Boundary Potentials for Hybrid Quantum Mechanical / Molecular Mechanical Simulations of Solvated Biomolecules

Inaugural-Dissertation

zur
Erlangung des Doktorgrades der
Mathematisch-Naturwissenschaftlichen Fakultät
der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Tobias Benighaus

aus Münster

Mülheim an der Ruhr, September 2010

Gedruckt mit der Genehmigung der Mathematisch-Naturwissenschaftlichen
Fakultät der Heinrich-Heine-Universität Düsseldorf

Referent: Prof. Dr. Walter Thiel
Koreferentin: Prof. Dr. Christel Marian
Tag der mündlichen Prüfung:

Zusammenfassung

Die vorliegende Arbeit befasst sich mit der Entwicklung und Implementierung von Randpotentialen für quantenmechanisch/molekularmechanische (QM/MM) Methoden. Ziel dieser Arbeit ist es, den zweilagigen QM/MM-Ansatz zu einem dreilagigen Modell zu erweitern, um eine zuverlässige Beschreibung langreichweitiger elektrostatischer Wechselwirkungen mit hoher Effizienz zu ermöglichen.
Ausgangspunkt ist ein für klassische Kraftfeldmethoden benutztes Randpotential (GSBP, *generalized solvent boundary potential*), das für semiempirische QM/MM-Verfahren adaptiert und implementiert wird. Die Genauigkeit des GSBP wird an Hand eines Modellsystems (Threonin in Wasser) geprüft und für adäquat befunden. Die Untersuchung der Effizienz des GSBP zeigt, dass es die Rechenzeit ab einer Systemgröße von ca. 12.500 Atomen reduziert. Bei größeren Systemen von ca. 25.000 Atomen, die in QM/MM-Anwendungen üblich sind, verringert das GSBP die Rechenzeit um 70%. Die Anwendung des GSBP erfordert zur Initialisierung umfangreiche Poisson-Boltzmann-Rechnungen, für die drei algorithmische Verbesserungen vorgestellt werden, welche die Rechenzeiten um 60 % reduzieren.
Danach wird ein neues Randpotential (SMBP, *solvated macromolecule boundary potential*) eingeführt, welches im Gegensatz zum GSBP für Geometrieoptimierungen mit Dichtefunktional- oder ab-initio-Methoden entwickelt wurde. Das SMBP weist eine hohe konzeptionelle Ähnlichkeit zum GSBP auf: Die äußere Makromolekülregion wird durch ein elektrostatisches Potentials beschrieben, das durch Lösung der Poisson-Boltzmann-Gleichung bestimmt wird. Die äußeren Solvensmoleküle werden als kontinuierliches Dielektrikum behandelt. Mittels einer selbstkonsistenten Reaktionsfeldprozedur in Kombination mit einer punktladungsbasierten Repräsentation des Randpotentials in den QM-Rechnungen wird eine modulare Implementation erzielt, die die Verwendung des SMBP mit allen QM/MM-Methoden ermöglicht. Das SMBP wird an einem Modellsystem (Glycin in Wasser) und drei enzymatischen Systemen (*p*-Hydroxybenzoat-Hydroxylase, Cytochrom P450cam und Chorismat-Mutase) getested. Die Anwendung des SMBP im

Glycin-Testsystem erweist sich als problematisch, da Geometrieoptimierungen mit dem QM/MM- und dem QM/MM/SMBP-Ansatz zu unterschiedlichen lokalen Minima führen, die trotz großer struktureller Ähnlichkeit zu unterschiedlichen Reaktionsenergien führen. Bei der Anwendung in den enzymatischen Systemen zeigt sich hingegen, dass das SMBP das elektrostatische Potential mit hoher Genauigkeit reproduziert und die Abweichungen bei berechneten Potentialenergiedifferenzen selten oberhalb von 0,3 kcal/mol liegen. Die molekularen und elektronischen Strukturen, die aus QM/MM/SMBP-Geometrieoptimierungen resultieren, können als Ausgangspunkt für Berechnungen freier Energiedifferenzen mittels des QM/MM-*free energy perturbation*-Ansatzes verwendet werden. Aufgrund der konzeptionellen Ähnlichkeit von SMBP und GSBP kann das GSBP unter diesen Umständen beim *sampling* der Konfigurationen benutzt werden. Dadurch sinken die Rechenanforderungen für diesen Schritt um bis zu 90 %.
Langreichweitige elektrostatische Wechselwirkungen in Enzymen haben zwei Quellen: das äußere Makromolekül und das umgebende Solvens. Mit Hilfe von GSBP und SMBP werden die Auswirkungen der beiden Beiträge auf enzymatische Reaktionen getrennt untersucht und quantifiziert. Die Ergebnisse zeigen, dass beide Beiträge die Energetik nur dann deutlich beeinflussen, wenn ein signifikanter Ladungstransfer mit der Reaktion verbunden ist. In solchen Fällen ist jedoch eine genaue Beschreibung beider Effekte für eine zuverlässige Simulation notwendig. GSBP und SMBP bieten diese genaue Beschreibung bei reduziertem Rechenaufwand.

Abstract

This thesis presents the development and implementation of boundary potentials for hybrid quantum mechanical/molecular mechanical (QM/MM) methods. The dual-layer QM/MM method is extended to a three-layer method with the objective of providing an accurate and efficient description of long range electrostatic interactions.
First, a generalized solvent boundary potential (GSBP) originally developed for classical force field simulations is adapted for hybrid QM/MM methods with semiempirical QM Hamiltonians. The GSBP is tested on a model system (threonine in water) and is found to yield accurate results. The efficiency of the GSBP is studied and the breakeven point with standard QM/MM calculations is located at system sizes of around 12,500 atoms. The GSBP reduces the computational costs by 70 % for systems with about 25,000 atoms which are common in QM/MM studies. Since application of the GSBP is connected with a significant overhead, three algorithmic improvements are introduced that reduce the computation time of the overhead by 60 % with only minimal loss of accuracy.
Thereafter, a novel solvated macromolecule boundary potential (SMBP) is introduced which, in contrast to the GSBP, targets geometry optimizations and can be applied with density functional theory or ab initio methods for the QM region. The SMBP is conceptually similar to the GSBP: The outer macromolecule region is represented by a boundary potential obtained from solution of the Poisson-Boltzmann equation; the outer solvent molecules are modeled as a dielectric continuum. A modular implementation that allows application with any QM/MM Hamiltonian is achieved by combining a self consistent reaction field procedure with a point charge-based representation of the boundary potential in the QM calculations. The SMBP is tested on a model system (glycine in water) and three enzymatic systems (*p*-hydroxybenzoate hydroxylase, cytochrome P450cam, and chorismate mutase). In the case of solvated glycine, application of the SMBP turns out to be problematic since QM/MM and QM/MM/SMBP optimizations lead to different local minima with different energetics despite their structural similarity. In the enzymatic

