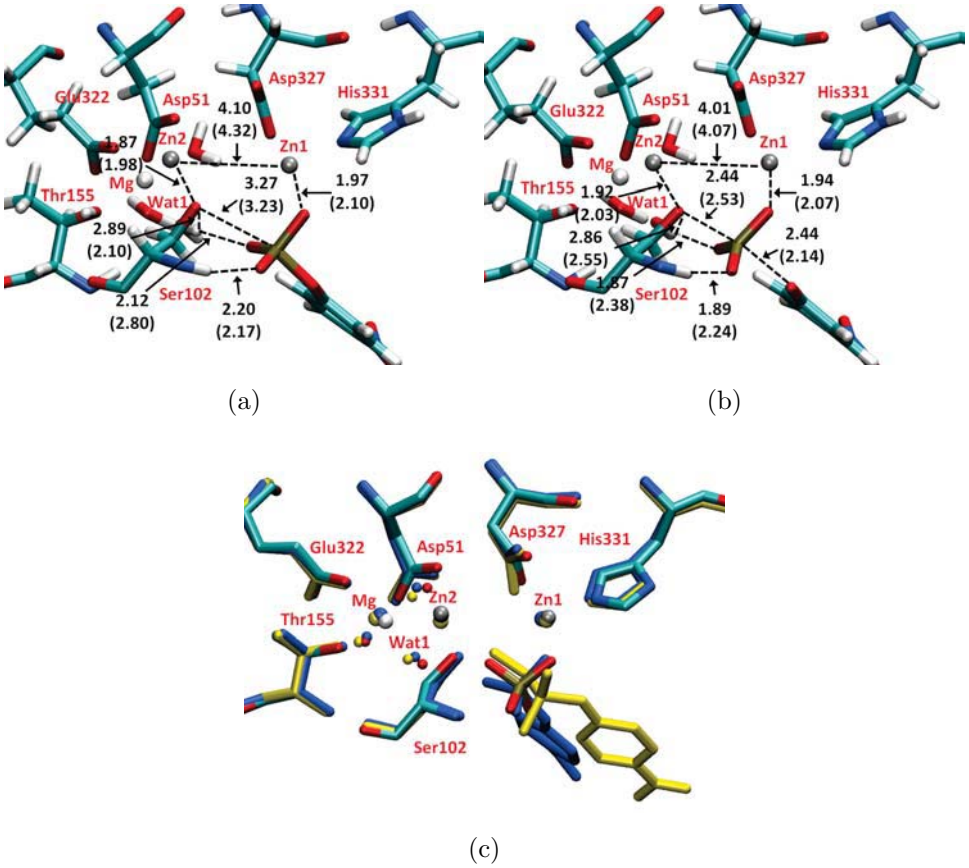


Figure 6.2: Benchmark calculations for $pNPP^{2-}$ in R166S AP. Key distances are labeled in Å. Numbers without parenthesis are obtained with B3LYP/6-31G*/MM optimization; those with parentheses are obtained by SCC-DFTBPR/MM optimization with KO scheme. Asp369, His370, and His412 are omitted for clarity. (a) The reactant state in R166S AP; (b) The transition state in R166S AP by adiabatic mapping; (c) The overlay of crystal structure with PO_4^{3-} (colorful), B3LYP/6-31G*/MM optimized structures with $pNPP^{2-}$ (blue) and $MpNPP^{-}$ (yellow). Hydrogen atoms are omitted.

Table 6.1: $pNPP^{2-}$ hydrolysis reactions in solution, R166S AP and NPP from experiments and calculations

	Exp[a]	SCC/MM	M06/MM[b]
Solution	31.8	32	
R166S AP	12.1 (18.0[c])	13.5 (12.2/6.2)[d]	20.2
NPP	17.5	14.0 (12.4/8.5)[d]	

a. Free energy barriers (kcal/mol) calculated by transition state theory at 300 K based on experimental k_{cat}/K_M values; b. PMF results after M06/MM FEP corrections; c. free energy barriers (kcal/mol) calculated by transition state theory at 300 K based on experimental k_{cat} values; d. adiabatic mapping barriers with B3LPY/MM/6-31G* and SCC-DFTBPR/MM with KO scheme.

experimental estimation by 4.5 kcal/mol, consistent with the MEP benchmark results that SCC-DFTBPR/MM tends to underestimate the reaction barrier. Therefore, we explore the M06/MM corrections by a one-step FEP that has been successfully applied in our previous work to improve the quantitative agreement with experimental data for phosphate diesters. Indeed, the reaction barrier after the M06/MM correction is 20.2 kcal/mol (Table 6.1), much closer to the experimental results. Compared with the 24.4 kcal/mol barrier for the calculated chemical step of $MpNPP^-$, the monoester reaction is favored by 4.2 kcal/mol. Overall, these results are qualitatively consistent with the fact that AP favors $pNPP^{2-}$ over $MpNPP^-$.

Several important structural properties are plotted with the RC (Fig.6.3). The bond lengths of P-O^{lg} and P-O^{nu} change smoothly and intersect at RC around 0 Å. The Zn-Zn distance fluctuates around 4 Å, close to the value in crystal structures. The TS locates at RC equals to -0.4 Å (Table 6.2), more negative than $MpNPP^-$, with averaged P-O^{lg} and P-O^{nu} bond lengths as 2.04 and 2.46 Å, respectively. Compared with the TS of $MpNPP^-$, both bonds are elongated and the Tightness Coordinate (TC) also increases from 3.89 to

Table 6.2: Key structural properties for the TS of the first step of phosphate monoester and diester hydrolysis in solution, AP and NPP

	Substrate	RC[a]	TC[b]	P-O^{lg}	P-O^{nu}	Zn^{2+}-Zn^{2+}
Solution	$pNPP^{2-}$	-0.31	4.21	1.95	2.26	
	$MpNPP^{-}$	-0.20	4.66	2.23	2.43	
R166S AP	$pNPP^{2-}$	-0.41±0.07	4.50±0.19	2.04±0.11	2.46±0.10	4.10±0.21
	$MpNPP^{-}$[c]	-0.11±0.07	3.89±0.14	1.89±0.07	2.00±0.09	3.89±0.14
NPP	$pNPP^{2-}$	-0.41±0.07	4.63±0.23	2.11±0.13	2.52±0.11	4.11±0.21
	$MpNPP^{-}$	-0.20±0.07	3.86±0.14	1.83±0.06	2.03±0.09	3.92±0.17

a. The Reaction coordinate (RC) is defined as the difference between P-O^{lg} and P-O^{nu}; b. The Tightness coordinate (TC) is defined as the sum of P-O^{lg} and P-O^{nu}; c. The two substrate orientations result in very similar structural properties, therefore only one is shown.

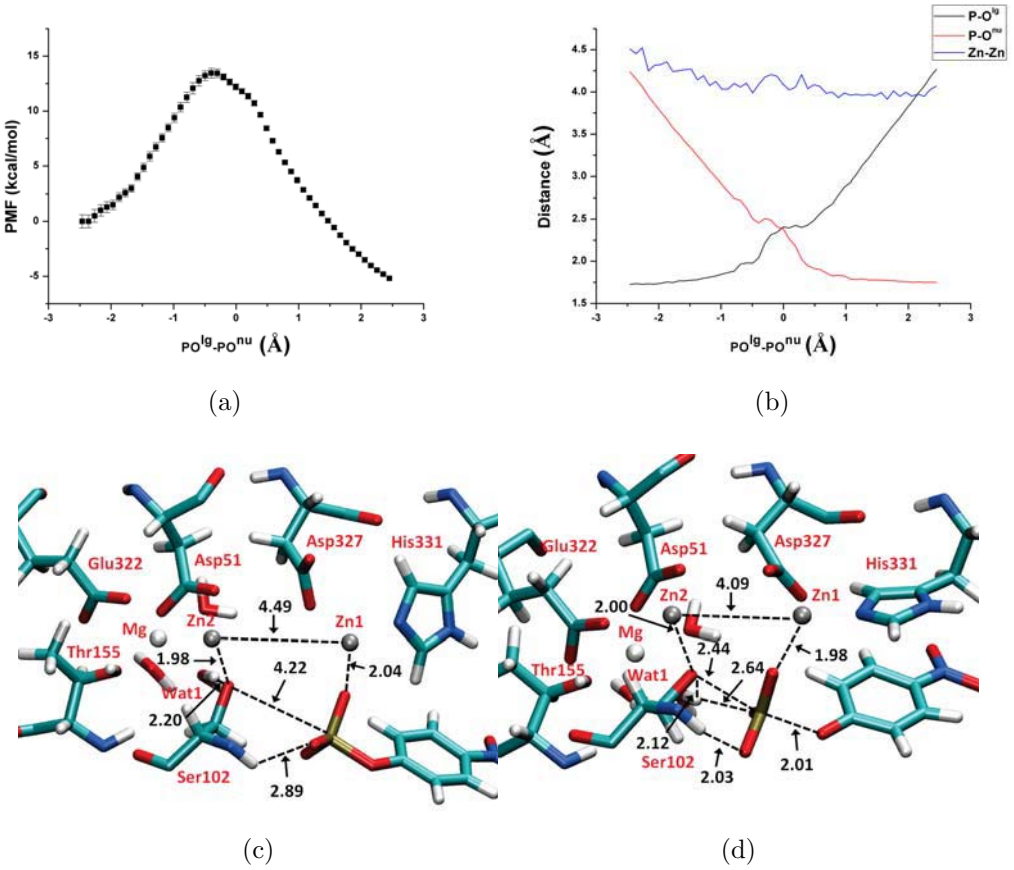


Figure 6.3: Potential of Mean Force (PMF) calculation results for $pNPP^{2-}$ hydrolysis in R166S AP. Key distances are labeled in Å and energies are in kcal/mol. (a) PMF along the reaction coordinate with error bar included; (b) Changes of average key distances along the reaction coordinate; (c) A snapshot for the reactant state, with average key distances labeled; (d) A snapshot for the TS, with average key distances labeled. Asp369, His370, and His412 are omitted for clarity.

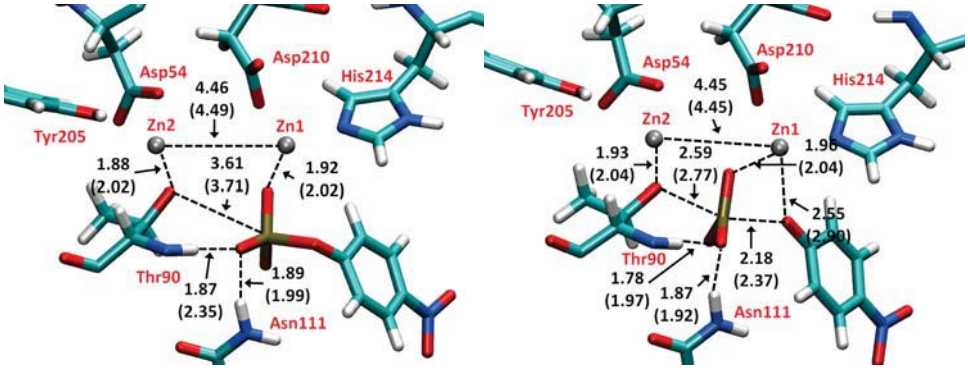
4.50 Å. Therefore, $pNPP^{2-}$ hydrolysis goes through a loose TS, clearly different from diester reactions.

In the reactant state (Fig.6.3(c)), the substrate binds with Zn1 via a nonbridging oxygen and forms a hydrogen bond with a backbone amide. Different from the experimental expectation, [28] Wat1 forms a hydrogen bond with the deprotonated Ser102, instead of the substrate, similar to our observation for $MpNPP^{-}$, probably due to the increased $PO^{n}u$ distance in the reactant state than the crystal structure. The Zn-Zn distance is slightly increased to 4.49 Å, about the largest value we observed in our calculations. Later, in the TS (Fig.6.3(d)) the Ser102 goes to attack the substrate while Wat1 partially breaks the hydrogen bond with Ser102 and forms a new hydrogen bond with a $pNPP^{2-}$ nonbridging oxygen, which has been also observed in $MpNPP^{-}$ reactions and proposed to help lower the reaction barrier by providing extra stabilization of TS.

A very interesting fact for the reaction process is that the leaving group oxygen does not directly interact with Zn1 but solvated by water instead, which is similar to our previous observations for $MpNPP^{-}$ but at odds with the crystal structure of a vanadate TS analog. To clarify this point, we carry out one calculation with the initial structure prepared so that the leaving group oxygen is constrained to bind with Zn1 and later remove the constraint in the PMF calculations. The results are very similar to the original simulation starting from the unconstrained structure and the leaving group oxygen quickly becomes solvated by water after the removal of constraint. By these comparisons, we believe this observation is not subject to the bias of the simulation. Actually, if we compare the TC of TS in the simulation (4.50 Å) and the average zinc-zinc distance (4.10 Å), the bi-metallo zinc motif cannot completely accommodate the TS due to the geometric constraint. Alternatively, the vanadate has a TC of 3.64 Å in the crystal structure that can be perfectly fit into the zinc site. Therefore, the vanadate may not be a good choice for phosphate monoester TS analog.

6.3.2 First step of $pNPP^{2-}$ hydrolysis in NPP

Different from AP, NPP catalyzes $pNPP^{2-}$ hydrolysis promiscuously with lower proficiency than $MpNPP^{-}$. The measured reaction barrier including binding process (k_{cat}/K_m) equals to 17.5 kcal/mol, slight higher than the 14.3 kcal/mol barrier of $MpNPP^{-}$. There is no available data for the chemical step (k_{cat}).



(a) (b)

Figure 6.4: Benchmark calculations for $pNPP^{2-}$ in NPP. Key distances are labeled in Å. Numbers without parenthesis are obtained with B3LYP/6-31G*/MM optimization; those with parentheses are obtained by SCC-DFTBPR/MM optimization with KO scheme. (a) The reactant state in NPP; (b) The transition state in NPP by adiabatic mapping. Asp257, His258, and His363 are omitted for clarity.

Similar to the comparisons made above for AP, SCC-DFTBPR/MM minimizations for $pNPP^{2-}$ in NPP also give similar reactant complex structure to B3LYP/MM calculations (Fig. 6.4a). The O_{Thr90}-P distance increases from 3.2 Å in crystal, which contains AMP as the inhibitor, to 3.6 (3.7) Å at the B3LYP/MM (SCC-DFTBPR/MM) level. The substrate O2 coordinates with Zn1, while O4 forms hydrogen bonds with Asn111 and the backbone amide of Thr90. The optimized Zn^{2+}-Zn^{2+} distance is 4.46 (4.49) at the B3LYP/MM (SCC-DFTBPR/MM) level. The two hydrogen bonds formed between O4-Asn111 and O4-Thr90-backbone-amide are also in decent agreement at different levels of theory. Similar to the

adiabatic mapping results in AP, the SCC-DFTBPR/MM with KO scheme tends to underestimate the reaction barrier compared to B3LYP/MM (8.5 vs. 12.4 kcal/mol). However, the transition state geometries are quite consistent at the two levels of theory.

The calculated PMF (Fig.6.5) also indicates an exothermic reaction maximizing at RC slightly less than 0 Å. The reaction barrier height is 14.0 kcal/mol, lower than the experimental value. Together with our AP results, these discrepancies indicate some systematic errors in our calculation methods that may require further improvement. Similar to AP catalysis, the TS is at RC equals to -0.4 Å, with a TC of 4.63 Å (Table 6.2) which is much looser than the $MpNPP^-$ reaction in NPP (3.86 Å). Therefore, these observations indicate that NPP also catalyzes phosphate mono- and di-esters via different mechanisms. In the reactant state (Fig.6.5(c)), the substrate binds with Zn1 via a nonbridging oxygen and forms two hydrogen bonds with a backbone amide and Asn111. The zinc-zinc distance is also slight elongated to 4.52 Å. The deprotonated Thr90 serves as the nucleophile and attacks the substrate via a loose TS (Fig.6.5(d)). Similar to in AP, the leaving group oxygen does not interaction with Zn1, but solvated by water instead.

6.3.3 Comparisons of AP superfamily catalysis for phosphate mono- and di-esters

Together with our previous studies of phosphate diester reactions, we obtain a complete view of the strategy that AP and NPP employ for phosphate hydrolysis. Our calculation results show that although AP and NPP feature different specificity and promiscuity, they catalyze the same type of substrates via similar mechanisms: although AP is evolved for phosphate monoester reactions via a loose TS, it can recognize and catalyze the synchronous TS of phosphate diesters; similarly, the active site of NPP is evolutionarily shaped for the synchronous TS of phosphate diesters, but it can also accommodate the loose TS of phosphate monoesters and catalyze it as well. These results are consistent with experimental findings and different from previous theoretical studies which claimed that AP and NPP loosen the TS of phosphate diesters.