systems, the SMBP reproduces the electrostatic potential with high accuracy and computed potential energy differences rarely deviate by more than 0.3 kcal/mol from the full QM/MM results. Molecular and electronic structures resulting from QM/MM/SMBP geometry optimizations can be used as input for free energy computations following the QM/MM-free energy perturbation scheme. The conceptual similarity of GSBP and SMBP permits application of the GSBP during configurational sampling thereby reducing the computational costs of this step by up to 90%.
Long range electrostatic interactions in enzymes can have two sources: the outer macromolecule and the surrounding solvent. The effect of both contributions on enzymatic reactions is studied by means of SMBP and GSBP. It is found that both contributions influence reaction energetics considerably only if there is significant charge transfer during the reaction. In such cases an accurate description of both contributions is necessary. GSBP and SMBP offer such accuracy at reduced computational costs.

Danksagung

Zuerst möchte ich Prof. Dr. Walter Thiel meine Dankbarkeit dafür ausdrücken, dass er sein profundes Wissen bereitwillig teilte und mir die Möglichkeit gab, in einem spannenden Bereich unter perfekten Bedingungen zu forschen. Die gelungene Balance von Vorgaben und Freiheit erlaubten es mir, eigene Ideen und Interessen zu verfolgen ohne das Ziel aus den Augen zu verlieren.

Ich danke Prof. Dr. Christel Marian dafür, bei dieser Promotion als Koreferentin zu fungieren.

Des Weiteren gilt mein Dank allen Mitgliedern der Theorie-Abteilung des Max-Planck-Instituts für Kohlenforschung, die eine stimulierende, interessante und sehr internationale Stimmung erzeugt haben, die ich in meiner Zeit in Mülheim sehr genossen habe. Mein besonderer Dank gilt Sebastian Metz für die Einführung in "praktische" QM/MM-Rechnungen, Tom Keal, Axel Koslowski und Johannes Kästner für die kompetente Unterstützung bei technischen Problemen mit der ChemShell oder Fortran sowie Horst Lenk für die Antworten auf viele Computerfragen. Besonders danken möchte ich auch Muhannad Altarsha für die sehr fruchtbare und freundschaftliche Kooperation im Cytochrom P450-Projekt sowie Mark Waller und Marko Schreiber für motivierende Diskussionen. Mein Dank gilt ebenfalls Jürgen Breidung, Stephan Thiel, Hans Martin Senn und Ahmet Altun, die detaillierte Informationen über vorhergehende Projekte bereitstellten.

Ich danke dem Fonds der Chemischen Industrie für die Gewährung eines Kekulé-Stipendiums.

Schließlich gilt mein tiefer Dank meinen Eltern und Dani, deren kontinuierliche, unbedingte und liebevolle Unterstützung mir diese Arbeit ermöglichten.

Erklärung

Hiermit erkläre ich, Tobias Benighaus, geboren am 22.12.1981 in Münster, dass ich diese Arbeit eingeständig verfasst und ohne unerlaubte Hilfe angefertigt und diese in der vorgelegten oder ähnlichen Form noch keiner anderen Institution eingereicht habe.

Mülheim an der Ruhr, den 28.09.2010

"So lang die dicke Frau noch singt, ist die Oper nicht zuende."
(Kettcar)

Contents

Chapter 1

Introduction

Understanding enzymatic reactions on an atomistic level is an important goal of biochemical research. In many enzymes, the key step being catalyzed is a single chemical reaction such as proton transfer to the active site, substrate oxidation, dehalogenation, or a similar process. Detailed knowledge of these reaction steps allows researchers to understand, manipulate or copy enzymatic mechanisms. The atomistic insight may be applied to the design of new efficient catalysts that are analogous to active sites of enzymes, the identification of targets for mutagenesis, or the development of new inhibitors.
However, the information about enzymatic processes on an atomistic level that is accessible by experiment is very limited. This stimulated the development of theoretical methods to simulate biomolecular systems. In the past decades theoretical methods have matured to a valuable tool for this purpose. Examples of significant contributions from computational studies range from the first simulation of the bovine pancreatic trypsin inhibitor[1] to simulations of processes that occur on long time scales in complex systems like protein folding pathways.[2]
Due to their size of several thousand atoms, biomolecular systems are normally simulated by empirical molecular mechanics (MM) force fields. In these methods, an empirical potential energy function is constructed using analytical potentials that depend on bond lengths, bond angles, and dihedral angles. In most force fields, electrostatic interactions are modeled as interactions of fixed point charges, and van der Waals interactions are described by the 12-6 Lennard-Jones potential. Through continuous refinement of parameters and models, force fields have developed into reliable methods that describe the structure and dynamics of biomolecules with sufficient accuracy.[3,4] Significant research efforts have been undertaken to go beyond the fixed point charge approximation, and hence the development of polarizable force fields has been an active area of research in the past two decades.[5–7]

Force field methods are by construction unable to describe all chemical events that involve changes in the electronic structure, such as bond breaking and forming or electronic excitations. While quantum mechanical (QM) methods are suitable to model those events, their computational intensity restricts reliable QM methods to systems of at most a few hundred atoms such that they are not directly applicable to study large biomolecular systems.
Already in 1976, Warshel and Levitt envisioned a combination of QM and MM methods that permits a realistic description of bond breaking processes in enzymatic systems. They laid the foundation for the development of hybrid quantum mechanical/molecular mechanical (QM/MM) methods with their seminal work on lysozyme catalysis.[8] In this ansatz, the active region (active site including substrate and cofactors) is modeled by an accurate QM method while the rest of the enzyme and solvent is described by a fast and approximate MM force field. In this way, hybrid QM/MM methods offer the necessary accuracy to study chemical reactions in an enzymatic environment at tolerable computational costs. Despite the convincing ansatz, QM/MM methods found widespread acceptance only much later in the 1990s after a report by Field, Bash, and Karplus on a QM/MM implementation that was thoroughly validated against ab initio and experimental data.[9] Since then QM/MM methods underwent rapid progress and QM methods of increasing complexity were combined with MM force fields. While Warshel and Levitt used the semiempirical modified intermediate neglect of differential overlap (MINDO) method for the QM part,[8] more accurate semiempirical[9] and Hartree-Fock (HF)[10] methods were applied in QM/MM studies in the following years. Nowadays, density functional theory (DFT) methods are the most common choice since they offer the best price/performance ratio. In recent years also very complex and computationally demanding correlated ab initio methods were used in QM/MM approaches to provide more accurate estimates of reaction and activation energies in enzymes.[11–13]
Due to the considerable computational expense of the QM part, QM/MM methods were mainly applied to study the minimum energy path of enzymatic reactions on the potential energy surface. Dynamic effects were neglected although entropic contributions can be significant since many biomolecular systems are flexible. With the steady increase in computational resources and the development of new methods to compute free energies that are specifically designed for QM/MM, the application of QM/MM Hamiltonians to compute free energy differences has gained popularity.[14–24]
In the context of this development, the treatment of long range electrostatic interactions in QM/MM simulations attracted considerable attention since these interactions can have a significant influence on the relative energies of different configurations. Therefore, an accurate description of these inter-

actions is indispensable for meaningful computations of properties that require extensive configurational sampling, such as free energy differences.[25–28] While the development of efficient and accurate methods to treat these interactions has been an active area of research in the field of classical simulations for a long time, these techniques have only recently been adapted to QM/MM methods due to the technical difficulties introduced by the QM atoms.

For the case of periodic boundary conditions (PBC), Ewald summation is an established method to compute the electrostatic energy and forces of an infinite periodic array of systems without significant truncations.[29–31] Therefore, the applicability of the Ewald summation method has been extended to hybrid QM/MM simulations with semiempirical[32–34] and DFT[35–37] QM Hamiltonians. Unfortunately, application of these methods to large non-periodic biomolecules is affected by serious problems. The artificially imposed periodicity may lead to significant artifacts[38–40] and even qualitatively wrong results unless the molecule is solvated in a solvent box of adequate size.[41] Thus, the number of solvent molecules is necessarily large and increases the computational costs massively, such that Ewald summation can only be used for small to medium sized biomolecules.

Very often, however, one is interested in simulating the behavior of a large biomolecule in infinite dilution, and alternative approaches were devised to facilitate these computations. For biochemical reactions that proceed in a localized region of the macromolecule, boundary potentials are an especially attractive approach.[42–53] Within this approach, the system is subdivided into an inner region, containing the active site and the adjacent part of the enzyme, and an outer region, containing the rest of the enzyme and the outer solvent molecules. While the inner region is simulated atomistically, the effect of the outer region onto the inner region is mimicked by the boundary potential. Ideally, the boundary potential is designed such that the statistical properties of the inner region interacting with the boundary potential are the same as those of the full solvated macromolecule. Although this may be formulated rigorously as an integration over the outer region degrees of freedom,[53] an efficient implementation necessitates the introduction of further approximations.

In the generalized solvent boundary potential (GSBP), developed by Im *et al.* in 2001, the outer region solvent molecules are described by a polarizable dielectric continuum (PDC) and the outer region charge distribution by fixed point charges.[54] Electrostatic interactions with the outer region are separated into a solvent-shielded static field created by the outer region point charges interacting with the PDC, and a dynamic reaction field induced by interaction of the inner region charge distribution with the PDC. The great advantage of the GSBP is the analytical expression for the dy-

namic reaction field response which can handle irregularly shaped macromolecule/solvent boundaries. Accuracy and efficiency of the GSBP in classical simulations were validated by studies on aspartyl-tRNA synthetase[54,55] and the KcsA potassium channel.[54] In 2005, Cui and co-workers adapted the GSBP method to the QM/MM framework as a means to treat long range electrostatic interactions in QM/MM simulations accurately and to describe QM/MM and MM/MM interactions in a balanced way.[56] Here, the self--consistent-charge density-functional tight-binding (SCC-DFTB)[57] method was chosen as the QM Hamiltonian. The SCC-DFTB/MM/GSBP approach was evaluated by comparison to results from Ewald/PBC calculations on small model systems. It was found to provide quantitatively very similar results at significantly lower computational costs compared to Ewald/PBC methods.[58–60] The fixation of the outer region atoms is a fundamental assumption in the GSBP that allows for a closed-form expression for the electrostatics.[54] While this assumption is valid in many cases, the use of the GSBP was found to be problematic if the macromolecule underwent major conformational changes during the course of a reaction.[60] For the investigation of localized processes in large macromolecules, however, the SCC-DFTB/MM/GSBP approach proved to be an efficient and accurate method, and was applied subsequently to study several biological systems.[61–64]
These encouraging results provided the motivation to advance the development of boundary potentials in hybrid QM/MM methods. In this thesis, the GSBP was first adapted and implemented for QM/MM Hamiltonians with semiempirical QM methods. Thereafter, a novel boundary potential on the basis of the GSBP was developed which can be used in combination with any QM/MM Hamiltonian and is efficient in geometry optimizations. Both approaches extend QM/MM to a three-layer method. They were validated on test systems ranging from small molecules to enzymatic systems.
This thesis is structured as follows: In chapter 2, the theoretical background of the underlying methods is reviewed to provide the foundation for the development of the boundary potentials. The implementation of the GSBP for semiempirical QM/MM Hamiltonians is described in chapter 3. In the following chapter, the new solvated macromolecule boundary potential (SMBP) is introduced. Both boundary potentials are applied to study the effect of long range electrostatics and bulk solvent on two real-life enzymatic systems in chapter 5. Moreover, a protocol to determine optimal values for the inherent parameters of GSBP and SMBP is presented in this chapter. Finally, a short conclusion is given in chapter 6.

Chapter 2

Theoretical Background

2.1 The QM/MM Approach

The QM/MM approach is based on the fundamental concept that only a small subset of atoms of an enzyme directly participate in the chemical event that is catalyzed. Based on this idea, it seems sensible to use the most accurate and computationally intense methods only for this subset of atoms and apply a more approximate approach for the remainder of the system.