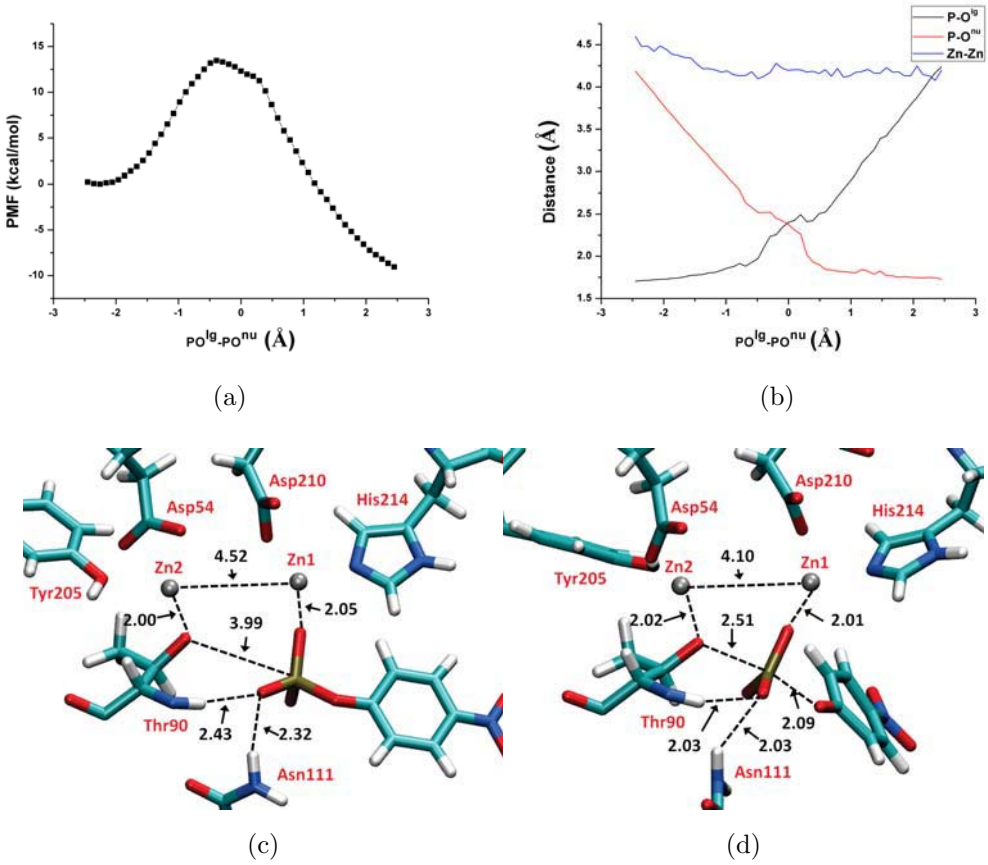


Figure 6.5: Potential of Mean Force (PMF) calculation results for $pNPP^{2-}$ hydrolysis in NPP. Key distances are labeled in Å and energies are in kcal/mol. (a) PMF along the reaction coordinate; (b) Changes of average key distances along the reaction coordinate; (c) A snapshot for the reactant state, with average key distances labeled; (d) A snapshot for the TS, with average key distances labeled. Asp257, His258, and His363 are omitted for clarity.

For the solution reactions which serve as the reference for enzyme catalysis, it is interesting that the TS of phosphate monoester is not necessarily looser than phosphate diester; on contrary, it is actually tighter for the pair ($pNPP^{2-}$ vs. $MpNPP^{-}$) that we studied (3.94 vs. 4.66 Å). Considering previous theoretical work, although there are quantitative differences on the TC of the TSs, the fact that phosphate diester hydrolysis is not tighter than monoester is consistent. The reason might be due to the difference in nucleophiles: for phosphate monoester, it is typically water while for diester it is usually hydroxide. For monoester reactions, before the nucleophilic attacking, the water actually transfers one proton to the phosphate monoester which effectively becomes a diester-like substrate. Hence, it may not be very meaningful to compare the solution reactions directly due to the difference in the nucleophiles.

6.4 Concluding remarks

In this work, we studied the hydrolysis of $pNPP^{2-}$ in R166S AP and wild type NPP using SCC-DFTBPR/MM simulations and a state-dependent QM/MM interaction scheme. Together with our previous studies of phosphate monoester reactions in solution and diester reactions in solution and enzymes, it provides us the first complete view from theoretical perspective of AP superfamily catalysis.

Our calculated reaction barriers for the chemical steps are qualitatively consistent with experimental results. The direct comparison of TSs for AP and NPP reactions show that the similar loose TSs are employed in both enzymes, although phosphate monoester is the cognate substrate of AP but promiscuous substrate of NPP. The loose TS is clearly different from the more synchronous TS of diester reactions in solution and enzyme. Therefore, our results support the proposal that AP superfamily are able to recognize different TSs and catalyze them via similar mechanisms to solution reactions, hence consistent with the conclusion from previous experimental studies. Our monoester results are fundamentally different from the two-step mechanism in a previous theoretical work for an alkyl phosphate

monoester in AP. [179] Actually, the two-step mechanism is not the typical mechanism for phosphate monoester aqueous reactions, contrary to the claim from the authors.

For phosphate monoester enzyme reaction, previous crystal structure of a TS analog, vanadate, suggests that the leaving group oxygen directly interact with one zinc ion. In our previous diester studies, we did not observe this direct interaction and the reason is due to the difference of atomic charge of the diester and vanadate: the diester only bears -1 charge while vanadate has -3. Therefore the leaving group oxygen is significantly less charged compared with the vanadate in enzyme active site, suggesting that vanadate is not a good analog for phosphate diesters. In this study, the phosphate monoester $pNPP^{2-}$ bears -2 charge, therefore more similar to vanadate for chemical properties. However, the TC in the TS is more than 4.5 Å, much larger than that for vanadate and the zinc-zinc distance in AP/NPP. So it is impossible to the bi-metallo site to completely accommodate the TS. These results suggest that vanadate is neither a good analog for phosphate monoester.

Chapter 7

Concluding Remarks

The long-term goal of our research is to develop state-of-the-art computational approaches of studying the catalysis mechanisms for phosphoryl transfer reactions, which arguably represent the most important chemical transformation in biology. Together with experimental techniques, the computational studies target at understanding the strategies that the biological systems adopt to catalyze the reactions with high substrate specificity and promiscuity and providing useful guidance of modifying or developing enzyme functions in engineering field.

In Chapter 2, an implicit solvent model for approximate density functional theory, SCC-DFTB, has been developed, motivated by the need to rapidly explore the potential energy surface of aqueous chemical reactions that involve highly charged species, which are the typical references for enzyme catalysis. The solvation free energy is calculated using a popular model that employs Poisson-Boltzmann for electrostatics and a surface-area term for nonpolar contributions. To balance the treatment of species with different charge distributions, we make the atomic radii that define that dielectric boundary and solute cavity depend on the solute charge distribution. Specifically, the atomic radii are assumed to be linearly dependent on the Mulliken charges and solved self-consistently together with the solute electronic structure. Benchmark calculations indicate that the model leads to solvation free energies of comparable accuracy to the SM6 model (especially for ions), which requires much more expensive DFT calculations. With analytical first derivatives and favorable computational speed, the SCC-DFTB based solvation model can be effectively used, in conjunction with

high-level QM calculations, to explore the mechanism of solution reactions. This is illustrated with a brief analysis of the hydrolysis of monomethyl monophosphate ester and trimethyl monophosphate ester.

In Chapter 3, we develop a novel QM/MM interaction scheme by employing a modified Klopman-Ohno functional in electrostatic interactions and a set of element type dependent vdW parameters for condense phase chemical reactions. Extensive benchmarks of solute-solvent interactions for amino acid and phosphate hydrolysis transition state analogs demonstrate the improvement in accuracy for highly charged species and a good parameter transferability. Equipped with this method, the hydrolysis reactions of two phosphate monoesters, MMP^{2-} and $pNPP^{2-}$, are studied and significant improvements of reaction energetics are obtained compared with conventional QM/MM interactions and previous experimental and computational results. These aqueous reaction studies indicate that the nature of transition states of phosphate monoesters is not necessarily looser than that of diesters in solution, since different nucleophiles are involved in reactions. Therefore the previous experimental view of the aqueous reactions may overlook the intrinsic complexity of this problem and result in oversimplified picture.

In Chapter 4, we study the hydrolysis of a phosphate diester, $MpNPP^{-}$, in solution, two experimentally well-characterized variants of AP (R166S AP, R166S/E322Y AP) and wild type NPP by QM/MM calculations and SCC-DFTB method. The general agreements found between these calculations and available experimental data for both solution and enzymes support the use of SCC-DFTB/MM for a semiquantitative analysis of the catalytic mechanism and nature of transition state in AP and NPP. Although phosphate diesters are cognate substrates for NPP but promiscuous substrates for AP, the calculations suggest that their hydrolysis reactions catalyzed by AP and NPP feature similar synchronous transition states that are slightly tighter in nature compared to those in solution, due in part to the geometry of the bimetallic zinc motif. Therefore, this study provides the first directly computational support to the hypothesis that enzymes in the AP superfamily catalyze cognate and promiscuous substrates via similar transition states to those in solution. Our calculations

for different phosphate diester orientations and phosphorothioate diesters highlight that the interpretation of thio-substitution experiments is not always straightforward.

In Chapter 5, we study two more aryl phosphate diesters, MmNPP$^-$ and MPP$^-$, hydrolysis reactions in R166S AP and NPP by SCC-DFTB method and the QM/MM framework. Together with our previous work of MpNPP$^-$, this work composes the computational efforts of exploring the experimental LFER of phosphate diester reactions in AP and NPP. With our enzyme model, we are able to qualitatively reproduce the trend of reaction energetics in AP and NPP for the series of phosphate diesters. By including high level DFT corrections via a one-step free energy perturbation approach for the intrinsic errors in SCC-DFTB, the overestimation of the substrate substitution effects can be partially reduced, resulting in further improvement of the computational accuracy.

In Chapter 6, we study a phosphate monoester, pNPP^{2-}, hydrolysis in R166S AP and NPP with the Klopman-Ohno scheme developed in Chapter 3. By including a similar correction scheme to Chapter 5 via a one-step free energy perturbation and the M06 density functional, the calculated reaction kinetics qualitatively agrees with experimental observations and is consistent with previous results for phosphate diesters. Our studies indicate that AP and NPP employ similar loose TS for phosphate monoester reactions, fundamentally different from the two-step mechanism proposed from a previous theoretical work and clearly distinct from the more synchronous TS for phosphate diester hydrolysis. Therefore, our results support the hypothesis that AP and NPP can recognize different nature of TSs and catalyze them via similar mechanisms to corresponding aqueous reactions. In addition, our results suggest that vanadate may not be a good TS analog for phosphate monoesters due to their differences in the tightness coordinates.

Based on the fruitful results in this project, what are the implications for the future? From the computational method developments and applications in this work, it is obvious that a central line of computational studies of biological systems is the balance of computational accuracy and efficiency. In biological system, the environment affects chemical reactions via electrostatic interactions, hydrogen bond interactions, or hydrophobic interactions that

are crucial to finely tuning the reaction mechanisms. Therefore, it is important to use an accurate method to capture these complicated effects and their influence on enzyme catalysis.

The computational overhead is another major concern. For biological systems, the environment has crucial effects for the chemical events. Therefore, the cluster type of model that has achieved remarkable success in other fields has severe limitations due to the neglect of the surroundings. The typically theoretical models, including not only the proteins and substrates, but the solvent and ions as well, range from at least thousands of atoms to millions of atoms. From the time scale, large amount of samplings are imperative to account for the the functional events that take place within from a few picoseconds to a few seconds. Based on these requirements, the much cheaper molecular mechanics method is still among the top choices in theoretical studies of biological systems. Numerous efforts are also paid to improve the accuracy of molecular mechanics, such as the development of polarizable force fields.

Combing the strength of quantum mechanics and molecular mechanics, the QM/MM framework can employ the highly level quantum mechanics on the central part of the system, such as the enzyme active site, while still allows the inclusion of the surroundings at a modest cost. With the emerging of GPU computing which accelerates the conventional calculations by hundreds of times, the QM region in the QM/MM scheme can be significantly increased to thousands of atoms instead of tens of atoms at current stage while still treats the rest via a much cheaper MM method. Therefore, it enables completely new power to allow computational methods handle bigger system with better accuracy and faster speed.

LIST OF REFERENCES

[1] V. López-Canut, M. Roca, J. Bertrán, V. Moliner, and I. Tuñón, "Theoretical study of phosphodiester hydrolysis in nucleotide pyrophosphatase/phosphodiesterase. environmental effects on the reaction mechanism," *J. Am. Chem. Soc.*, vol. 132, no. 20, pp. 6955–6963, 2010.

[2] V. López-Canut, M. Roca, J. Bertrán, V. Moliner, and I. Tuñón, "Promiscuity in alkaline phosphatase superfamily. unreveling evolution through molecular simulations," *J. Am. Chem. Soc.*, vol. 133, pp. 12050–12062, 2011.

[3] M. Bianciotto, J. C. Barthelat, and A. Vigroux, "Reactivity of phosphate monoester monoanions in aqueous solution. 1. quantum mechanical calculations support the existence of "anionic zwitterion" meo(h)po as a key intermediate in the dissociative hydrolysis of the mehtyl phosphate anion," *J. Am. Chem. Soc.*, vol. 124, no. 25, pp. 7573–7587, 2002.

[4] J. Florián and A. Warshel, "Phosphate ester hydrolysis in aqueous solution: Associative versus dissociative mechanisms," *J. Phys. Chem. B*, vol. 102, no. 4, pp. 719–734, 1998.

[5] M. Klhn, E. Rosta, and A. Warshel, "On the mechanism of hydrolysis of phosphate monoester dianions in solutions and proteins," *J. Am. Chem. Soc.*, vol. 128, no. 47, pp. 15310–15323, 2006.

[6] E. Rosta, S. C. L. Kamerlin, and A. Warshel, "On the interpretation of the observed linear free energy relationship in phosphate hydrolysis: A thorough computational study of phosphate diester hydrolysis in solution," *Biochemistry*, vol. 47, no. 12, pp. 3725–3735, 2008.

[7] P. J. O'Brien and D. Herschlag, "Catalytic promiscuity and the evolution of new enzymatic activities," *Chemistry & Biology*, vol. 6, no. 4, pp. R91–R105, 1999.

[8] S. D. Copley, "Enzymes with extra talents: moonlighting functions and catalytic promiscuity," *Curr. Opin. Chem. Biol.*, vol. 7, no. 2, pp. 265–272, 2003.

[9] D. M. Z. Schmidt, E. C. Mundorff, M. Dojka, E. Bermudez, J. E. Ness, S. Govindarajan, P. C. Babbitt, J. Minshull, and J. A. Gerlt, "Evolutionary potential of (beta/alpha)(8)-barrels: Functional promiscuity produced by single substitutions in the enolase superfamily," *Biochem.*, vol. 42, no. 28, pp. 8387–8393, 2003.

[10] J. G. Zalatan and D. Herschlag, "The far reaches of enzymology," *Nat. Chem. Biol.*, vol. 5, no. 8, pp. 516–520, 2009.

[11] S. Jonas and F. Hollfelder, "Mapping catalytic promiscuity in the alkaline phosphatase superfamily," *Pure & Appl. Chem.*, vol. 81, no. 4, pp. 731–742, 2009.

[12] A. Aharoni, L. Gaidukov, O. Khersonsky, S. M. Gould, C. Roodveldt, and D. S. Tawfik, "The 'evolvability' of promiscuous protein functions," *Nat. Genet.*, vol. 37, no. 1, pp. 73–76, 2005.

[13] O. Khersonsky, C. Roodveldt, and D. S. Tawfik, "Enzyme promiscuity: evolutionary and mechanistic aspects," *Curr. Opin. Chem. Biol.*, vol. 10, no. 5, pp. 498–508, 2006.

[14] T. M. Penning and J. M. Jez, "Enzyme redesign," *Chemical Reviews*, vol. 101, no. 10, pp. 3027–3046, 2001.

[15] R. J. Kazlauskas, "Enhancing catalytic promiscuity for biocatalysis," *Current Opinion in Chemical Biology*, vol. 9, no. 2, pp. 195–201, 2005.

[16] M. E. Glasner, J. A. Gerlt, and P. C. Babbitt, "Evolution of enzyme superfamilies," *Current Opinion in Chemical Biology*, vol. 10, no. 5, pp. 492–497, 2006.

[17] K. Hult and P. Berglund, "Enzyme promiscuity: mechanism and applications," *Trends in Biotechnology*, vol. 25, no. 5, pp. 231–238, 2007.

[18] J. A. Gerlt and P. C. Babbitt, "Enzyme (re)design: lessons from natural evolution and computation," *Current Opinion in Chemical Biology*, vol. 13, no. 1, pp. 10–18, 2009.

[19] I. Nobeli, A. D. Favia, and J. M. Thornton, "Protein promiscuity and its implications for biotechnology," *Nature Biotechnology*, vol. 27, no. 2, pp. 157–167, 2009.

[20] M. Galperin, A. Bairoch, and E. Koonin, "A subperfamily of metalloenzymes unifies phosphopentomutase and cofactor-independent phosphoglycerate mutase with alkaline phosphatases and sulfatases," *Prot. Sci.*, vol. 7, pp. 1829–1835, 1998.

[21] M. Galperin and M. Hedrzejas, "Conserved core structure and active site residues in alkaline phosphatase superfamily enzymes," *Proteins: Struct., Funct., and Bioinf.*, vol. 45, pp. 318–324, 2001.

[22] J. Coleman, "Structure and mechanism of alkaline phosphatase," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 21, pp. 441–483, 1992.

[23] J. R. Knowles, "Enzyme catalyzed phosphoryl transfer reactions," *Annu. Rev. Biochem.*, vol. 49, pp. 877–919, 1980.

[24] F. H. Westheimer, "Why nature chose phosphates," *Science*, vol. 235, pp. 1173–1178, 1987.

[25] P. O'Brien and D. Herschlag, "Sulfatase activity of e-coli alkaline phosphatase demonstrates a functional link to arylsulfatases, an evolutionarily related enzyme family," *J. Am. Chem. Soc.*, vol. 120, pp. 12369–12370, 1998.

[26] P. O'Brien and D. Herschlag, "Functional interrelationships in the alkaline phosphatase superfamily: phosphodiesterase activity of escherichia coli alkaline phosphatase," *Biochem.*, vol. 40, pp. 5691–5699, 2001.

[27] J. Zalatan, T. Fenn, A. Brunger, and D. Herschlag, "Structural and functional comparisons of nucleotide pyrophosphatase/phosphodiesterase and alkaline phosphatase: Implicaitons for mechanism and evolution," *Biochem.*, vol. 45, pp. 9788–9803, 2006.