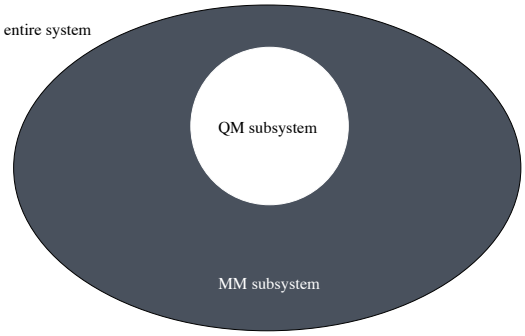


Figure 2.1: QM/MM system partitioning.

In the QM/MM approach, the system is subdivided into a QM region and an MM region (as illustrated in figure 2.1). All atoms in the QM region are modeled by an accurate QM method capable of describing chemical events that involve changes of the electronic structure. Atoms which are not directly involved in the reaction constitute the MM region and are modeled with a force field. The QM and MM regions are frequently labeled "inner" and "outer" region in the QM/MM literature. In this thesis, a different nomen-

clature is adopted since the terms "inner" and "outer" region are used with a different meaning in the context of boundary potentials (see section 2.2).

2.1.1 Additive and Subtractive Scheme

Combination of QM and MM approaches permits the construction of a potential energy function to study chemical events in enzymatic systems at affordable computational costs. Two fundamentally different ways of combining QM and MM methods have emerged.
In the subtractive scheme, the MM energy for the full system and the QM energy of the QM region are added. To avoid double counting, the energy of the QM region computed with the MM method is subtracted.

$$E^{sub}_{QM/MM} = E_{MM}(r_{QM}, r_{MM}) + E_{QM}(r_{QM}) - E_{MM}(r_{QM}) \tag{2.1}$$

Here, E_{MM}, E_{QM}, r_{MM}, r_{QM} denote the MM and QM energies and coordinates, respectively. By construction, all interactions between QM and MM region are described at the MM level. Electrostatic QM-MM interactions are then usually computed as Coulombic interactions of fixed point charges (corresponding to mechanical embedding in section 2.1.2). This is problematic since the electrostatic interactions can be very important, and fixed point charges rarely provide an adequate representation of the QM charge density. Simplicity and ease of implementation are the main advantages of the subtractive scheme. QM and MM programs can be used without any modifications.
In the alternative additive scheme, the energies of the MM and the QM region are computed separately with the respective methods. An extra term that describes only the interactions between the two regions (E_{QM-MM}) is added.

$$E^{add}_{QM/MM} = E_{QM}(r_{QM}) + E_{MM}(r_{MM}) + E_{QM-MM}(r_{QM}, r_{MM}) \tag{2.2}$$

2.1.2 QM-MM Interactions

The treatment of the QM-MM interactions defines a specific QM/MM method within the additive scheme. The QM-MM coupling term can be separated into three individual contributions since QM and MM region interact via covalent bonds (E^{bond}_{QM-MM}) as well as nonbonded electrostatic (E^{elec}_{QM-MM}) and van der Waals (E^{vdW}_{QM-MM}) interactions.

$$E_{QM-MM} = E^{bond}_{QM-MM} + E^{elec}_{QM-MM} + E^{vdW}_{QM-MM} \tag{2.3}$$

Bonded interactions occur at the QM-MM boundary between QM and MM atoms that are connected by a covalent bond. In QM/MM, such bonds are described at the MM level by a harmonic potential so that bond parameters have to be defined for this bond type. Nonbonded van der Waals interactions are normally also described at the MM level by a Lennard-Jones potential:

$$E_{QM-MM}^{vdW} = \sum_{A \in QM} \sum_{B \in MM} \epsilon_{AB} \left[\left(\frac{\sigma_{AB}}{r_{AB}} \right)^{12} - \left(\frac{\sigma_{AB}}{r_{AB}} \right)^{6} \right], \qquad (2.4)$$

where ϵ_{AB} and σ_{AB} are the van der Waals parameters of atom pair AB and r_{AB} is its distance. Although this may appear to be a strong approximation, several factors advocate this approach. First, in the vast majority of systems, the QM-MM van der Waals interactions will not contribute significantly to energy differences that are considered when studying a biochemical process.[65] Moreover, the accuracy that is provided by using the Lennard-Jones potential in combination with a set of empirical parameters is sufficient, even in the context of self consistent field (SCF) QM methods which do not treat electron correlation explicitly. Hence, similar force field terms have even been introduced into DFT methods to improve the description of van der Waals interactions.[66]
The outer (MM) region of the enzyme induces an electrostatic potential which in many systems has an impact on reaction energetics, *e.g.*, by stabilizing the transition state. Therefore, the electrostatic interactions have to be described accurately. A well-defined hierarchy of electrostatic coupling methods has emerged that is defined by the extent of mutual polarization:

- Mechanical embedding: Electrostatic QM-MM interactions are described at the MM level. Usually, Coulombic interactions of fixed point charges are applied. Therefore, charges for the QM atoms have to be found that represent the QM charge density with acceptable accuracy during the whole course of the reaction. This might be problematic or even impossible since the electronic structure of the QM region may change significantly during the reaction while the point charges are constant. Moreover, the QM region does not experience the electrostatic potential that is induced by the MM charges. In many enzymes, however, this electrostatic potential is vital for the function of the enzyme and its neglect can bias the results significantly.

- Electronic embedding: Electronic embedding (also called electrostatic embedding) remedies most of the shortcomings of mechanical embedding. The QM density is embedded into the electrostatic potential of the MM region which in practice translates into performing the QM

calculation in the presence of the MM charges. The one-electron part of the QM Hamiltonian is augmented by an additional term:

$$E_{QM-MM}^{elec} = -\sum_{i}^{electrons} \sum_{B \in MM} \frac{q_B}{r_{iB}} + \sum_{A \in QM} \sum_{B \in MM} \frac{Z_A q_B}{r_{AB}} \tag{2.5}$$

Here, q_B is the charge of MM atom B, r_{iB} is the distance between electron i and MM atom B, and Z_A is the nuclear charge of QM atom A. Using electronic embedding, the QM density can adapt to the electrostatic potential of the MM charges. However, computational costs increase significantly, especially for computation of the gradient due to the contributions of the numerous MM atoms.

- Polarized embedding without self consistency: In the next step, electronic embedding is combined with a polarizable MM charge model. The QM charges polarize the MM charges but the effect of MM polarization is neglected in the QM calculation.

- Polarized embedding with self consistency: This model offers the highest degree of mutual polarization. The potential of the polarizable MM charges is included in the QM Hamiltonian. QM wave function and MM charges are optimized self consistently.

Electronic embedding has become the most popular embedding scheme and has been applied successfully in numerous QM/MM studies.[24] However, one aspect has to be considered when using electronic embedding: Charge transfer across the QM-MM boundary is impossible. Therefore, the QM region has to be of adequate size to accommodate fluctuations of the charge density in the chemically relevant region. QM/MM calculations should be validated by comparing results obtained with QM regions of different size. Throughout this thesis, the additive scheme in combination with electronic embedding is used.

2.1.3 QM-MM Boundary

If the QM-MM boundary cuts through a covalent bond, a free valency is created in the QM subsystem. Several boundary schemes have been developed to saturate the free valency that may be categorized as link atom,[9,10] boundary atom,[67–72] or localized orbital schemes.[8,73–75] If the boundary cuts through non-polar single bond at a reasonable distance to the reactive region, the accuracy is not sensitive with respect to the boundary scheme.[24] In this work, hydrogen link atoms are employed, *i.e.*, an additional hydrogen

atom (L) with adequate basis functions is placed on the axis defined by the atoms Q^1 and M^1 (see figure 2.2 for labeling conventions). Forces acting on the link atom are projected onto Q^1 and M^1 to remove the additional degrees of freedom from the system.

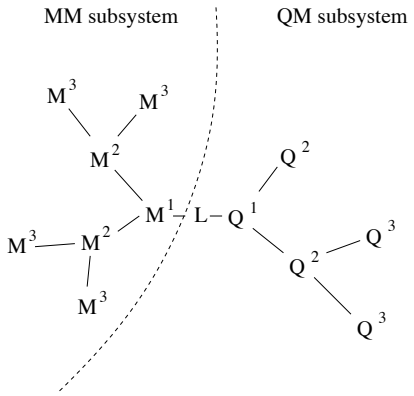


Figure 2.2: Definition of atom labels at the QM-MM boundary. QM and MM atoms that are covalently bonded to each other are labeled Q^1 and M^1, respectively. QM atoms separated from Q^1 by one or two bonds are labeled Q^2 and Q^3, respectively. The same rules apply for MM atoms. L is a link atom.

An artificial proximity of the QM density and the closest MM charges results from the introduction and positioning of the link atom. Unless measures are taken, the M^1 charges will overpolarize the QM density and affect the electronic structure. In this work, the charge-shift scheme is employed throughout to avoid overpolarization. This scheme distributes the charge of the M^1 atom equally over all M^2 atoms. The change in dipole moment that results from shifting the M^1 charges is compensated by sets of two point charges that are placed in the vicinity of each M^2 atom so that the change in dipole moment is neutralized.[76]

2.2 Generalized Solvent Boundary Potential

In this subsection, the GSBP is introduced for classical simulations, *i. e.*, for those simulations employing a force field method for the whole system. The terms "inner" and "outer" region are not to be mistaken for the QM and MM region, respectively. Instead, they have a different meaning. Nomenclature, differences, and common aspects of the boundary potential and the QM/MM approach are discussed in more detail in section 3.1. Later, when the GSBP

is introduced for QM/MM Hamiltonians, the "inner" region is further subdivided into QM and MM region.
The GSBP is derived for a macromolecule R surrounded by N solvent molecules. The system is subdivided into an inner region that contains the inner part of the macromolecule (R_i) and the n inner solvent molecules, and an outer region that contains the outer part of the macromolecule (R_o) and the $N-n$ outer solvent molecules. Statistical expectation values depending only on the degrees of freedom of the inner region ($\mathbf{R}_i, \mathbf{1}, ..., \mathbf{n}$) can be calculated by integrating out the outer region contributions. The influence of the outer region on the inner region can be described rigorously by means of the potential of mean force (PMF) $W(\mathbf{R}_i, \mathbf{1}, ..., \mathbf{n})$.

$$e^{-\beta W(\mathbf{R}_i,1,..n)} = \frac{1}{C}\int' d\mathbf{R}_o d(\mathbf{n}+\mathbf{1}) \cdots d\mathbf{N} e^{-\beta U(\mathbf{R},\mathbf{1},...,\mathbf{N})} \tag{2.6}$$

Here, C denotes an arbitrary integration constant, U is the potential energy, and the primed integral indicates integration over the degrees of freedom of the outer region ($\mathbf{R}_o$, $\mathbf{n}+\mathbf{1}$, ..., $\mathbf{N}$) including only those configurations with all outer region atoms outside the inner region. Beglov and Roux demonstrated that the PMF is related to the reversible thermodynamic work necessary to assemble the inner region.[53]

$$\begin{aligned} W(\mathbf{R}_i, \mathbf{1}, ..., \mathbf{n}) &= U(\mathbf{R}_i, \mathbf{1}, ..., \mathbf{n}) + \Delta W_{cr} + \Delta W_{np}(\mathbf{R}_i, \mathbf{1}, ..., \mathbf{n}) \\ &\quad + \Delta W_{elec}(\mathbf{R}_i, \mathbf{1}, ..., \mathbf{n}) \end{aligned} \tag{2.7}$$

The contributions to the PMF that arise from the configurational restrictions, the non-polar and the electrostatic interactions are denoted ΔW_{cr}, ΔW_{np}, and ΔW_{elec}, respectively. U is the potential energy of the isolated inner region that includes bonded and non-bonded (van der Waals and electrostatic) terms.
The goal of the GSBP is to provide an efficient and accurate approximation to the electrostatic contribution to the PMF. Therefore, the outer region solvent molecules are described by a PDC and the outer region macromolecule by fixed point charges. The electrostatic contributions to the PMF now consist of the direct Coulombic interactions of inner and outer region (U^{io}_{elec}), and the solvation free energy resulting from interaction with the PDC (ΔW^{solv}_{elec}).