[28] J. Zalatan, A. Fenn, and D. Herschlag, "Comparative enzymology in the alkaline phosphatase superfamily to determine the catalytic role of an active-site metal ion," *J. Mol. Biol.*, vol. 384, pp. 1174–1189, 2008.

[29] F. Hollfelder and D. Herschlag, "The nature of the transition-state for enzyme-catalyzed phosphoryl transfer-hydrolysis of o-aryl phosphorothioates by alkaline-phosphatase," *Biochem.*, vol. 38, pp. 12255–12264, 1995.

[30] P. O'Brien and D. Herschlag, "Does the active site arginine change the nature of the transition state for alkaline phosphatase-catalyzed phosphoryl transfer?," *J. Am. Chem. Soc.*, vol. 121, pp. 11022–11023, 1999.

[31] I. Nikolic-Hughes, D. Rees, and D. Herschlag, "Do electrostatic interactions with positively charged active site groups tighten the transition state for enzymatic phosphoryl transfer?," *J. Am. Chem. Soc.*, vol. 126, pp. 11814–11819, 2004.

[32] J. Zalatan and D. Herschlag, "Alkaline phosphatase mono- and diesterase reactions: Comparative transition state analysis," *J. Am. Chem. Soc.*, vol. 128, pp. 1293–1303, 2006.

[33] J. Zalatan, I. Catrina, R. Mitchell, P. Grzyska, P. O'Brien, and D. Herschlag, "Kinetic isotope effects for alkaline phosphatase reactions: Implications for the role of active-site metal ions in catalysis," *J. Am. Chem. Soc.*, vol. 129, pp. 9789–9798, 2007.

[34] J. Åqvist, K. Kolmodin, J. Florián, and A. Warshel, "Mechanistic alternatives in phosphate monoester hydrolysis: what conclusions can be drawn from available experimental data?," *Chem. Bio.*, vol. 6, no. 3, pp. R71–R80, 1999.

[35] T. Glennon and A. Warshel, "How does gap catalyze the gtpase reaction of ras?: A computer simulation study," *Biochem.*, vol. 39, pp. 9641–9651, 2000.

[36] W. Jencks, *Catalysis in chemistry and enzymology.* New York: Dover publications, 1987.

[37] A. Fersht, *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding.* W.H. Freeman and Company, 1999.

[38] D. Draut, K. Carroll, and D. Herschlag, "Challenges in enzyme mechanism and energetics," *Annu. Rev. Biochem.*, vol. 72, pp. 517–571, 2003.

[39] V. Schramm, "Enzymatic transition states and transition state analog design," *Annu. Rev. Biochem.*, vol. 67, pp. 693–720, 1998.

[40] M. Garcia-Viloca, J. Gao, M. Karplus, and D. Truhlar, "How enzymes work: Analysis by modern rate theory and computer simulations," *Science*, vol. 303, pp. 186–195, 2004.

[41] A. Warshel, P. Sharma, M. Kato, Y. Xiang, H. Liu, and M. Olsson, "Electrostatic basis for enzyme catalysis," *Chem. Rev.*, vol. 106, pp. 3210–3235, 2006.

[42] W. Cleland and A. Hengge, "Enzymatic mechanisms of phosphate and sulfate transfer," *Chem. Rev.*, vol. 106, pp. 3252–3278, 2006.

[43] A. Hengge, "Mechanistic studies on enzyme-catalyzed phosphoryl transfer," *Adv. Phys. Org. Chem.*, vol. 40, pp. 49–108, 2005.

[44] A. Hengge, W. Edens, and H. Elsing, "Transition-state structures for phosphoryl-transfer reactions of p-nitrophenyl phosphate," *J. Am. Chem. Soc.*, vol. 116, pp. 5045–5049, 1994.

[45] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert, "Self-consistent-charge density-funcitonal tight-binding method for simulations of complex materials properties," *Phys. Rev. B*, vol. 58, no. 11, pp. 7260–7268, 1998.

[46] Y. Yang, H. Yu, D. York, M. Elstner, and Q. Cui, "Description of phosphate hydrolysis reactions with the self-consistent-charge density-functional-tight-binding (scc-dftb) theory. 1. parameterization," *J. Chem. Theo. Comp.*, vol. 4, no. 12, pp. 2067–2084, 2008.

[47] Y. Yang, H. Yu, and Q. Cui, "Extensive conformational changes are required to turn on atp hydrolysis in myosin," *J. Mol. Biol.*, vol. 381, pp. 1407–1420, 2008.

[48] Y. Yang and Q. Cui, "The hydrolysis activity of adenosine triphosphate in myosin: A theoretical analysis of anomeric effects and the nature of the transition state," *J. Phys. Chem. A*, vol. 113, no. 45, pp. 12439–12446, 2009.

[49] Y. Yang and Q. Cui, "Does water relayed proton transfer play a role in phosphoryl transfer reactions? a theoretical analysis of uridine 3'-*m*-nitrobenzyl phosphate isomerization in water and *tert*-butanol," *J. Phys. Chem. B*, vol. 113, pp. 4930–4933 NIHMS:103392, 2009.

[50] A. Kirby and A. Varvoglis, "Reactivity of phosphate esters. monoester hydrolysis," *J. Am. Chem. Soc.*, vol. 89, pp. 415–423, 1967.

[51] G. Thatcher and R. Kluger, "Mechanism and catalysis of nucleophilic substitution in phosphate esters," *Adv. Phys. Org. Chem.*, vol. 25, pp. 99–265, 1989.

[52] G. H. Hou, X. Zhu, and Q. Cui, "An implicit solvent model for scc-dftb with charge-dependent radii," *J. Chem. Theo. Comp.*, vol. 6, no. 8, pp. 2303–2314, 2010.

[53] A. Warshel and M. Levitt, "Theoretical studies of enzymic reactions-dielectric, electrostatic and steric stabilization of carbonium-ion in reaction of lysozyme," *J. Mol. Biol.*, vol. 103, pp. 227–249, 1976.

[54] M. J. Field, P. A. Bash, and M. Karplus, "A combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations," *Journal of Computational Chemistry*, vol. 11, no. 6, pp. 700–733, 1990.

[55] Q. Cui, M. Elstner, E. Kaxiras, T. Frauenheim, and M. Karplus, "A qm/mm implementation of the self-consistent charge density functional tight binding (scc-dftb) method," *J. Phys. Chem. B*, vol. 105, no. 2, pp. 569–585, 2001.

[56] M. Freindorf and J. L. Gao, "Optimization of the lennard-jones parameters for a combined ab initio quantum mechanical and molecular mechanical potential using the 3-21g basis set," *Journal of Computational Chemistry*, vol. 17, no. 4, pp. 386–395, 1996.

[57] D. Riccardi, G. H. Li, and Q. Cui, "Importance of van der waals interactions in qm/mm simulations," *Journal of Physical Chemistry B*, vol. 108, no. 20, pp. 6467–6478, 2004.

[58] G. H. Hou and Q. Cui, "Qm/mm analysis suggests that alkaline phosphatase (ap) and nucleotide pyrophosphatase/phosphodiesterase slightly tighten the transition state for phosphate diester hydrolysis relative to solution: Implication for catalytic promiscuity in the ap superfamily," *Journal of the American Chemical Society*, vol. 134, no. 1, pp. 229–246, 2012.

[59] J. L. Gao, S. H. Ma, D. T. Major, K. Nam, J. Z. Pu, and D. G. Truhlar, "Mechanisms and free energies of enzymatic reactions," *Chem. Rev.*, vol. 106, pp. 3188–3209, 2006.

[60] D. Riccardi, P. Schaefer, Y. Yang, H. Yu, H. Ghosh, X. Prat-Resina, P. König, G. Li, D. Xu, H. Guo, M. Elstner, and Q. Cui, "Development of effective quantum mechanical/molecular mechanical (qm/mm) methods for complex biological processes," *J. Phys. Chem. B*, vol. 110, no. 13, pp. 6458–6469, 2006.

[61] Y. K. Zhang, "Pseudobond ab initio QM/MM approach and its applications to enzyme reactions," *Theo. Chem. Acc.*, vol. 116, pp. 43–50, 2006.

[62] S. C. L. Kamerlin, M. Haranczyk, and A. Warshel, "Progress in ab initio QM/MM free-energy simulations of electrostatic energies in proteins: Accelerated QM/MM studies of pK(a), redox reactions and solvation free energies," *J. Phys. Chem. B*, vol. 113, pp. 1253–1272, 2009.

[63] H. Hu and W. T. Yang, "Free energies of chemical reactions in solution and in enzymes with ab initio quantum mechanics/molecular mechanics methods," *Annu. Rev. Phys. Chem.*, vol. 59, pp. 573–601, 2008.

[64] H. M. Senn and W. Thiel, "QM/MM methods for biomolecular systems," *Angew. Chem. Int. Ed.*, vol. 48, pp. 1198–1229, 2009.

[65] D. Marx and J. Hutter, *Ab initio molecular dynamics: Basic theory and advanced methods*. Cambridge, UK: Cambridge University Press, 2009.

[66] C. J. Cramer and D. G. Truhlar, "Implicit solvation models: Equilibria, structure, spectra, and dynamics," *Chem. Rev.*, vol. 99, no. 8, pp. 2161–2200, 1999.

[67] C. J. Cramer and D. G. Truhlar, "A universal approach to solvation modeling," *Acc. Chem. Res.*, vol. 41, pp. 760–768, 2008.

[68] H. Sato, F. Hirata, and S. Kato, "Analytical energy gradient for the reference interaction site model multiconfigurational self-consistent-field method: Application to 1,2-difluoroethylene in aqueous solution," *J. Chem. Phys.*, vol. 105, pp. 1546–1551, 1996.

[69] D. J. Tannor, B. Marten, R. Murphy, R. A. Friesner, D. Sitkoff, A. Nicholls, M. RingaldaI, W. A. Goddard, and B. Honig, "Accurate first principles calculation of molecular charge-distributions and solvation energies from ab-initio quantum-mechanics and continuum dielectric theory," *J. Am. Chem. Soc.*, vol. 116, no. 26, pp. 11875–11882, 1994.

[70] B. Marten, K. Kim, C. Cortis, and R. A. Friesner, "New model for calculation of solvation free energies: correction of self-consistent reaction field continuum dielectric theory for short-range hydrogen-bond effects," *J. Phys. Chem.*, vol. 100, no. 8, pp. 11775–11788, 1996.

[71] S. Miertus and J. Tomasi, "Approximatie evaluations of the electrostatic free energy and internal energy changes in solution processes," *Chem. Phys.*, vol. 65, pp. 239–245, 1982.

[72] M. Cossi, V. Barone, R. Cammi, and J. Tomasi, "Ab initio study of solvated molecules: a new implementation of the polarizable continuum model," *Chem. Phys. Lett.*, vol. 255, pp. 327–335, 1996.

[73] V. Barone, M. Cossi, and J. Tomasi, "A new definition of cavities for the computation of solvation free energies by the polarizable continuum model," *J. Chem. Phys.*, vol. 107, no. 8, pp. 3210–3221, 1997.

[74] E. Cancès, B. Mennucci, and J. Tomasi, "A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics," *J. Chem. Phys.*, vol. 107, no. 8, pp. 3032–3041, 1997.

[75] B. Mennucci and J. Tomasi, "Continuum solvation models: A new approach to the problem of solute's charge distribution and cavity boundaries," *J. Chem. Phys.*, vol. 106, pp. 5151–5158, 1997.

[76] C. Amovilli and B. Mennucci, "Self-consistent-field calculation of pauli repulsion and dispersion contributions to the solvation free energy in the polarizable continuum model," *J. Phys. Chem. B*, vol. 101, pp. 1051–1057, 1997.

[77] M. Cossi, V. Barone, B. Mennucci, and J. Tomasi, "Ab initio study of ionic solutions by a polarizable continuum dielectric model," *Chem. Phys. Lett.*, vol. 286, pp. 253–260, 1998.

[78] V. Barone, M. Cossi, and J. Tomasi, "Geometry optimization of molecular structures in solution by the polarizable continuum model," *J. Comput. Chem.*, vol. 19, no. 4, pp. 404–417, 1998.

[79] H. Li and J. H. Jensen, "Improving the efficiency and convergence of geometry optimization with the polarizable continuum model: New energy gradients and molecular surface tessellation," *J. Comp. Chem.*, vol. 25, pp. 1449–1462, 2004.

[80] M. Cossi, N. Rega, G. Scalmani, and V. Barone, "Polarizable dielectric model of solvation with inclusion of charge penetration effects," *J. Chem. Phys.*, vol. 114, pp. 5691–5701, 2001.

[81] M. Cossi, G. Scalmani, N. Rega, and V. Barone, "New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution," *J. Chem. Phys.*, vol. 117, pp. 43–54, 2002.

[82] M. Cossi, N. Rega, G. Scalmani, and V. Barone, "Energies, structures, and electronic properties of molecules in solution with the c-pcm solvation model," *J. Chem. Comput.*, vol. 24, pp. 669–681, 2003.

[83] A. V. Marenich, C. J. Cramer, and D. G. Truhlar, "Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions," *J. Phys. Chem. B*, vol. 113, pp. 6378–6396, 2009.

[84] G. D. Hawkins, C. J. Cramer, and D. G. Truhlar, "Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium," *J. Phys. Chem.*, vol. 100, no. 51, pp. 19824–19839, 1996.

[85] D. Qiu, P. S. Shenkin, F. P. Hollinger, and W. C. Still, "The gb/sa continuum model for solvation. a fast analytical method for the calculation of approximate born radii," *J. Phys. Chem. A*, vol. 101, no. 16, pp. 3005–3014, 1997.

[86] A. Ghosh, C. S. Rapp, and R. A. Friesner, "Generalized born model based on a surface integral formulation," *J. Phys. Chem. B*, vol. 102, pp. 10983–10990, 1998.

[87] M. S. Lee, F. R. Salsbury, and C. L. Brooks, "Novel generalized born methods," *J. Chem. Phys.*, vol. 116, pp. 10606–10614, 2002.

[88] W. P. Im, M. S. Lee, and C. L. Brooks, "Generalized born model with a simple smoothing function," *J. Comput. Chem.*, vol. 24, pp. 1691–1702, 2003.

[89] C. P. Kelly, C. J. Cramer, and D. G. Truhlar, "Sm6: A density functional theory continuum solvation model for calculating aqueous solvation free energies of neutrals, ions, and solute-water clusters," *J. Chem. Theory Comput.*, vol. 1, no. 6, pp. 1133–1152, 2005.

[90] A. V. Marenich, R. M. Olson, C. P. Kelly, C. J. Cramer, and D. G. Truhlar, "Self-consistent reaction field model for aqueous and nonaqueous solutions based on accurate polarized partial charges," *J. Chem. Theo. Comp.*, vol. 3, pp. 2011–2033, 2007.

[91] A. Klamt and G. Schuurmann, "Cosmo: A new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient," *J. Chem. Soc. Perkin Trans.*, vol. 2, pp. 799–805, 1993.

[92] A. Klamt, "Conductor-like screening model for real solvent: A new approach to the quantitative calculation of solvation phenomena," *J. Phys. Chem.*, vol. 99, no. 7, pp. 2224–2235, 1995.

[93] A. klamt, V. Jonas, T. Burger, and J. C. W. Lohrenz, "Refinement and parametetrization of cosmo-rs," *J. Phys. Chem. A*, vol. 102, pp. 5074–5085, 1998.

[94] V. Barone and M. Cossi, "Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model," *J. Phys. Chem. A*, vol. 102, pp. 1995–2001, 1998.

[95] D. M. York and M. Karplus, "A smooth solvation potential based on the conductor-like screening model," *J. Phys. Chem. A*, vol. 103, pp. 11060–11079, 1999.

[96] D. M. Dolney, G. D. Hawkins, P. Winget, D. A. Liotard, C. J. Cramer, and D. G. Truhlar, "Universal solvation model based on conductor-like screening model," *J. Comput. Chem.*, vol. 21, pp. 340–366, 2000.

[97] J. Florián and A. Warshel, "Langevin dipoles model for ab initio calculations of chemical processes in solution: parametrization and application to hydration free energies of neutral and ionic solutes and conformational analysis in aqueous solution," *J. Phys. Chem.*, vol. 101, no. 28, pp. 5583–5595, 1992.

[98] D. Wales, *Energy Landscapes*. Cambridge, UK: Cambridge University Press, 2004.

[99] V. Barone, M. Cossi, and J. Tomasi, "A new definition of cavities for the computation of solvation free energies by the polarizable continuum model," *J. Chem. Phys.*, vol. 107, pp. 3210–3221, 1997.

[100] J. B. Foresman, T. A. Keith, K. B. Wiberg, J. Snoonian, and M. J. Frisch, "Solvent effects .5. influence of cavity shape, truncation of electrostatics, and electron correlation ab initio reaction field calculations," *J. Phys. Chem.*, vol. 100, pp. 16098–16104, 1996.

[101] M. J. Vilkas and C. G. Zhan, "An efficient implementation for determining volume polarization in self-consistent reaction field theory," *J. Chem. Phys.*, vol. 129, p. 194109, 2008.

[102] C. G. Zhan and D. M. Chipman, "Cavity size in reaction field theory," *J. Chem. Phys.*, vol. 109, pp. 10543–10558, 1998.

[103] T. Kruger, M. Elstner, P. Schiffels, and T. Frauenheim, "Validation of the density functional based tight-binding approximation method for the calculation of reaction energies and other data," *J. Chem. Phys.*, vol. 122, p. 114110, 2005.

[104] K. W. Sattelmeyer, J. Tirado-Rives, and W. Jorgensen, "Comparison of scc-dftb and nddo-based semiempirical molecular orbital methods for organic molecules," *J. Phys. Chem. A*, vol. 110, pp. 13551–13559, 2006.