$$\Delta W_{elec} = U^{io}_{elec} + \Delta W^{solv}_{elec} \tag{2.8}$$

Representing the charge distribution of the outer macromolecule and the inner region by point charges q_A, the electrostatic solvation free energy can be calculated as

$$\Delta W^{solv}_{elec} = \frac{1}{2}\sum_A q_A \phi_{rf}(\mathbf{r}_A), \tag{2.9}$$

where the reaction field potential $\phi_{rf}(\mathbf{r})$ is the difference of a reference electrostatic potential computed in vacuum, $\phi_v(\mathbf{r})$, and the electrostatic potential computed in solution, $\phi_s(\mathbf{r})$. The electrostatic potentials are obtained by solving the linearized Poisson-Boltzmann (PB) equation

$$\nabla\left[\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})\right] - \bar{\kappa}^2(\mathbf{r})\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}), \tag{2.10}$$

with the charge density of all explicit atoms $\rho(\mathbf{r})$, the space-dependent dielectric constant $\epsilon(\mathbf{r})$, and the modified Debye-Hückel screening factor $\bar{\kappa}(\mathbf{r})$ (see section 2.3).[77] The solvation free energy term is problematic, since during sampling of the inner region configurations the PB equation would have to be solved for each configuration which is prohibitively expensive. To isolate the dynamic properties, the charge distribution is separated into an inner and outer part.

$$\rho(\mathbf{r}) = \rho_i(\mathbf{r}) + \rho_o(\mathbf{r}) \tag{2.11}$$

In consequence, the electrostatic solvation free energy splits up into three terms: outer-outer, inner-outer, and inner-inner contributions.

$$\Delta W_{elec}^{solv} = \Delta W_{elec}^{oo} + \Delta W_{elec}^{io} + \Delta W_{elec}^{ii} \tag{2.12}$$

The first term, ΔW_{elec}^{oo}, stems from the interaction of the outer region charge distribution with the self-induced reaction field, and is constant throughout sampling. The inner-outer contribution arises from the interaction of the inner region charge distribution with the reaction field that is induced by the outer region charge distribution. Calculation of the Coulombic interaction of the inner and outer region can be combined very efficiently with the calculation of the inner-outer contribution to the solvation free energy.

$$\begin{aligned}
\Delta W_{elec}^{io} + U_{elec}^{io} &= \sum_{A\in inner} q_A\phi_{rf}^o(\mathbf{r}_A) + U_{elec}^{io} \\
&= \sum_{A\in inner} q_A\phi_s^o(\mathbf{r}_A) - \sum_{A\in inner} q_A\phi_v^o(\mathbf{r}_A) + U_{elec}^{io} \\
&= \sum_{A\in inner} q_A\phi_s^o(\mathbf{r}_A)
\end{aligned} \tag{2.13}$$

Interaction of the outer electrostatic potential in vacuum, ϕ_v^o, with the inner region charges corresponds to the inner-outer electrostatic contribution to the potential energy, U_{elec}^{io}. Hence, both inner-outer interactions can be evaluated in one step.

The outer region being fixed, the electrostatic potential of the outer region charges in solution, $\phi_s^o(\mathbf{r})$, has to be computed only once and is valid for all inner region configurations. As the interaction with all outer region point

charges and solvent molecules is substituted by an interaction with a static potential, computational costs are reduced massively. However, computation of the inner-inner contribution,

$$\Delta W_{elec}^{ii} = \frac{1}{2} \sum_{A \in inner} q_A \phi_{rf}^i(\mathbf{r}_A), \tag{2.14}$$

remains demanding because $\phi_{rf}^i(\mathbf{r})$ depends on the inner region configuration. To circumvent repeated solution of the PB equation, the configuration-independent reaction field Green's function is introduced

$$\phi_{rf}^i(\mathbf{r}) = \int d\mathbf{r}' \rho_i(\mathbf{r}') G_{rf}(\mathbf{r}, \mathbf{r}'). \tag{2.15}$$

Since G_{rf} is a six-dimensional quantity, its direct computation and storage is not feasible. Therefore, G_{rf} has to be projected onto a normalized basis set $\{b_m(\mathbf{r})\}$ that is used to express the charge distribution of the inner region.

$$\rho_i(\mathbf{r}) = \sum_m c_m b_m(\mathbf{r}) \tag{2.16}$$

The coefficients c_m are calculated as

$$c_m = \sum_n S_{mn}^{-1} Q_n, \tag{2.17}$$

where S_{mn} is the overlap matrix and Q_n are the generalized multipole moments

$$Q_n = \sum_A^{\in inner} q_A b_n(\mathbf{r}_A). \tag{2.18}$$

If the inner region has a spherical shape, an orthonormal set of basis functions can be constructed using spherical harmonics. The basis function of the multipole moment with order l and component m is defined as:

$$b_{lm}(\mathbf{r}) = \left[\frac{(2l+3)}{R^{2l+3}} \right]^{1/2} r^l Y_{lm}(\theta, \phi) \tag{2.19}$$

Here, r, θ, and ϕ define the point $\mathbf{r}$ in spherical coordinates, and R is the radius of the spherical inner region. Y_{lm} is the spherical harmonics defined as

$$Y_{lm}(\theta, \phi) = (-1)^m \left[\frac{(2l+1)}{4\pi} \frac{(l-m)!}{(l+m)!} \right]^{1/2} P_l^m(\cos\theta) e^{im\phi}, \tag{2.20}$$

using the associated Legendre polynomial P_l^m.
In the basis set representation, ΔW_{elec}^{ii} becomes

$$\begin{aligned}
\Delta W_{elec}^{ii} &= \frac{1}{2}\int d\mathbf{r}d\mathbf{r}' \left[\sum_{im} S_{im}^{-1} Q_m b_i(\mathbf{r})\right] G_{rf}(\mathbf{r},\mathbf{r}') \left[\sum_{jn} S_{jn}^{-1} Q_n b_j(\mathbf{r}')\right] \\
&= \frac{1}{2}\sum_{im}\sum_{jn} S_{im}^{-1} Q_m \left[\int d\mathbf{r}d\mathbf{r}' b_i(\mathbf{r}) G_{rf}(\mathbf{r},\mathbf{r}') b_j(\mathbf{r}')\right] S_{jn}^{-1} Q_n \\
&= \frac{1}{2}\sum_{im}\sum_{jn} S_{im}^{-1} Q_m M_{mn}^{*} S_{jn}^{-1} Q_n \\
&= \frac{1}{2}\sum_{mn} Q_m M_{mn} Q_n \qquad (2.21)
\end{aligned}$$

This leads to the final expression for the electrostatic contribution to the PMF,

$$\Delta W_{elec} = \sum_{A\in inner} q_A \phi_s^o(\mathbf{r}_A) + \frac{1}{2}\sum_{mn} Q_m M_{mn} Q_n. \qquad (2.22)$$

In MD simulations employing the GSBP, the inner region atoms move on the PMF surface that is defined as

$$\begin{aligned}
W(\mathbf{R}_i, \mathbf{1}, ..., \mathbf{n}) &= U(\mathbf{R}_i, \mathbf{1}, ..., \mathbf{n}) + \Delta W_{cr} + \Delta W_{np} \\
&\quad + \sum_{A\in inner} q_A \phi_s^o(\mathbf{r}_A) + \frac{1}{2}\sum_{mn} Q_m M_{mn} Q_n. \qquad (2.23)
\end{aligned}$$

The last two terms correspond to the total electrostatic influence on the inner region and constitute the core of the GSBP.

2.3 Poisson-Boltzmann Equation

The linearized form of the PB equation provides an accurate estimate of the electrostatic potential if the response of the system can be described by a continuous dielectric.

$$\nabla\left[\epsilon(\mathbf{r})\nabla\phi(\mathbf{r})\right] - \bar{\kappa}^2(\mathbf{r})\phi(\mathbf{r}) = -4\pi\rho(\mathbf{r}), \qquad (2.24)$$

The electrostatic potential $\phi(\mathbf{r})$ is determined by the spatial dielectric function $\epsilon(\mathbf{r})$, the modified Debye-Hückel screening factor $\bar{\kappa}(\mathbf{r})$, and the charge density $\rho(\mathbf{r})$. Analytic solution of the PB equation is only possible for a few idealized shapes of the dielectric interface such as spheres or cylinders. Therefore, a numerical approach is employed in practice to solve the PB

equation. In a finite-difference ansatz, the continuous functions are approximated by distinct values at points on a regular cubic grid of mesh size h. The potential at grid point j is

$$\phi_j = \int \phi(\mathbf{r}) d\mathbf{r} \tag{2.25}$$

where $\mathbf{r}$ goes over all points in space that are closer to grid point j than to any other grid point. Discretizing all functions, the potential at a specific grid point 0, ϕ_0, takes the following form in the finite-difference ansatz:

$$\phi_0 = \frac{\sum_{i=1}^{6} \epsilon_i \phi_i + 4\pi q_0/h}{\sum_{i=1}^{6} \epsilon_i + \bar{\kappa}_0^2 h^2} \tag{2.26}$$

Here, ϕ_i is the potential value at the six neighboring grid points and ϵ_i is the dielectric constant at the midpoint between ϕ_0 and ϕ_i. The PB equation is solved by making an initial guess for all values of ϕ and then updating the values at all grid points based on equation 2.26. The updating procedure is repeated until convergence is reached. This simple approach is also described as Jacobian relaxation.[78]

Moreover, the values of ϕ at the boundary of the cubic grid have to be specified before the iterative computation of ϕ. Since exact values are not available, approximate values for the boundary points have to be found. If the grid is of adequate size, the dielectric boundary is far from the grid boundary such that the Debye-Hückel expression can be used to determine the boundary values:

$$\phi_i = \sum_j \frac{q_j e^{-\kappa r_{ij}}}{\epsilon r_{ij}}, \tag{2.27}$$

Dirichlet boundary conditions are applied, *i. e.*, the boundary values are not relaxed.[79]

Using Jacobian relaxation, the value of ϕ_0 in the nth iteration depends on the values ϕ_i in the $(n-1)$th iteration. Since the new value of ϕ_0 depends only on the the values of ϕ at the surrounding grid points, it is more efficient to split each iteration into two steps and the set of all grid points into two subsets. Subset A contains all even grid points, *i. e.*, those points where the sum of grid indices is even. Subset B contains all odd grid points. Equation 2.26 implies that any grid point of set A depends only on ϕ values at grid points of set B. In the first step of iteration n, all grid points of subset A are

updated based on the ϕ values of the grid points of subset B from iteration $n-1$. In the second step, the grid points of subset B are updated based on the values of the updated grid points in subset A from iteration n (instead of $n-1$ as in Jacobian relaxation). Since the values of subset B are one iteration ahead, the next update of subset A will also lead to better values (compared to Jacobian relaxation). This procedure converges twice as fast as Jacobian relaxation and is named Gauss-Seidel relaxation.[80]
Convergence can be further accelerated by means of the successive over-relaxation method (SOR). Using this approach, the value of ϕ_0 after the nth iteration, ϕ_o^n, is first computed using equation 2.26 and then extrapolated on the basis of the results from the current iteration and the previous one.

$$\phi_0^n = \omega\phi_0^n + (1-\omega)\phi_0^{n-1} \tag{2.28}$$

Here, ω is the relaxation parameter, and for $\omega > 1$ over-relaxation is employed. The optimal value for ω can only be estimated once the solution is known.[80] Therefore, Chebyshev acceleration is used to approximate the optimal value of ω during the iterative procedure. ω is initially chosen to be one and then updated iteratively.

$$\begin{aligned} \omega^0 &= 1 \\ \omega^1 &= \frac{1}{1-\gamma^2/2} \\ \omega^{n+1} &= \frac{1}{1-\omega^n\gamma^2/4} \\ \omega^\infty &\rightarrow \omega_{optimal} \end{aligned} \tag{2.29}$$

Here, γ is the spectral radius of the matrix representation of the finite-difference solution which under Dirichlet boundary conditions can be estimated as

$$\gamma = 1 - \frac{\pi^2}{2N^2}, \tag{2.30}$$

where N is the number of grid points in one dimension.[81]
A new module was added to the ChemShell program which uses the Gauss-Seidel relaxation method in combination with SOR to solve the PB equation. This module was used in all following applications.