[105] N. Otte, M. Scholten, and W. Thiel, "Looking at self-consistent-charge density functional tight binding from a semiempirical perspective," *J. Phys. Chem. A*, vol. 111, pp. 5751–5755, 2007.

[106] M. Elstner, "Scc-dftb: what is the proper degree of self-consistency?," *J. Phys. Chem. A*, vol. 111, no. 26, pp. 5614–5621, 2007.

[107] Y. Yang, H. Yu, D. York, Q. Cui, and M. Elstner, "Extension of the self-consistent-charge density-functional tight-binding method: third-order expansion of the density functional theory total energy and introduction of the modified effective coulomb interaction," *J. Phys. Chem. B*, vol. 111, no. 42, pp. 10861–10873, 2007.

[108] M. Elstner, Q. Cui, P. Munih, E. Kaxiras, T. Frauenheim, and M. Karplus, "Modeling zinc in biomolecules with the self consistent charge-density functional tight binding (scc-dftb) method: applications to structural and energetic analysis," *J. Comput. Chem.*, vol. 24, no. 5, pp. 565–581, 2003.

[109] Z. Cai, P. Lopez, J. R. Reimers, Q. Cui, and M. Elstner, "Application of the computationally efficient self-consistent-charge density-functional-tight-binding method to magnesium-containing molecules," *J. Phys. Chem. A*, vol. 111, pp. 5743–5750, 2007.

[110] G. S. Zheng, H. A. Witek, P. Bobadova-Parvanova, S. Irle, D. G. Musaev, R. Prabhakar, and K. Morokuma, "Parameter calibration of transition-metal elements for the spin-polarized self-consistent-charge density-functional tight-binding (DFTB) method: Sc, Ti, Fe, Co, and Ni," *J. Chem. Theo. Comp.*, vol. 3, pp. 1349–1367, 2007.

[111] N. H. Moreira, G. Dolgonos, B. Aradi, A. L. da Roasa, and T. Frauenheim, "Toward an accurate density-functional tight-binding description of zinc-containing compounds," *J. Chem. Theo. Comp.*, vol. 5, pp. 605–614, 2009.

[112] D. M. York, T. S. Lee, and W. T. Yang, "Parameterization and efficient implementation of a solvent model for linear-scaling semiempirical quantum mechanical calculations of biological macromolecules," *Chem. Phys. Lett.*, vol. 263, no. 1-2, pp. 297–304, 1996.

[113] V. Gogonea and K. M. Merz, "Fully quantum mechanical description of proteins in solution. combining linear scaling quantum mechanical methodologies with the poisson-boltzmann equation," *J. Phys. Chem. A*, vol. 103, no. 26, pp. 5171–5188, 1999.

[114] M. E. Davis and J. A. McCammon, "Electrostatics in biomolecular structure and dynamics," *Chem. Rev.*, vol. 90, p. 509, 1990.

[115] B. Honig and A. Nicholls, "Classical electrostatics in biology and chemistry," *Science*, vol. 268, pp. 1144–1149, 1995.

[116] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "CHARMM: A program for macromolecular energy, minimization and dynamics calculations," *J. Comput. Chem.*, vol. 4, no. 2, pp. 187–217, 1983.

[117] W. Im, D. Beglov, and B. Roux, "Continuum solvation model: computation of electrostatic forces from numerical solutions to the poisson-boltzmann equation," *Comp. Phys. Comm.*, vol. 111, no. 1-3, pp. 59–75, 1998.

[118] M. A. Aguilar and F. J. O. del Valle, "Solute-solvent interactions. a simple procedure for constructing the solvent cavity for retaining a molecular solute," *Chem. Phys.*, vol. 129, pp. 439–450, 1989.

[119] B. Ginovska, D. M. Camaioni, M. Dupuis, C. A. Schwerdtfeger, and Q. Gil, "Charges-dependent cavity radii for an accurate dielectric continuum model of solvation with emphasis on ions: Aqueous solute with oxo, hydroxo, amino, methyl, chloro, bromo, and fluoro functionalities," *J. Phys. Chem. A*, vol. 112, no. 42, pp. 10604–10613, 2008.

[120] B. Ginovska, D. M. Camaioni, and M. Dupuis, "The h_2o_2 + oh −− > ho_2 + h_2o reaction in aqueous solution from a charge-dependent continuum model of solvation," *J. Chem. Phys.*, vol. 129, p. 014506, 2008.

[121] M. Bianciotto, J. C. Barthelat, and A. Vigroux, "Reactivity of phosphate monoester monoanions in aqueous solution. 2. a theoretical study of the elusive zwitterion intermediates $ro^+(h)po_3^{2-}$," *J. Phys. Chem. A*, vol. 106, no. 27, pp. 6521–6526, 2002.

[122] B. Roux and T. Simonson, "Implicit solvent models," *Bio. Chem.*, vol. 78, pp. 1–20, 1999.

[123] J. D. Jackson, *Classical Electrodynamics*. New York: John Wiley & Sons, 3rd ed., 2001.

[124] W. Im, S. Bernèche, and B. Roux, "Generalized solvent boundary potential for computer simulations," *J. Chem. Phys.*, vol. 114, no. 7, pp. 2924–2937, 2001.

[125] D. A. McQuarrie, *Statistical Mechanics*. New York: Harper & Row, 1976.

[126] Y. Yamaguchi, J. D. Goddard, Y. Osamura, and H. Schaefer, *A new dimension to quantum chemistry: Analytic derivative methods in Ab initio molecular electronic structure theory*. Oxford, UK: Oxford University Press, 1994.

[127] M. Feig and C. L. I. Brooks, "Gb review," *Curr. Opin. Struct. Biol.*, vol. 14, pp. 217–224, 2004.

[128] L. Xie and H. Liu, "The treatment of solvation by a generalized born model and a self-consistent charge-density functional theory-based tight-binding method," *J. Comput. Chem.*, vol. 23, no. 15, pp. 1404–1415, 2002.

[129] D. E. Goldberg, *Genetic algorithms in search, optimization, and machine learning*. Addison-Wesley: Reading, MA, 1989.

[130] D. L. Carroll, "http://cuaerospace.com/carroll/ga.html."

[131] S. A. Ba-Saif, A. M. Davis, and A. Williams, "Effective charge distribution for attack of phenoxide ion on aryl methyl phosphate monoanion: studies related to the action of ribonuclease," *J. Org. Chem.*, vol. 54, no. 23, pp. 5483–5486, 1989.

[132] J. A. Barnes, J. Wilkie, and I. H. Williams, "Transition-state structure variation and mechanistic change," *J. Chem. Soc. Faraday Trans.*, vol. 90, no. 12, pp. 1709–1714, 1994.

[133] S. Fischer and M. Karplus, "Conjugate peak refinement: an algorithm for finding reaction paths and accurate transition states in systems with many degrees of freedom," *Chem. Phys. Lett.*, vol. 194, no. 3, pp. 511–527, 1992.

[134] S. C. L. Kamerlin, M. Haranczyk, and A. Warshel, "Are mixed explicit/implicit solvation models reliable for studying phosphate hydrolysis? a comparative study of continuum, explicit and mixed solvation models," *ChemPhyschem*, vol. 10, pp. 1125–1134, 2009.

[135] Q. Cui and M. Karplus, "Quantum mechanical/molecular mechanical studies of the triosephosphate isomerase-catalyzed reaction: Verification of methodology and analysis of reaction mechanisms," *J. Phys. Chem B*, vol. 106, pp. 1768–1798, 2002.

[136] J. Florián and A. Warshel, "A fundamental assumption about oh^- attack in phosphate ester hydrolysis is not fully justified," *J. Am. Chem. Soc.*, vol. 119, no. 23, pp. 5473–5474, 1997.

[137] A. Bondi, "vad der wall volumes and radii," *J. Phys. Chem.*, vol. 68, pp. 441–451, 1964.

[138] P. W. C. Barnard, C. A. Bunton, D. R. Llewellyn, and K. Oldham *Chem. Ind. (London)*, vol. 760, pp. 2420–2423, 1955.

[139] W. W. Butcher and F. H. Wesheimer *J. Am. Chem. Soc.*, vol. 77, pp. 2420–, 1955.

[140] C. A. Bunton, D. R. Llewellyn, K. G. Oldham, and C. A. Vernon *J. Chem. Soc*, pp. 3574–, 1958.

[141] C. A. Bunton, D. R. Llewellyn, K. G. Oldham, and C. A. Vernon, "The reaction of organic phosphate," *J. Chem. Soc.*, pp. 3574–3587, 1958.

[142] T. J. Giese and D. M. York, "Charge-dependent model for many-body polarization, exchange, and dispersion interactions in hybrid quantum mechanical/molecular mechanical calculations," *J. Chem. Phys.*, vol. 127, p. 194101, 2007.

[143] P. W. C. Barnard, C. A. Bunton, D. R. Llewellyn, C. A. Vernon, and V. A. Welch, "The reactions of organic phosphates.," *J. Chem. Soc.*, pp. 2670–2676, 1961.

[144] Q. Cui, "Combining implicit solvation models with hybrid quantum mechanical/molecular mechanical methods: A critical test with glycine," *J. Chem. Phys.*, vol. 117, no. 10, pp. 4720–4728, 2002.

[145] H. Li and M. S. Gordon, "Polarization energy gradients in combined quantum mechanics, effective fragment potential, and polarizable continuum model calculations," *J. Chem. Phys.*, vol. 126, p. 124112, 2007.

[146] M. J. Field, P. A. Bash, and M. Karplus, "A combined quantum-mechanical and molecular mechanical potential for molecular-dynamics simulations," *Journal of Computational Chemistry*, vol. 11, no. 6, pp. 700–733, 1990.

[147] N. Reuter, A. Dejaegere, B. Maigret, and M. Karplus, "Frontier bonds in qm/mm methods: A comparison of different approaches," *Journal of Physical Chemistry A*, vol. 104, no. 8, pp. 1720–1735, 2000.

[148] J. L. Gao, P. Amara, C. Alhambra, and M. J. Field, "A generalized hybrid orbital (gho) method for the treatment of boundary atoms in combined qm/mm calculations," *Journal of Physical Chemistry A*, vol. 102, no. 24, pp. 4714–4721, 1998.

[149] T. J. Giese and D. M. York, "Charge-dependent model for many-body polarization, exchange, and dispersion interactions in hybrid quantum mechanical/molecular mechanical calculations," *Journal of Chemical Physics*, vol. 127, no. 19, 2007.

[150] G. Klopman *Journal of the American Chemical Society*, vol. 86, pp. 4550–, 1964.

[151] K. Ohno *Theor. Chim. Acta*, vol. 2, pp. 219–, 1964.

[152] M. Kolb and W. Thiel, "Beyond the mndo model - methodical considerations and numerical results," *Journal of Computational Chemistry*, vol. 14, no. 7, pp. 775–789, 1993.

[153] M. Gaus, Q. A. Cui, and M. Elstner, "Dftb3: Extension of the self-consistent-charge density-functional tight-binding method (scc-dftb)," *Journal of Chemical Theory and Computation*, vol. 7, no. 4, pp. 931–948, 2011.

[154] D. Das, K. P. Eurenius, E. M. Billings, P. Sherwood, D. C. Chatfield, M. Hodoscek, and B. R. Brooks, "Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method," *Journal of Chemical Physics*, vol. 117, no. 23, pp. 10534–10547, 2002.

[155] P. Politzer, R. Parr, and D. Murphy, "Relationships between atomic chemical potentials, electrostatic potentials and covalent radii," *J. Chem. Phys.*, vol. 79, pp. 3859–3861, 1983.

[156] R. Pearson, "Absolute electronegativity and hardness-application to inorganic-chemistry," *Inorg. Chem.*, vol. 27, pp. 734–740, 1988.

[157] D. Ghosh and R. Biswas, "Theoretical calculations of absolute radii of atoms and ions. part 2. the ionic radii," *Int. J. Mol. Sci.*, vol. 4, pp. 379–407, 2003.

[158] P. Politzer, J. Murray, and P. Lane, "Electrostatic potentials and covalent radii," *J. Comp. Chem.*, vol. 24, pp. 505–511, 2003.

[159] C. Lad, N. H. Williams, and R. Wolfenden, "The rate of hydrolysis of phosphomonoester dianions and the exceptional catalytic proficiencies of protein and inositol phosphatases," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 10, pp. 5607–5610, 2003.

[160] P. O'Brie and D. Herschlag, "Alkaline phosphatase revisited: hydrolysis of alkyl phosphates," *Biochem.*, vol. 41, pp. 3207–3225, 2002.

[161] C. Boorks and M. Karplus, "Deformable stochastic boundaries in molecular dynamics," *J. Chem. Phys.*, vol. 79, pp. 6312–6325, 1983.

[162] W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey, and M. Klein, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.*, vol. 79, pp. 926–935, 1983.

[163] P. Schaefer, D. Riccardi, and Q. Cui, "Reliable treatment of electrostatics in combind qm/mm simulation of macromolecules," *J. Chem. Phys.*, vol. 123, p. 014905, 2005.

[164] P. Steinbach and B. Brooks, "New spherical-cutoff methods for long-range forces in macromolecular simulation," *J. Comput. Chem.*, vol. 15, pp. 667–683, 1994.

[165] C. L. Brooks and M. Karplus, "Deformable stochastic boundaries in molecular-dynamics," *Journal of Chemical Physics*, vol. 79, no. 12, pp. 6312–6325, 1983.

[166] J. Rychaert, G. Ciccotti, and H. Berendsen, "Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes," *J. Comput. Phys.*, vol. 23, pp. 327–341, 1977.

[167] G. M. Torrie and J. P. Valleau, "Non-physical sampling distributions in monte-carlo free-energy estimation - umbrella sampling," *Journal of Computational Physics*, vol. 23, no. 2, pp. 187–199, 1977.

[168] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, "The weighted histogram analysis method for free-energy calculations on biomolecules .1. the method," *Journal of Computational Chemistry*, vol. 13, no. 8, pp. 1011–1021, 1992.

[169] G. Hou and Q. Cui *Journal of the American Chemical Society*, vol. in press, 2011.

[170] T. J. Giese, B. A. Gregersen, Y. Liu, K. Nam, E. Mayaan, A. Moser, K. Range, A. N. Faza, C. S. Lopez, A. R. de Lera, G. Schaftenaar, X. Lopez, T. S. Lee, G. Karypis, and D. M. York, "Qcrna 1.0: A database of quantum calculations for rna catalysis," *Journal of Molecular Graphics & Modelling*, vol. 25, no. 4, pp. 423–433, 2006.

[171] J. Lassila, J. Zalatan, and D. Herschlag, "Biological phosphoryl-transfer reactions: understanding mechanism and catalysis," *Annu. Rev. Biochem.*, vol. 80, pp. 669–702, 2011.

[172] R. A. Jensen, "Enzyme recruitment in evolution of new function," *Annu. Rev. Microbio.*, vol. 30, pp. 409–425, 1976.

[173] O. Khersonsky and D. S. Tawfik, "Enzyme promiscuity: A mechanistic and evolutionary perspective," *Annu. Rev. Biochem.*, vol. 79, pp. 471–505, 2010.

[174] B. van Loo, S. Jonas, A. C. Babtie, A. Benjdia, O. Berteau, M. Hyvönen, and F. Hollfelder, "An efficient, multiply promiscuous hydrolase in the alkaline phosphatase superfamily," *Proc. Nat. Acad. Sci. USA*, vol. 107, pp. 2740–2745, 2010.

[175] C. Lad, N. H. Williams, and R. Wolfenden, "The rate of hydrolysis of phosphomonoester dianions and the exceptional catalytic proficiencies of protein and inositol phosphatases," *Proc. Natl. Acad. Sci. USA*, vol. 100, pp. 5607–5610, 2003.

[176] J. Lassila and D. Herschlag, "Promiscuous sulfatase activity and thio-effects in a phosphodiesterase of the alkaline phosphatase superfamily," *Biochem.*, vol. 47, pp. 12853–12859, 2008.

[177] B. Stec, K. Holtz, and E. Kantrowitz, "A revised mechanism for the alkaline phosphatase reaction involving three metal ions," *J. Mol. Biol.*, vol. 299, pp. 1303–1311, 2000.

[178] J. K. Lassila, J. G. Zalatan, and D. Herschlag, "Biological phosphoryl transfer reactions: Understanding mechanism and catalysis," *Annu. Rev. Biochem.*, vol. 80, pp. 669–702, 2011.

[179] V. López-Canut, S. Marti, J. Bertrán, V. Moliner, and I. Tuñón, "Theoretical modeling of the reaction mechanism of phosphate monoester hydrolysis in alkaline phosphatase," *J. Phys. Chem. B*, vol. 113, no. 22, pp. 7816–7824, 2009.

[180] K. Nam, Q. Cui, J. Gao, and D. York, "Specific reaction parameterization of the am1/d hamiltonian for phosphoryl transfer reactions: H, o, and p atoms," *J. Chem. Theory Comput.*, vol. 3, pp. 486–504, 2007.

[181] C. McWhirter, E. A. Lund, E. A. Tanifum, G. Feng, Q. I. Sheikh, A. C. Hengge, and N. H. Williams, "Mechanistic study of protein phosphatase-1 (pp1), a catalytically promiscuous enzyme," *J. Am. Chem. Soc.*, vol. 130, pp. 13673–13682, 2008.

[182] P. O'Brien, J. Lassila, T. Fenn, J. Zalatan, and D. Herschlag, "Arginine coordination in enzymatic phosphoryl transfer: evaluation of the effect of arg166 mutations in escherichia coli alkaline phosphatase," *Biochem.*, vol. 47, pp. 7663–7672, 2008.

[183] W. Thiel, "Perspectives on semiempirical molecular orbital theory," *Adv. Chem. Phys.*, vol. 93, pp. 703–757, 1996.

[184] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, J. A. Montgomery, J. T. Vreven, K. N. Kudin, J. C. Burant, J. M. Millam, S. S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G. A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J. E. Knox, H. P. Hratchian, J. B. Cross, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, P. Y. Ayala, K. Morokuma, G. A. Voth, P. Salvador, J. J. Dannenberg, V. G. Zakrzewski, S. Dapprich, A. D. Daniels, M. C. Strain, O. Farkas, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. V. Ortiz, Q. Cui, A. G. Baboul, S. Clifford, J. Cioslowski, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, M. Challacombe, P. M. W. Gill, B. Johnson, W. Chen, M. W. Wong, C. Gonzalez, and J. A. Pople, "Gaussian 03," 2003.

[185] G. H. Li and Q. Cui, "pk(a) calculations with qm/mm free energy perturbations," *J. Phys. Chem. B*, vol. 107, no. 51, pp. 14521–14528, 2003.

[186] S. Jonas, B. van Loo, M. Hyvönen, and F. Hollfelder, "A new member of the alkaline phosphatase superfamily with a formylglycine nucleophile: Structural and kinetic characterisation of a phosphonate monoester hydrolase/phosphodiesterase from rhizobium leguminosarum," *J. Mol. Biol.*, vol. 384, pp. 120–136, 2008.

[187] K. M. Holtz, I. E. Catrina, A. C. Hengge, and E. R. Kantrowitz, "Mutation of arg-166 of alkaline phosphatase alters the thio effect but not the transition state for phosphoryl transfer. implications for the interpretation of thio effects in reactions of phosphatases," *Biochemistry*, vol. 39, no. 31, pp. 9451–9458, 2000.

[188] A. Brunger and M. Karplus, "Polar hydrogen positions in proteins-empirical energy placement and neutron-diffraction comparison," *Protein Struct. Funct. Genet.*, vol. 4, pp. 148–156, 1988.

[189] B. Boorks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus, "Charmm-a program for macromolecular energy, minimization, and dynamics calculations," *J. Comput. Chem.*, vol. 4, pp. 187–217, 1983.

[190] A. MacKerell, D. Bashford, M. Bellott, R. Dunbrack, J. Evanseck, M. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. Lau, C. Mattos, S. Michnick, T. Ngo, D. Nguyen, B. Prodhom, W. Reiher, B. Roux, M. Schlenkrich, J. Smith, R. Stote, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," *J. Chem. Phys.*, vol. 102, pp. 3586–3616, 1998.

[191] P. Konig, M. Hoffmann, T. Frauenheim, and Q. Cui, "A critical evaluation of different qm/mm frontier treatments with scc-dftb as the qm method," *J. Phys. Chem. B*, vol. 109, pp. 9082–9095, 2005.

[192] G. Arantes and M. Loos, "Specific parameterization of a hybrid potential to simulate reactions in phosphatases," *Phys. Chem. Chem. Phys.*, vol. 8, pp. 347–353, 2006.

[193] C. Brooks and M. Karplus, "Solvent effects on protein motion and protein effects on solvent motion: Dynamics of the active-site region of lysozyme," *J. Mol. Biol.*, vol. 208, pp. 159–181, 1989.

[194] M. Nina, D. Beglov, and D. Roux, "Atomic radii for continuum electrostatics calculations based on molecular dynamics free energy simulations," *J. Phys. Chem. B*, vol. 101, pp. 5239–5248, 1997.

[195] M. Nina, W. Im, and D. Roux, "Optimized atomic radii for protein continuum electrostatics solvation forces," *Biophys. Chem.*, vol. 78, pp. 89–96, 1999.

[196] A. Becke, "Density-functional exchange-energy approximation with correct asymptotic-behavior," *Phys. Rev. A*, vol. 38, pp. 3098–3100, 1988.

[197] A. Becke, "Density-functional thermochemistry .3. the role of exact exchange," *J. Chem. Phys.*, vol. 98, pp. 5648–5652, 1993.

[198] C. Lee, W. Yang, and R. Parr, "Development of the colle-salvetti correlation-energy formula into a functional of the electron-density," *Phys. Rev. B*, vol. 37, pp. 785–789, 1988.

[199] G. Petersson, A. Bennett, T. Tensfeldt, M. Allaham, W. Shirley, and J. Mantzaris, "A complete basis set model chemistry .1. the total energies of closed-shell atoms and hydrides of the 1st-row elements," *J. Chem. Phys.*, vol. 89, pp. 2193–2218, 1988.

[200] Y. Shao, L. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. Brown, A. Gilbert, L. Slipchenko, D. O'Neill, R. DiStasio, R. Lochan, T. Wang, G. Beran, N. Besley, J. Herbert, C. Lin, T. Van Voorhis, S. Chien, A. Sodt, R. Steele, V. Rassolov, P. Maslen, P. Korambath, R. Adamson, B. Austin, J. Baker, E. Byrd, H. Bachsel, R. Doerksen, A. Dreuw, B. Dunietz, A. Dutoi, T. Furlani, S. Gwaltney, A. Heyden, S. Hirata, C. Hsu, G. Kedziora, R. Khalliulin, P. Klunzinger, A. Lee, M. Lee, W. Liang, I. Lotan, N. Nair, B. Peters, E. Proynov, P. Pieniazek, Y. Rhee, J. Ritchie, E. Rosta, C. Sherrill, A. Simmonett, J. Subotnik, H. Woodcock, W. Zhang, A. Bell, A. Chakraborty, D. Chipman, F. Keil, A. Warshel, W. Hehre, H. Schaefer, J. Kong, A. Krylov, P. Gill, and M. Head-Gordon, "Advances in methods and algorithms in a modern quantum chemistry program package," *Phys. Chem. Chem. Phys.*, vol. 27, pp. 3172–3191, 2006.

[201] B. R. Brooks, C. L. B. III, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, and M. Karplus, "Charmm: The biomolecular simulation program," *J. Comp. Chem.*, vol. 30, pp. 1545–1614, 2009.

[202] D. Riccardi, P. Schaefer, and Q. Cui, "pka calculations in solution and proteins with qm/mm free energy perturbation simulations," *J. Phys. Chem. B*, vol. 109, pp. 17715–17733, 2005.

[203] M. Fujio, R. T. Mciver, and R. W. Taft, "Effects on the acidities of phenols from specific substituent-solvent interactions - inherent substituent parameters from gas-phase acidities," *J. Am. Chem. Soc.*, vol. 103, no. 14, pp. 4017–4029, 1981.

[204] D. R. Lide, ed., *CRC Handbook Chemistry and Physics*. CRC Press, 85 ed., 2005.

[205] A. C. Hengge, A. E. Tobin, and W. W. Cleland, "Studies of transition-state structures in phosphoryl transfer-reactions of phosphodiesters of p-nitrophenol," *J. Am. Chem. Soc.*, vol. 117, no. 22, pp. 5919–5926, 1995.

[206] M. E. Harris, A. G. Cassano, and V. E. Anderson, "Evidence for direct attack by hydroxide in phosphodiester hydrolysis," *J. Am. Chem. Soc.*, vol. 124, no. 37, pp. 10964–10965, 2002.

[207] I. Tunon, V. Lopez-Canut, J. Ruiz-Pernia, S. Ferrer, and V. Moliner, "Theoretical modeling on the reaction mechanism of p-nitrophenylmethylphosphate alkaline hydrolysis and its kinetic isotope effects," *J. Chem. Theo. Comp.*, vol. 5, no. 3, pp. 439–442, 2009.

[208] M. Gaus, Q. Cui, and M. Elstner, "Dftb-3rd: Extension of the self-consistent-charge density-functional tight-binding method SCC-DFTB," *J. Chem. Theo. Comp.*, vol. 7, pp. 931–948, 2011.

[209] M. Gaus, C. P. Chou, H. Witek, and M. Elstner, "Automatized parametrization of scc-dftb repulsive potentials: Application to hydrocarbons," *J. Phys. Chem. A*, vol. 113, pp. 11866–11881, 2009.

[210] K. M. Holtz, B. Stec, and E. R. Kantrowitz, "A model of the transition state in the alkaline phosphatase reaction," *J. Biol. Chem.*, vol. 274, pp. 8351–8354, 1999.

[211] K. Y. Wong and J. L. Gao, "The reaction mechanism of paraoxon hydrolysis by phosphotriesterase from combined qm/mm simulations," *Biochem.*, vol. 46, pp. 13352–13369, 2007.

[212] K. Y. Wong and J. L. Gao, "Insight into the phosphodiesterase mechanism from combined qm/mm free energy simulations," *FEBS J.*, vol. 278, pp. 2579–2595, 2011.

[213] D. Das, K. P. Eurenius, E. M. Billings, P. Sherwood, D. C. Chatfield, M. Hodoscek, and B. R. Brooks, "Optimization of quantum mechanical molecular mechanical partitioning schemes: Gaussian delocalization of molecular mechanical charges and the double link atom method," *J. Chem. Phys.*, vol. 117, pp. 10534–10547, 2002.

[214] E. E. Kim and H. W. Wyckoff, "Reaction-mechanism of alkaline-phosphatase based on crystal-structures - 2-metal ion catalysis," *J. Mol. Biol.*, vol. 218, pp. 449–464, 1991.

[215] N. Strater, W. N. Lipscomb, T. Klabunde, and B. Krebs, "Two-metal ion catalysis in enzymatic acyl- and phosphoryl-transfer reactions," *Angew. Chem. Int. Ed.*, vol. 35, pp. 2024–2055, 1996.

[216] T. A. Steitz and J. A. Steitz, "A general 2-metal-ion mechanism for catalytic RNA," *Proc. Natl. Acad. Sci. USA*, vol. 90, pp. 6498–6502, 1993.

[217] J. J. G. Tesmer, R. K. Sunahara, R. A. Johnson, G. Gosselin, A. G. Gilman, and S. R. Sprang, "Two-metal-ion catalysis in adenylyl cyclase," *Science*, vol. 285, pp. 756–760, 1999.

[218] M. J. Jedrzejas and P. Setlow, "Comparison of the binuclear metalloenzymes diphosphoglycerate-independent phosphoglycerate mutase and alkaline phosphatase: Their mechanism of catalysis via a phosphoserine intermediate," *Chem. Rev.*, vol. 101, pp. 607–618, 2001.

[219] I. Nikolic-Hughes, P. O'Brien, and D. Herschlag, "Alkaline phosphatase catalysis is ultrasensitive to charge sequestered between the active site zinc ions," *J. Am. Chem. Soc.*, vol. 127, pp. 9314–9315, 2005.

[220] H. Gao, Z. Ke, N. J. DeYonker, J. Wang, H. Xu, Z. Mao, D. L. Phillips, and C. Zhao, "Dinuclear zn(ii) complex catalyzed phosphodiester cleavage proceeds via a concerted mechanism: A density functional theory study," *J. Am. Chem. Soc.*, vol. 133, pp. 2904–2915, 2011.

[221] Y. B. Fan and Y. Q. Gao, "Coorperativity between metals, ligands and solvent: a dft study on the mechanism of a dizinc complex-mediated phosphodiester cleavage," *Acta Phys. Chim. Sinica*, vol. 26, pp. 1034–1042, 2010.

[222] J. C. Hermann, E. Ghanem, Y. Li, F. M. Raushel, J. J. Irwin, and B. K. Shoichet, "Predicting substrates by docking high-energy intermediates to enzyme structures," *J. Am. Chem. Soc.*, vol. 128, pp. 15882–15891, 2006.

[223] J. C. Hermann, R. Marti-Arbona, A. A. Fedorov, E. Fedorov, S. C. Almo, B. K. Shoichet, and F. M. Raushel, "Structure-based activity prediction for an enzyme of unknown function," *Nature*, vol. 448, pp. 775–779, 2007.

[224] M. D. Toscano, K. J. Woycechowsky, and D. Hilvert, "Minimalist active-site redesign: teaching old enzymes new tricks," *Angew. Chem. Int. Ed.*, vol. 46, pp. 3212–3236, 2007.

[225] L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Rothlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas, D. Hilvert, K. N. Houk, B. L. Stoddard, and D. Baker, "De novo computational design of retro-aldol enzymes," *Science*, vol. 319, pp. 1387–1391, 2008.

[226] D. G. Truhlar and Y. Zhao, "The m06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four m06-class functionals and 12 other functionals," *Theoretical Chemistry Accounts*, vol. 120, no. 1-3, pp. 215–241, 2008.

[227]

[228] M. Trajbl, G. Y. Hong, and A. Warshel, "Ab initio qm/mm simulation with proper sampling: "first principle" calculations of the free energy of the autodissociation of water in aqueous solution," *Journal of Physical Chemistry B*, vol. 106, no. 51, pp. 13333–13343, 2002.

[229] M. Elstner, T. Frauenheim, and S. Suhai, "An approximate dft method for qm/mm simulations of biological structures and processes," *J. Mol. Struct.: THEOCHEM*, vol. 632, pp. 29–41, 2003.

[230] M. Elstner, M. Gaus, M Gaus, and Q. A. Cui, "Dftb3: Extension of the self-consistent-charge density-functional tight-binding method (scc-dftb)," *Journal of Chemical Theory and Computation*, vol. 7, no. 4, pp. 931–948, 2011.

[231] G. Hou, X. Zhu, M. Elstner, and Cui, "Charge dependent qm/mm interactions with the self-consistent-charge tight-binding-density-functional theory," *to be submitted*.

Appendix A: Supporting Information: An implicit solvent model for SCC-DFTB with Charge-Dependent Radii

Table A.1: Error (in kcal/mol) Analysis of Solvation Free Energies for Training Set 1[a]

		Signed Error		
Solute	ΔG_{exp}	Single Point[b]	Optimization[c]	SM6[d]
Methane	2.0	-1.8	-1.8	0.0
Propane	2.0	-1.7	-1.7	-0.7
Neopentane	2.5	-2.2	-2.2	-0.4
n-Heptane	2.6	-2.2	-2.2	-0.7
Cyclohexane	1.2	-0.9	-0.9	-0.5
Ethene	1.3	-1.5	-1.5	0.2
Isobutene	1.2	-1.6	-1.6	0.1
1-Pentene	1.7	-1.9	-1.9	-0.1
Cyclopentene	0.6	-1.0	-1.0	-0.8
Propyne	-0.3	-1.7	-1.7	-0.4
1-Pentyne	0.0	-1.6	-1.7	0.2
Benzene	-0.9	-0.1	-0.1	-0.5
Ethylbenzene	-0.8	-0.1	-0.1	0.2
p-Xylene	-0.8	-0.1	-0.1	-0.1
Naphthalene	-2.4	1.1	1.1	-0.3
Anthracene	-4.2	2.6	2.6	0.3
Phenol	-6.6	2.5	2.7	1.4
p-Cresol	-6.1	2.4	2.3	1.2
Methanol	-5.1	1.3	1.2	0.2
Ethanol	-5.0	1.2	1.0	0.3

t-Butanol	-4.5	0.9	0.8	1.6
3-Pentanol	-4.3	1.1	0.9	1.6
Dimethyl ether	-1.9	-0.7	-0.8	0.2
Diethyl ether	-1.8	-0.8	-1.0	0.4
1,2-Dimethoxyethane	-4.8	0.7	0.5	1.4
Butanal	-3.2	-0.6	-1.0	0.0
Pentanal	-3.0	-0.7	-1.2	0.2
Benzaldehyde	-4.0	-0.2	-0.6	-0.7
Acetic acid	-6.7	-0.5	-1.4	0.6
Butanoic acid	-6.4	-0.5	-1.3	1.4
Hexanoic acid	-6.2	-0.6	-1.3	1.6
2-Butanone	-3.6	-0.9	-1.5	-0.4
3-Pentanone	-3.4	-1.1	-1.6	0.3
Cyclopentanone	-4.7	0.5	0.0	0.5
3-Methylindole	-5.9	1.8	1.7	1.2
n-Propylguanidine	-10.9	3.9	3.1	1.6
4-Methylimidazole	-10.3	4.2	4.0	2.6
Methylamine	-4.6	3.8	3.8	0.2
Ethylamine	-4.5	3.8	3.8	0.7
n-Butylamine	-4.3	3.7	3.6	0.9
Piperidine	-5.1	4.8	4.8	1.0
Diethylamine	-4.1	3.7	3.7	1.7
Aniline	-5.5	2.4	2.1	0.7
Acetonitrile	-3.9	0.3	0.2	-1.3
Ammonia	-4.3	3.2	3.2	-0.4
Formic acid (-1)	-78	0	-3	-1
Acetic acid (-1)	-80	2	-2	2

Hexanoic acid (-1)	-76	0	-4	3
Acrylic acid (-1)	-76	-1	-3	-1
Pyruvic acid (-1)	-70	-5	-7	5
Benzoic acid (-1)	-73	0	-3	0
Methanol (-1)	-97	12	5	6
Ethanol (-1)	-93	10	3	8
2-Propanol (-1)	-88	7	0	7
t-Butanol (-1)	-84	4	-2	9
Allyl alcohol (-1)	-88	8	2	6
Benzyl alcohol (-1)	-87	12	6	12
Phenol (-1)	-74	6	5	5
4-Methylphenol (-1)	-74	7	5	5
1,2-Ethanediol (-1)	-87	0	-4	1
4-Hydroxyphenol (-1)	-80	10	8	8
Acetaldehyde (-1)	-78	2	0	0
Acetone (-1)	-78	3	0	2
3-Pentanone (-1)	-76	5	2	6
Acetonitrile (-1)	-74	1	2	0
Cyanamide (-1)	-74	-2	-2	-2
Aniline (-1)	-65	2	1	-3
Diphenylamine (-1)	-56	3	3	-2
4-Nitrophenol (-1)	-60	0	-1	3
Nitromethane (-1)	-78	6	3	3
4-Nitroaniline (-1)	-59	2	1	1
Methanol (+1)	-91	10	9	9
Diethyl ether (+1)	-70	7	7	11
Acetone (+1)	-75	7	7	9

Acetophenone (+1)	-63	7	6	9
Methylamine (+1)	-74	-3	-3	-5
n-Propylamine (+1)	-70	-2	-3	-2
Cyclohexanamine (+1)	-67	-1	-1	1
Allylamine (+1)	-70	-3	-3	-1
Dimethylamine (+1)	-67	-2	-3	-1
Di-n-propylamine (+1)	-59	-1	-1	3
Diallylamine (+1)	-60	-2	-2	5
Trimethylamine (+1)	-59	-4	-4	-1
Tri-n-propylamine (+1)	-49	-3	-3	2
Aniline (+1)	-70	2	1	2
4-Methylaniline (+1)	-68	2	1	2
3-Aminoaniline (+1)	-64	-1	-2	-4
N-methylaniline (+1)	-61	-1	-1	2
N,N-dimethylaniline (+1)	-55	-1	-1	3
4-Methyl-N,N-dimethylaniline (+1)	-54	0	0	4
1-Aminonaphthalene (+1)	-66	1	1	2
Aziridine (+1)	-69	-2	-2	-4
Pyrrolidine (+1)	-64	-1	-1	0
Azacycloheptane (+1)	-61	-1	-1	1
Pyridine (+1)	-59	-1	-1	-1
Quinoline (+1)	-54	1	1	2
Piperazine (+1)	-64	0	0	-1
Acetonitrile (+1)	-73	3	3	3
4-Methoxyaniline (+1)	-69	4	3	2
Morpholine (+1)	-68	-2	-2	-1
Acetamide (+1)	-72	5	4	-6

Ammonia (+1)	-83	-3	-3	-9
Hydrazine (+1)	-83	4	3	-1
Error Analysis				
RMSE		3	3	3
MUE		3	2	2
MSE		1	0	1

a. RMSE: Root-Mean-Square-Error; MUE: Mean-Unsigned-Error; MSE: Mean-Signed-Error. All errors measured against experimental solvation free energies, which have typical uncertainties of 0.2 kcal/mol and 3 kcal/mol for neutral molecules and ions, respectively. b. With gas-phase geometries. c. With solution phase geometry optimizations (see **Methods**). d. Results are obtained by MPW1PW91/6-31+G(d,p).

Table A.2: Error (in kcal/mol) Analysis of Solvation Free Energies for Training Set 2

		Signed Error		
Solute	ΔG_{exp}	Single Point	Optimization	SM6
Propane	2.0	-1.7	-1.7	-0.7
Neopentane	2.5	-2.1	-2.1	-0.4
n-Heptane	2.6	-2.1	-2.1	-0.7
Cyclohexane	1.2	-0.8	-0.8	-0.5
Ethene	1.3	-1.4	-1.4	0.2
Cyclopentene	0.6	-0.7	-0.7	-0.8
Benzene	-0.9	0.3	0.3	-0.5
Ethylbenzene	-0.8	0.3	0.3	0.2
p-Xylene	-0.8	0.3	0.3	-0.1
Naphthalene	-2.4	1.6	1.6	-0.3

Anthracene	-4.2	3.2	3.2	0.3
Phenol	-6.6	1.5	2.5	1.4
p-Cresol	-6.1	2.0	0.9	1.2
Methanol	-5.1	-0.3	-0.7	0.2
Ethanol	-5.0	-0.5	-0.9	0.3
t-Butanol	-4.5	-0.9	-1.3	1.6
3-Pentanol	-4.3	-0.5	-0.9	1.6
Dimethyl ether	-1.9	-0.4	-0.6	0.2
Diethyl ether	-1.8	-0.5	-0.8	0.4
1,2-Dimethoxyethane	-4.8	1.2	0.9	1.4
Butanal	-3.2	-0.8	-2.1	0.0
Pentanal	-3.0	-0.9	-2.2	0.2
Benzaldehyde	-4.0	-0.4	-2.0	-0.7
Acetic acid	-6.7	-3.4	-5.2	0.6
Butanoic acid	-6.4	-3.0	-4.7	1.4
Hexanoic acid	-6.2	-3.0	-4.7	1.6
2-Butanone	-3.6	-2.0	-3.9	-0.4
3-Pentanone	-3.4	-1.9	-3.7	0.3
Cyclopentanone	-4.7	-0.6	-2.6	0.5
Phosphine	0.6	-0.3	-0.3	0.3
Trimethyl phosphate	-8.7	-0.5	-1.9	1.3
Methyl phosphonic diester	-10.1	-1.0	-4.6	2.9
Dimethyl hydrogen phosphite	-14.6	3.5	-0.1	7.4
Formic acid (-1)	-78	1	-1	-1
Acetic acid (-1)	-80	3	0	2
Hexanoic acid (-1)	-76	1	-2	3
Pyruvic acid (-1)	-70	5	2	5
Benzoic acid (-1)	-73	1	-2	0

Methanol (-1)	-97	11	4	6
Ethanol (-1)	-93	8	2	8
2-Propanol (-1)	-88	4	-1	7
t-Butanol (-1)	-84	1	-4	9
Allyl alcohol (-1)	-88	7	2	6
Benzyl alcohol (-1)	-87	11	6	12
Phenol (-1)	-74	7	5	5
4-Methylphenol (-1)	-74	7	5	5
1,2-Ethanediol (-1)	-87	-4	-5	1
4-Hydroxyphenol (-1)	-80	10	7	8
Acetone (-1)	-78	4	8	2
3-Pentanone (-1)	-76	6	4	6
Dihydrogen phosphate (-1)	-76	0	-5	-3
Dimethyl phosphate (-1)	-75	3	-2	0
Methanol (+1)	-91	9	8	9
Diethyl ether (+1)	-70	7	7	11
Acetone (+1)	-75	9	8	9
Acetophenone (+1)	-63	8	7	9
Phosphonium (+1)	-73	0	0	-4
Error Analysis				
RMSE		4	4	4
MUE		3	3	3
MSE		2	0	2

See Table A.1 for format.

Table A.3: Error (in kcal/mol) Analysis of Solvation Free Energies for Test Set 1

		Signed Error		
Solute	ΔG_{exp}	Single Point	Optimization	SM6
Ethane	1.8	-1.7	-1.7	-0.6
Cyclopropane	0.8	-0.8	-0.8	-0.8
1-butene	1.4	-1.5	-1.5	0.0
Ethyne	0.0	-2.0	-2.0	0.4
Toluene	-0.9	-0.1	-0.1	-0.2
1,2-ethanediol	-9.3	3.1	2.9	0.5
Cyclopentanol	-5.5	2.0	1.8	1.1
Tetrahydrofuran	-3.5	0.6	0.4	-0.1
Methyl isopropyl ether	-2.0	-0.5	-0.7	1.1
Ethanal	-3.5	-0.5	-1.0	-0.7
Acetone	-3.9	-0.9	-1.4	-1.1
Propanoic acid	-6.5	-0.5	-1.3	1.2
Methyl ethanoate	-3.3	-2.3	-2.9	-0.6
Trimethylamine	-3.2	3.1	3.1	0.0
Pyrrolidine	-5.5	3.6	3.6	-3.0
Pyridine	-4.7	3.3	3.3	-0.3
Hydrazine	-6.3	4.8	4.8	1.3
Acetamide	-9.7	0.3	-1.7	-0.7
Urea	-13.8	2.2	-1.3	-0.9
Propanoic acid (-1)	-78	1	-3	2
2-butanol (-1)	-86	7	-1	11
2-methoxyethanol (-1)	-91	7	2	9
Hydroxide (-1)	-107	2	2	-8
Ethanol (+1)	-86	9	8	11

Dimethyl ether (+1)	-78	7	7	9
t-butylamine (+1)	-65	-3	-3	0
Diethylamine (+1)	-62	-1	-2	2
2-methylaniline (+1)	-68	2	1	3
Azetidine (+1)	-66	-1	-1	-1
Piperidine (+1)	-62	-1	-1	1
Pyrrole (+1)	-60	-7	-7	-5
Benzamide (+1)	-65	9	7	-2
Error Analysis				
RMSE		4	3	4
MUE		3	3	2
MSE		1	0	1

See Table A.1 for format.

Table A.4: Error (in kcal/mol) Analysis of Solvation Free Energies for Test Set 2

		Signed Error		
Solute	ΔG_{exp}	Single Point	Optimization	SM6
Ethane	1.8	-1.6	-1.6	-0.6
Cyclopropane	0.8	-0.7	-0.7	-0.8
1-butene	1.4	-1.4	-1.4	0.0
Ethyne	0.0	-1.7	-1.7	0.4
Toluene	-0.9	0.3	0.3	-0.2
1,2-ethanediol	-9.3	0.1	-0.6	0.5
Cyclopentanol	-5.5	0.4	0.0	1.1
Tetrahydrofuran	-3.5	0.8	0.5	-0.1

Methyl isopropyl ether	-2.0	-0.3	-0.6	1.1
Ethanal	-3.5	-0.8	-2.3	-0.7
Acetone	-3.9	-2.1	-4.2	-1.1
Propanoic acid	-6.5	-3.1	-4.9	1.2
Methyl ethanoate	-3.3	-4.7	-6.3	-0.6
Triethylphosphate	-7.8	-1.9	-4.3	-1.5
Propanoic acid (-1)	-78	2	-1	2
2-butanol (-1)	-86	5	-2	11
2-methoxyethanol (-1)	-91	7	2	9
Hydroxide (-1)	-107	4	3	-8
Ethanol (+1)	-86	9	7	11
Dimethyl ether (+1)	-78	8	7	9
Methyl phosphine (+1)	-66	-11	-13	-1
Trimethyl phosphine (+1)	-57	-4	-5	3
Error Analysis				
RMSE		4	4	5
MUE		3	3	3
MSE		0	-1	2

See Table A.1 for format.

Appendix B: Supporting Information: Supporting Information: QM/MM analysis suggests that Alkaline Phosphatase and Nucleotide pyrophosphatase/phosphodiesterase slightly tighten the transition state for phosphate diester hydrolysis relative to solution

Table B.1: Solvation free energies for the leaving group in different protonation states (in kcal/mol)[a]

Diester	HA[b]	A^{-}[c]	$\Delta\Delta\Delta G^{d}_{solv}$
$MpNPP^{-}$	-10.6 (-16.1/-15.2/-12.4/-19.5)	-60 (-64/-63/-59/-61)	0
$MmNPP^{-}$	-9.6 (-13.1/-12.9/-11.0/-17.1)	-64 (-65/-66/-63/-68)	-5 (-5/-5/-5/-10)
MPP^{-}	-6.6 (-4.3/-3.8/-6.0/-9.9)	-74 (-71/-73/-73/-78)	-18 (-20/-21/-20/-26)

a. Numbers without parenthesis are experimental solvation free energies taken from Ref. [89]; with parentheses are SCC-DFTBPR/SCC-DFTB/SM6/UAKS calculated solvation free energies. The calculations for SM6 and UAKS are at B3LYP/6-31+G(d,p) and B3LYP/6-311++G(d,p) levels, respectively. b. Protonated form of the leaving groups. c. Deprotonated form of leaving groups. d. Difference between solvation free energies of the protonated and deprotonated forms, measured using $MpNPP^{-}$ as the reference.

Table B.2: Average Solvent Accessible Surface Area (in Å ^{2}) for sulfur of $MpNPPS^-$ and its equivalent oxygen of $MpNPP^-$ from R166S and R166S/E322Y AP simulations [a]

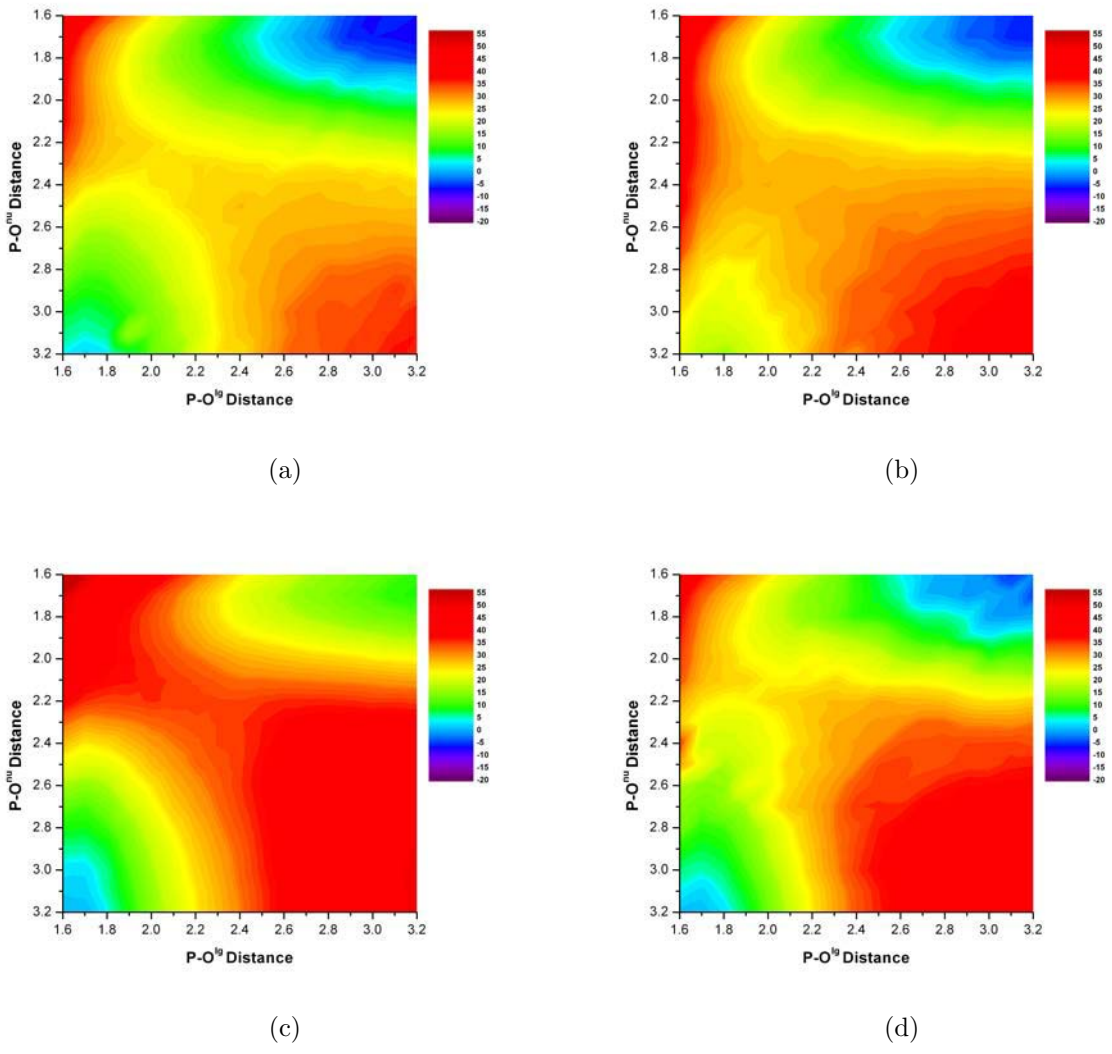
Enzyme + Substrate	α/R_p [b]	β/S_p [b]
R166S AP + $MpNPP^-$	2.4/1.2	7.0/0.0
R166S AP + $MpNPPS^-$	17.7/4.5	36.9/2.1
R166S/E332Y AP + $MpNPP^-$	3.3/1.8	6.2/1.9 (15.8/2.7)[c]
R166S/E332Y AP + $MpNPPS^-$	11.5/6.5	0.9/2.9

a. The results correspond to reactant state/transition state SASA, respectively. b. The R_p and S_p indicate different enantiomers of $MpNPPS^-$, and α/β refer to different orientations of $MpNPP^-$. c. Values in parentheses are from another independent set of simulations.

Table B.3: ^{18}O KIE of $MpNPP^-$ hydrolysis reaction in solution at 95 °C

	Exp[a]	Calc	B3LYP(PCM)[b]	AM1d/MM[b]
O_{lg}	1.0059±0.0005	1.0196	1.0047	1.0044±0.0033
O_{nu}	1.0227±0.0100	1.0408	1.0238	1.0125±0.0054
O_{nb}	0.9949±0.0006	0.9593	0.9977	0.9966±0.0032

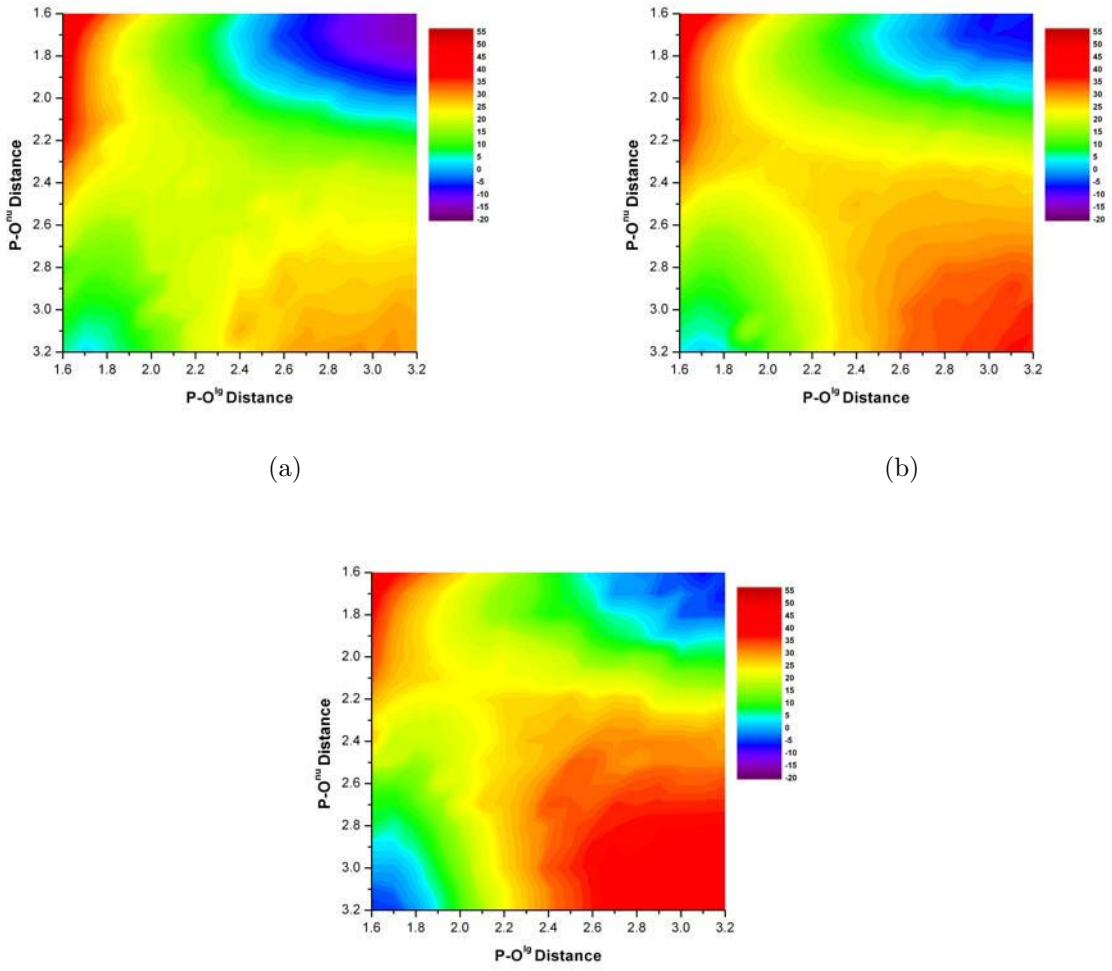
a. Results are taken from ref. [205, 206]. lg: leaving group oxygen; nu: nucleophile oxygen; nb: nonbond oxygen. b. Results are taken from ref. [207].



(a) (b)

(c) (d)

Figure B.1: Adiabatic mapping results for aqueous hydrolysis of phosphate diesters with hydroxide as the nucleophile. Energies are in kcal/mol. (a) $MmNPP^-$ by SCC-DFTBPR/PB; (b) $MmNPP^-$ by including single point gas phase correction at the MP2/6-311++G** level; (c) MPP^- by SCC-DFTBPR/PB; (d) MPP^- by including single point gas phase correction at the MP2/6-311++G** level.

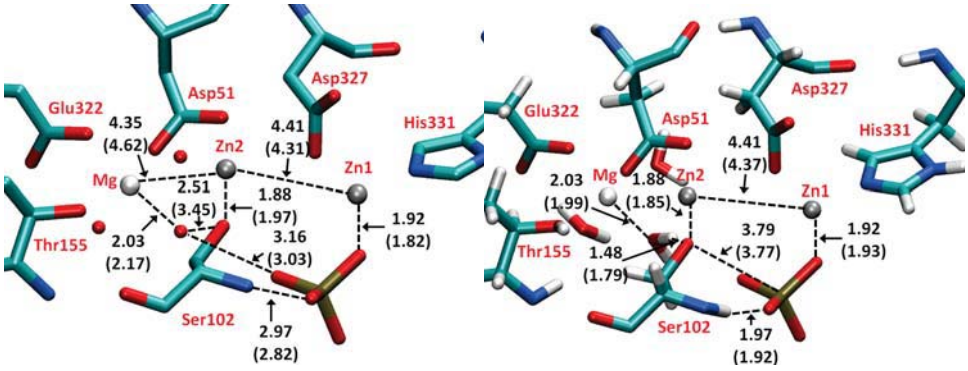


(a)

(b)

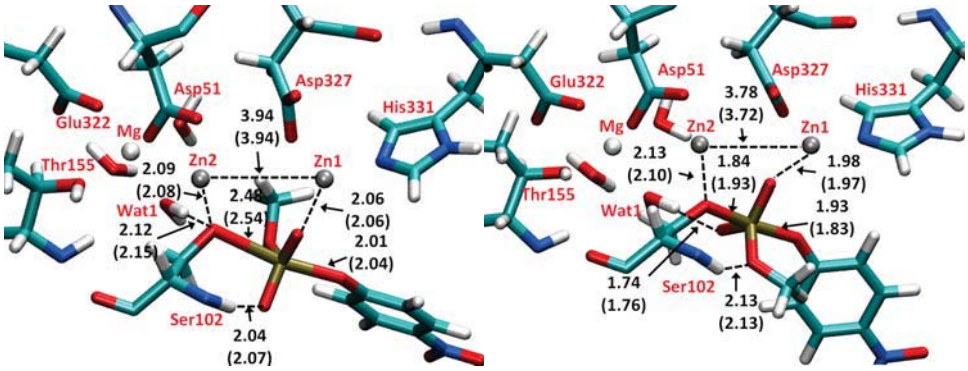
(c)

Figure B.2: Adiabatic mapping results for aqueous hydrolysis of phosphate diesters with hydroxide as the nucleophile. Energies are in kcal/mol. (a) $MpNPP^-$ by including single point gas phase correction at the B3LYP/6-311++G** level; (b) $MmNPP^-$ by including single point gas phase correction at the B3LYP/6-311++G** level; (c) MPP^- by including single point gas phase correction at the B3LYP/6-311++G** level.



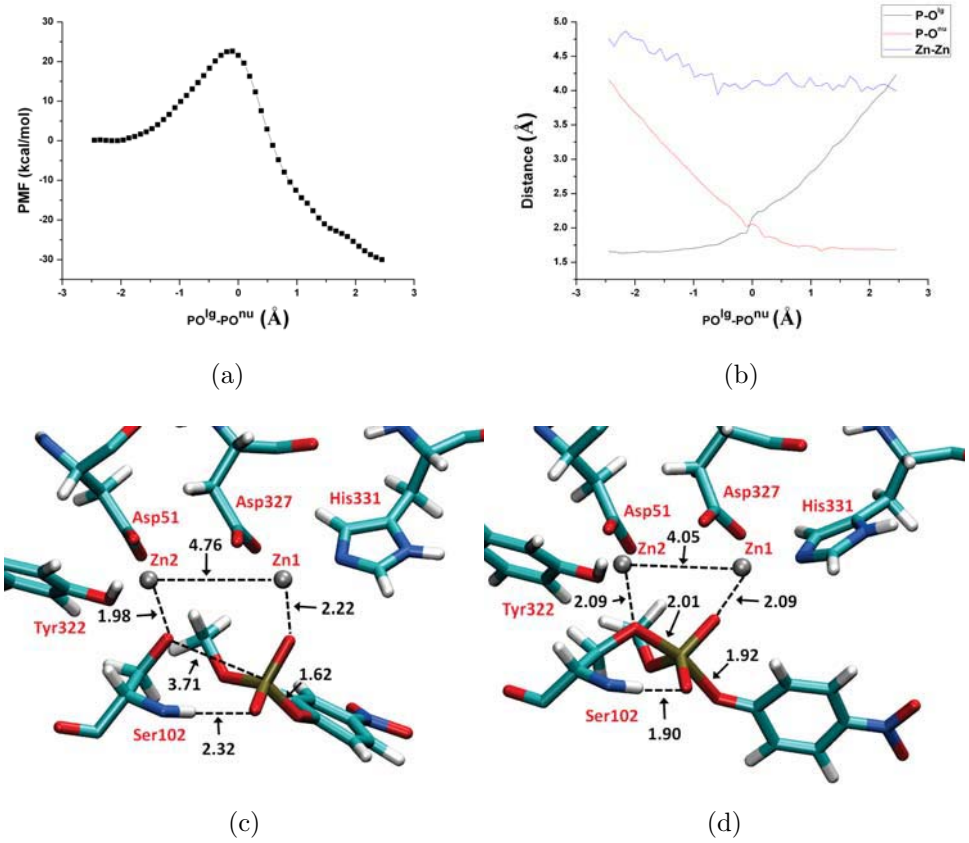
(a) (b)

Figure B.3: Benchmark calculations for an inorganic phosphate (-3 charge) bound to R166S AP with two different QM regions. Key distances are in Å. (a) Structural comparison between crystal structure (with parentheses) and optimized structure (without parentheses) with a large QM region. Hydrogen atoms are omitted. (b) Structural comparison between optimized structure by large (without parentheses) and small (within parentheses) QM region. Asp369, His370 and His412 are omitted for clarity. The smaller QM region, which is used in the main text, includes the two zinc ions and their 6 ligands (Asp51, Asp369, His370, Asp327, His412, His331), Ser102 and $MpNPP^{-}$. Only side chains of protein residues are included in the QM region and link atoms are added between C_α and C_β atoms. The larger QM region further incorporates the entire magnesium site, including Mg^{2+}, sidechains of Thr155, Glu322 and three ligand water molecules.



(a) (b)

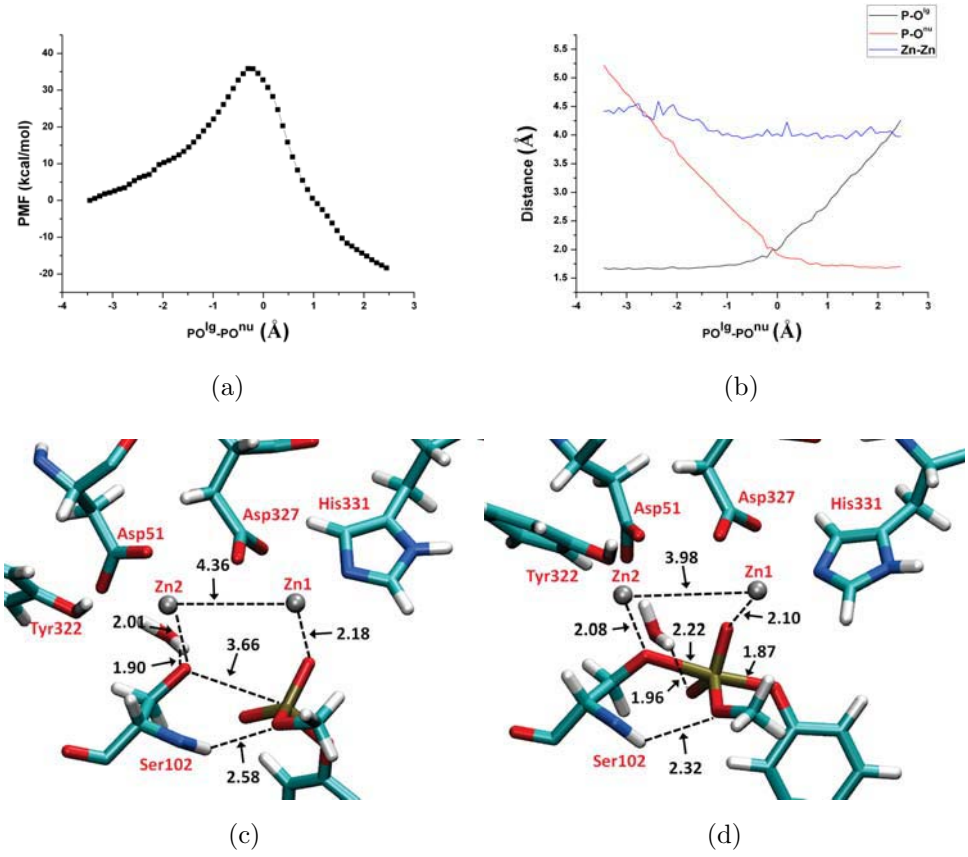
Figure B.4: Comparison of optimized transition state from adiabatic mapping (with parentheses) and CPR (without parentheses) calculations for $MpNPP^-$ in R166S AP with SCC-DFTBPR/MM. Key distances are in Å. (a) The substrate methyl group pointing toward the magnesium ion (the α orientation); (b) the substrate methyl group pointing toward Ser102 backbone (the β orientation). Asp369, His370 and His412 are omitted for clarity.



(a) (b)

(c) (d)

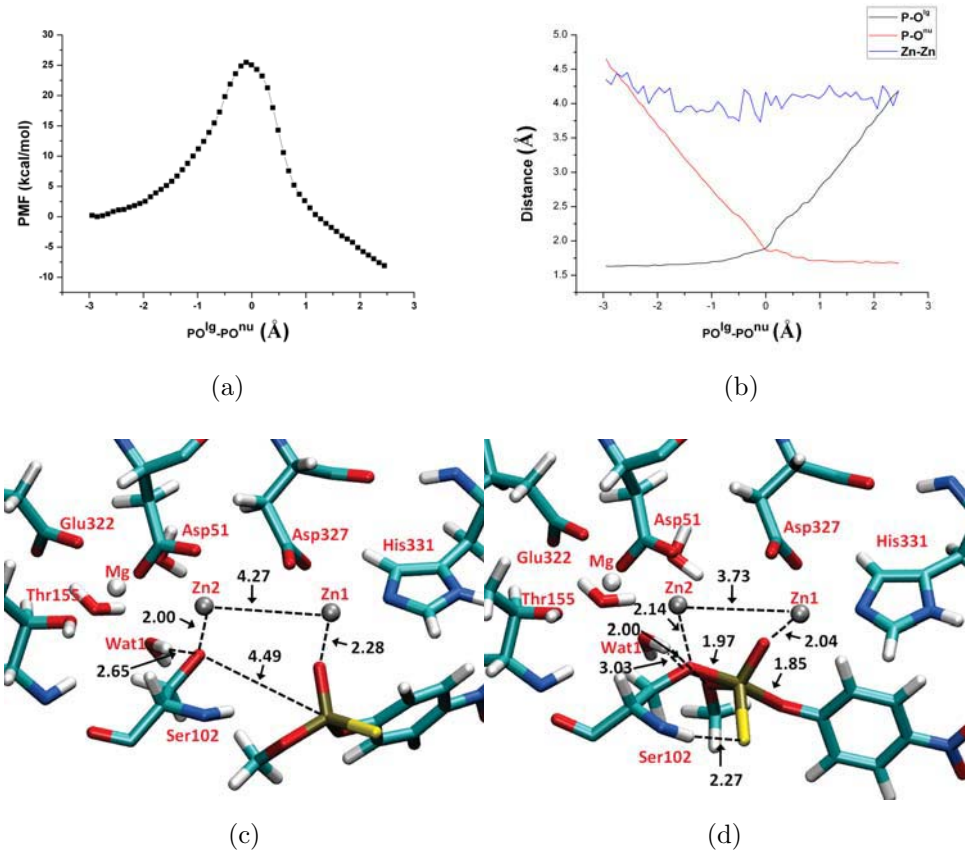
Figure B.5: Potential of Mean Force (PMF) calculation results for MpNPP$^-$ hydrolysis in R166S/E322Y AP with the substrate methyl group pointing toward the original magnesium site (the α orientation). Key distances are in Å and energies are in kcal/mol. (a) PMF along the reaction coordinate (the difference between P-O^{lg} and P-O^{nu}); (b) changes of average key distances along the reaction coordinate; (c) A snapshot for the reactant state, with average key distances labeled. (d) A snapshot for the TS, with average key distances labeled. In (c-d), Asp369, His370 and His412 are omitted for clarity.



(a) (b)

(c) (d)

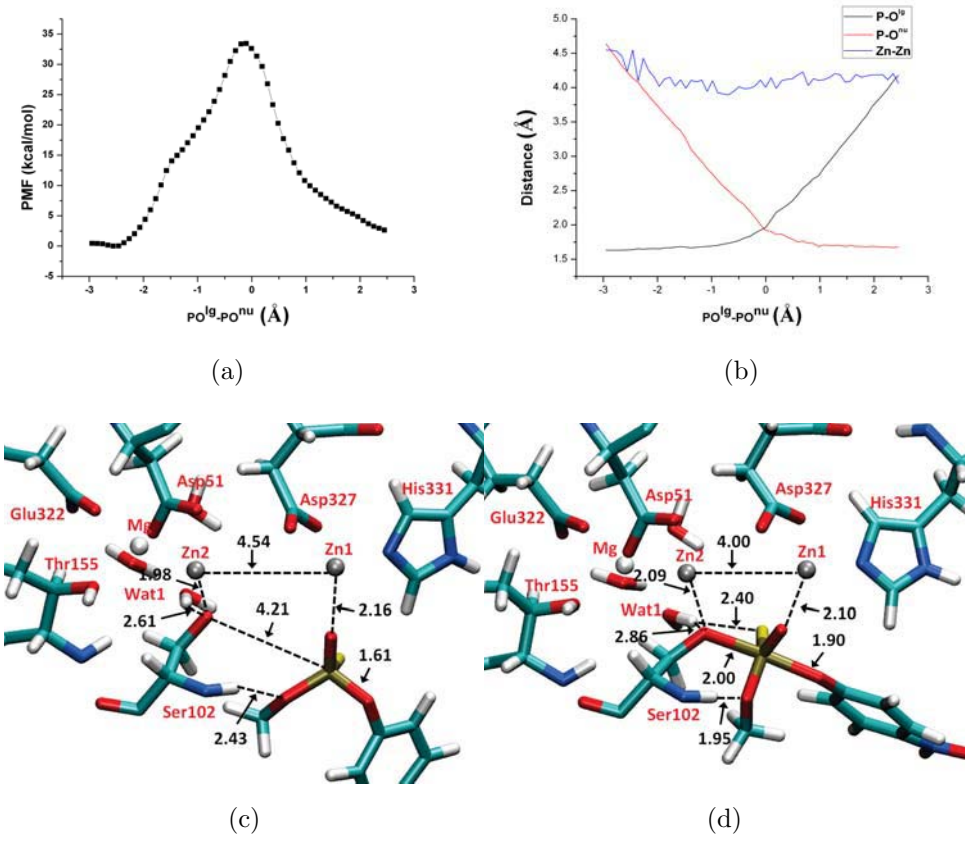
Figure B.6: Potential of Mean Force (PMF) calculation results for $MpNPP^-$ hydrolysis in R166S/E322Y AP with the substrate methyl group pointing toward Ser102 backbone (the β orientation). Other format details follow Fig.B.5.



(a) (b)

(c) (d)

Figure B.7: Potential of Mean Force (PMF) calculation results for R_p-MpNPPS$^-$ hydrolysis in R166S AP; the substrate methyl group points toward the magnesium ion (the α orientation of MpNPP$^-$). Other format details follow Fig.B.5.



(a) (b)

(c) (d)

Figure B.8: Potential of Mean Force (PMF) calculation results for S_p-MpNPPS$^-$ hydrolysis in R166S AP; the substrate methyl group pointing toward Ser102 backbone (the β orientation for MpNPP$^-$). Other format details follow Fig.B.5.

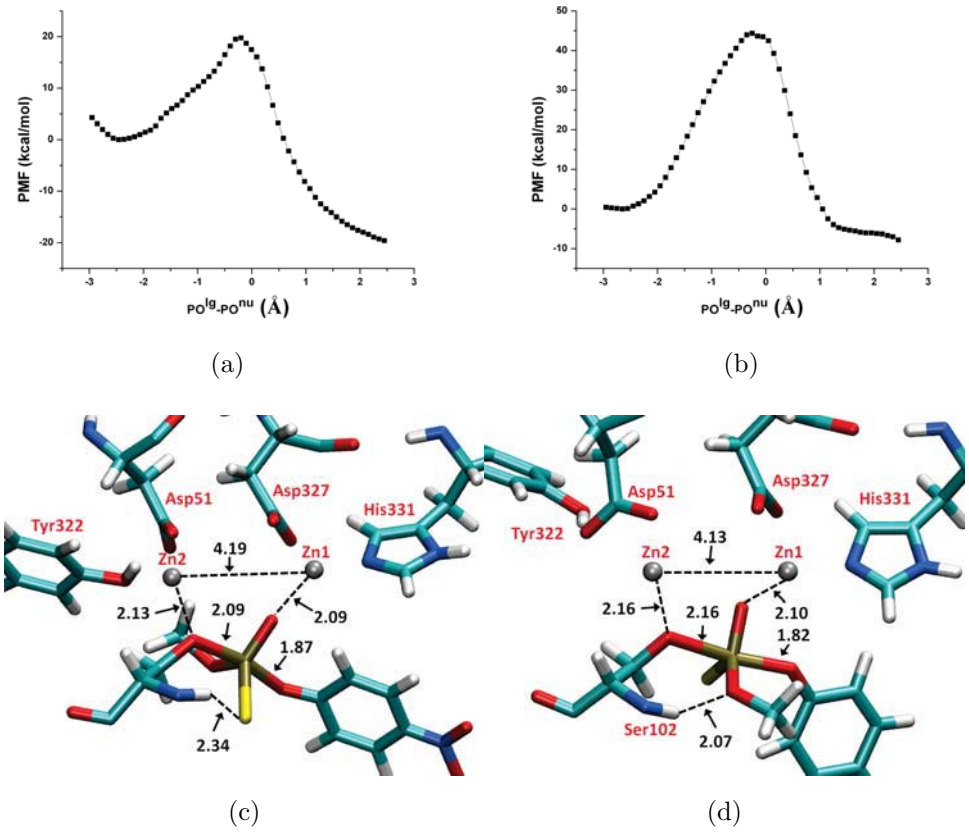
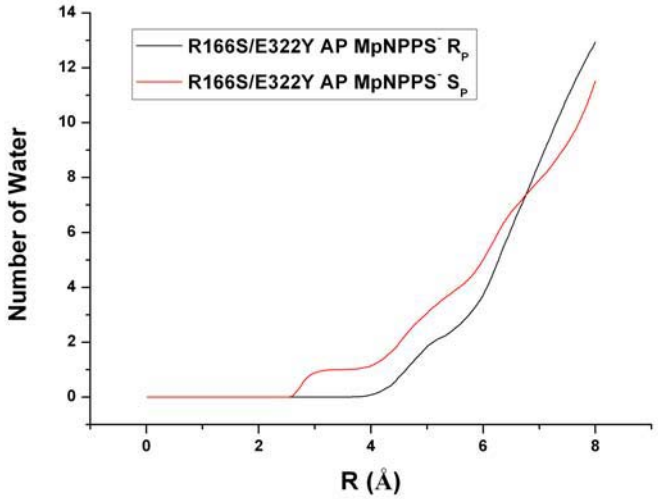
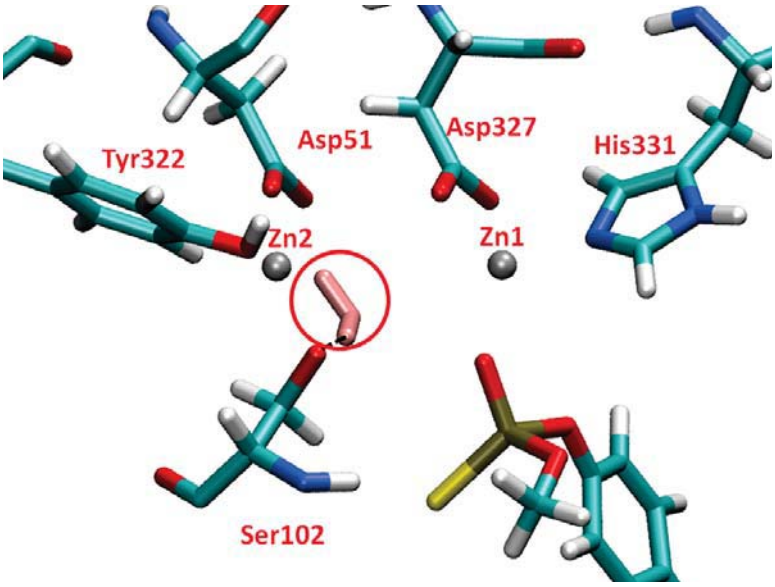


Figure B.9: Potential of Mean Force (PMF) calculation results for MpNPPS$^-$ hydrolysis in R166S/E322Y AP. Key distances are labeled in Å and energies are in kcal/mol. (a) PMF along the reaction coordinate (the difference between P-O^{lg} and P-O^{nu}) for R$_p$-MpNPPS$^-$; (b) PMF for S$_p$-MpNPPS$^-$; (c) A snapshot for the TS of R$_p$-MpNPPS$^-$, with average key distances labeled. (d) A snapshot for the TS of S$_p$-MpNPPS$^-$, with average key distances labeled. In (c-d), Asp369, His370 and His412 are omitted for clarity.



(a)



(b)

Figure B.10: Example of water penetration observed in some double mutant simulations. (a) Comparison of integrated radial distribution of water oxygen around Ser102 nucleophilic oxygen in the reactant state for R_p and S_p MpNPPS$^-$; water penetration is observed only for S_p. (b) A snapshot that illustrates the position of the penetrated water near Ser102; Asp369, His370 and His412 are omitted for clarity.

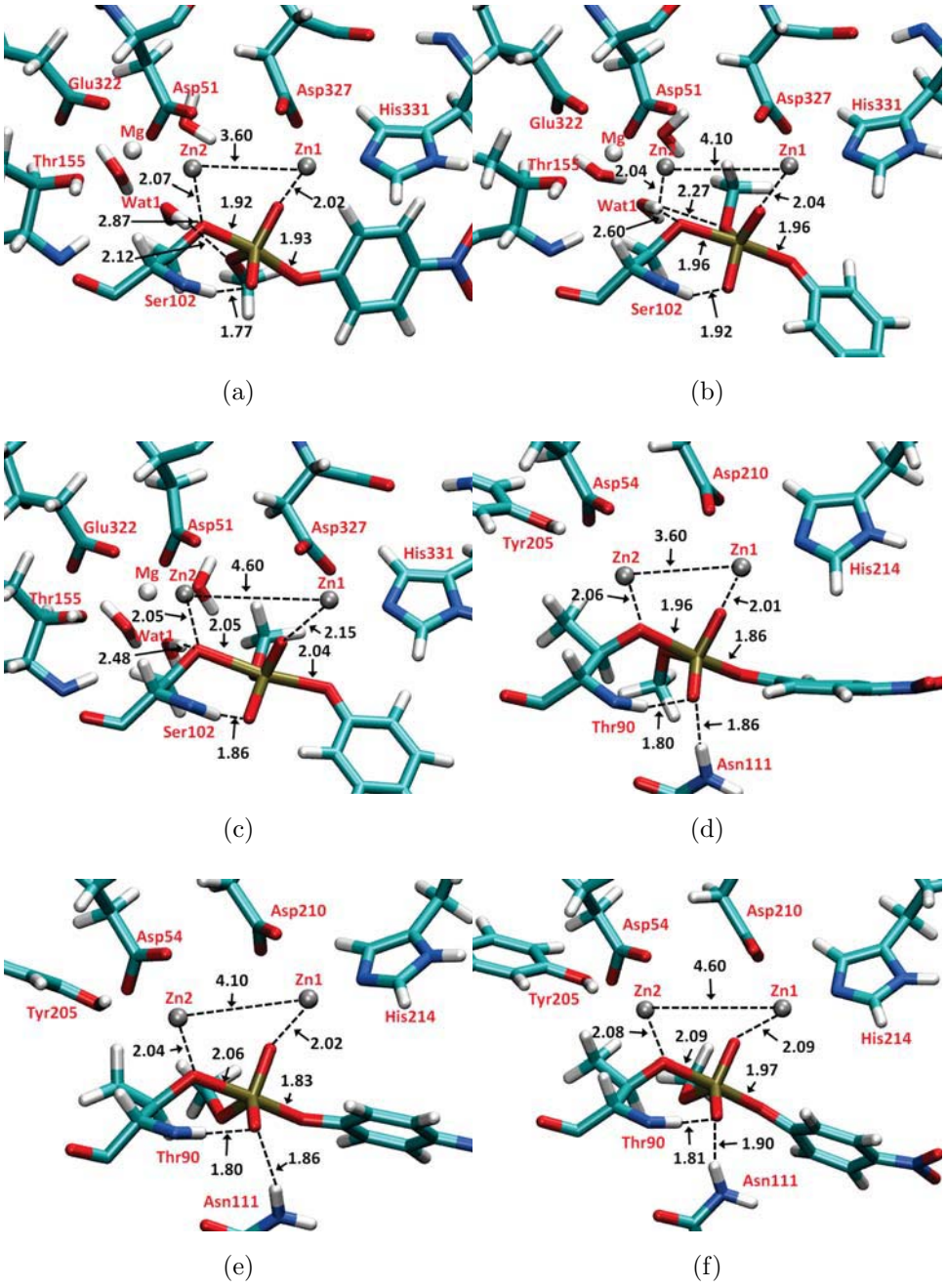
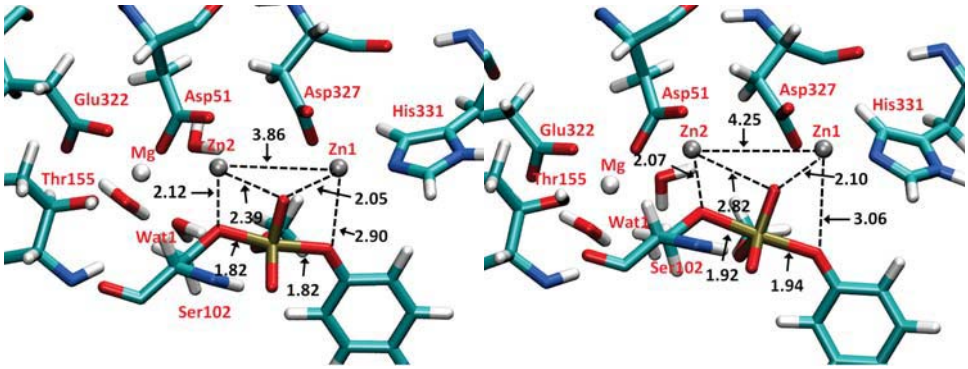


Figure B.11: Snapshots for the TS of $MpNPP^-$ in R166S AP and NPP from simulations in which the zinc-zinc distance is constrained to a specific value; average key distances are labeled in Å. Some nearby residues are omitted for clarity. (a-c) R166S AP with the zinc-zinc distance constrained at 3.6, 4.1 and 4.6 Å; (d-f) NPP with the zinc-zinc distance constrained at 3.6, 4.1 and 4.6 Å.



(a) (b)

Figure B.12: Snapshots for $MpNPP^-$ in R166S AP with α orientation. The reaction coordinate (P-O^{lg}-P-O^{nu}) is constrained at 0.0 Å by a restraint potential similar to the one used in PMF calculations. The initial substrate configuration is constructed similar to the crystal structure of vanadate in wt AP (see below). After optimization, the system is heated to 300 K within 100 ps, followed by a 200 ps production run. (a) The structure after geometry optimization; (b) a snapshot after equilibration run with average distances labeled in Å.

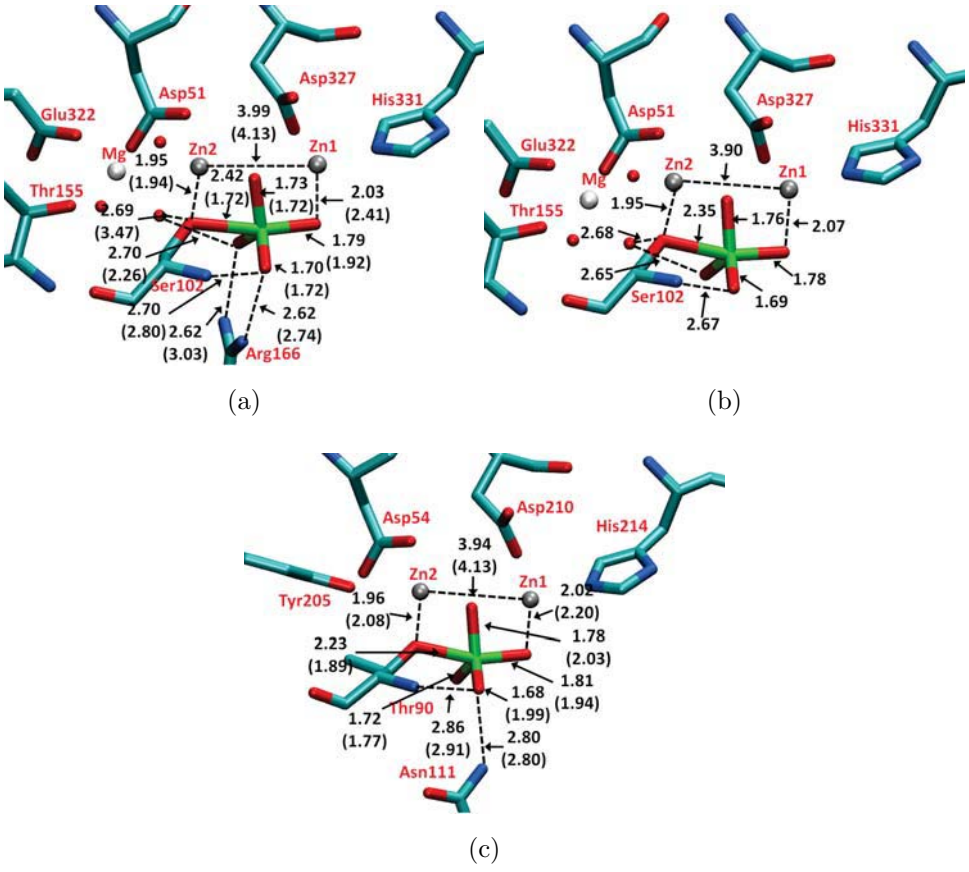
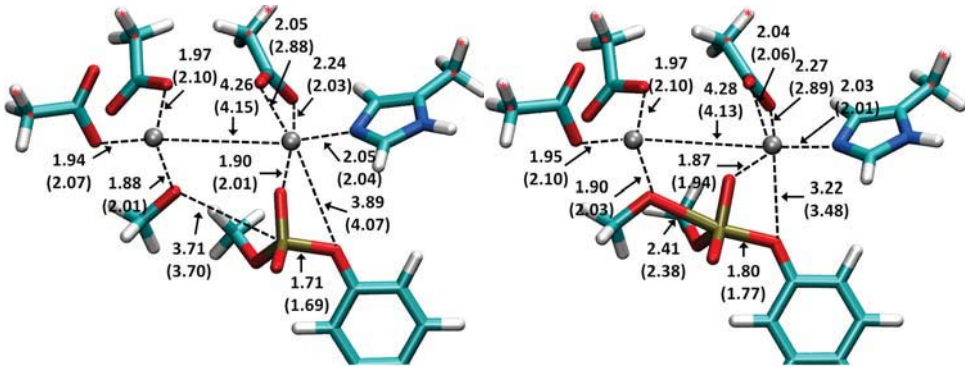


Figure B.13: Optimized structures for vanadate (VO_4^{3-}) in wt AP (a), R166S AP (b) and NPP (c). The numbers withou parenthesis are calculated values by B3LYP/6-31G*; those with parenthesis are values in crystal structures. Hydrogen atoms are omitted for clarity. Distances are in Å.



(a) (b)

Figure B.14: Active site model for $MpNPP^-$ in R166S AP. Atoms labeled by red star are fixed during structural optimization. The numbers without parenthesis are optimized at B3LYP/6-31G* level; those in parenthesis are optimized by SCC. The reaction barrier obtained by B3LYP/6-31+G**//B3LYP/6-31G* and SCC are both 6.7 kcal/mol. Distances are in Å(a) Reactant state; (b) transition state.