2.4 QM/MM-Free Energy Perturbation

The QM/MM-free energy perturbation (QM/MM-FEP) method was developed by Yang *et al.*[15] It is based on the perturbation expression for the free

energy difference that was proposed by Zwanzig.[82] In this ansatz, the free energy difference between two states a and b is calculated as the canonical average of the exponential potential energy difference sampled on the potential energy surface of state a.

$$\begin{aligned} \Delta A^{a\to b} &= -\frac{1}{\beta} ln \left[\int P_a(\mathbf{r}) exp\left(-\beta \left[E^b(\mathbf{r}) - E^a(\mathbf{r})\right]\right) d\mathbf{r} \right] \\ &= -\frac{1}{\beta} ln \left\langle exp\left(-\beta \Delta E^{a\to b}\right)\right\rangle_a , \end{aligned} \tag{2.31}$$

where β is $\frac{1}{k_B T}$, with k_B being the Boltzmann constant. In the limiting case of complete sampling, the FEP method is exact.
To apply the concept of FEP efficiently with QM/MM Hamiltonians, Yang *et al.* introduced three approximations: (1) the dynamics of the QM and MM subsystems are independent, (2) the entropy change in the QM region can be estimated from the harmonic approximation, and (3) the electrostatic QM-MM interactions are approximated by interactions between QM and MM point charges, with the QM charges fitted to the electrostatic potential (ESP).
For a QM/MM potential energy, the free energy difference takes the following form:

$$\Delta A^{a\to b} = -\frac{1}{\beta} ln \left\langle exp\left(-\beta \left[\Delta E^{a\to b}_{QM} + \Delta E^{a\to b}_{MM} + \Delta E^{a\to b}_{QM-MM}\right]\right)\right\rangle_a \tag{2.32}$$

Assuming that QM and MM degrees of freedom can be sampled independently and that sampling over the QM degrees of freedom can be neglected, the expression is simplified:

$$\Delta A^{a\to b} = \Delta E^{a\to b}_{QM} - \frac{1}{\beta} ln \left\langle exp\left(-\beta \left[\Delta E^{a\to b}_{MM} + \Delta E^{a\to b}_{QM-MM}\right]\right)\right\rangle_{MM,a} \tag{2.33}$$

Now, the states a and b are defined to correspond to different arrangements of the QM atoms such that the MM energy is not affected by the perturbation and only the QM-MM interaction energy has to be sampled.

$$\Delta A^{a\to b} = \Delta E^{a\to b}_{QM} - \frac{1}{\beta} ln \left\langle exp\left(-\beta \Delta E^{a\to b}_{QM-MM}\right)\right\rangle_{MM,a} \tag{2.34}$$

Using the electronic embedding scheme,[83] an exact calculation of the electrostatic QM/MM interactions necessitates solving the SCF equations for each MM configuration during sampling. In the QM/MM-FEP approach, computation of these interactions is drastically simplified by two assumptions: the QM density is frozen and approximated by atomic ESP charges.

Therefore, all QM calculations are avoided during sampling.
In practice, a QM/MM-FEP calculation is initiated by calculating the potential energy profile by means of constrained optimizations. A reaction coordinate ξ describing the reaction is defined and used to split the reaction into discrete windows characterized by a corresponding value ξ_i. For each window i, the reaction coordinate is constrained to some ξ_i and all other QM and MM degrees of freedom are optimized. This yields the potential energy profile of the reaction and a set of geometries along the reaction coordinate. Next, the difference of the free QM-MM interaction energy, $\Delta A^{i\rightarrow i+1}_{QM-MM}$, between every two adjacent windows i and $i+1$ is calculated. The difference is computed as a perturbation of the structure of window i with the QM structure of window $i+1$ (corresponding to arbitrary states a and b in the previous equations).

$$\Delta E^{i\rightarrow i+1}_{pert} = E_{QM-MM}(\mathbf{r}^{i+1}_{QM}, \mathbf{r}^{i}_{MM}) - E_{QM-MM}(\mathbf{r}^{i}_{QM}, \mathbf{r}^{i}_{MM}) \tag{2.35}$$

The change in the free energy that corresponds to the perturbation of the QM structure is obtained by sampling over the MM phase space at window i. This means that the MM forces refer to the interaction with the QM structure of window i which is frozen during sampling.

$$\Delta A^{i\rightarrow i+1}_{QM-MM} = -\frac{1}{\beta} ln \left\langle exp\left(-\beta \Delta E^{i\rightarrow i+1}_{pert}\right)\right\rangle_{MM,i} \tag{2.36}$$

Finally, the energy of the QM part is corrected for entropic and thermal effects. At the stationary points, the contributions from zero-point vibrational energies (ΔE^{ZPE}_{QM}), thermal contributions to the internal energy (ΔU^{th}_{QM}), and entropy (ΔS_{QM}) to the QM free energy difference can be estimated.

$$\Delta A_{QM} = \Delta E_{QM} + \Delta E^{ZPE}_{QM} + \Delta U^{th}_{QM} - T\Delta S_{QM} \tag{2.37}$$

The correction terms are evaluated from harmonic frequency calculations of the QM atoms and standard methods from statistical thermodynamics.[15,84] The total free energy difference is now:

$$\Delta A = \Delta A_{QM} + \Delta A_{QM-MM} \tag{2.38}$$

Chapter 3

Generalized Solvent Boundary Potential for Semiempirical QM/MM Hamiltonians

3.1 Motivation

In the previous chapter, two different dual-layer methods for simulations of biomolecular systems were introduced: the hybrid QM/MM ansatz (section 2.1) and the concept of boundary potentials such as the GSBP (section 2.2). Both methods have in common that they use a more accurate method for the central more important region of the system, and a more approximate and faster method for the rest of the system. In this way, both approaches reduce the computational costs without introducing significant errors since all regions are described at an appropriate level. They differ, however, in the position of the boundary. In the QM/MM ansatz, only those atoms *directly* involved in the reaction belong to the QM region which typically has a size of 50 to 100 atoms. Yet, the inner region in the boundary potential ansatz is significantly larger. Only those atoms that need not be represented atomistically are assigned to the outer region. The inner region typically has a spherical shape with a radius of about 15 to 20 Å, is centered on the active site of the enzyme, and encompasses 2,000 to 4,000 atoms.

Using the boundary definitions from both approaches, a biomolecular system may be divided into three regions as illustrated in figure 3.1: the active site (corresponding to the QM region), the intermediate region (inner region without active site), and the outer region. The three regions contribute in different ways to the catalytic effect of the enzyme. The residues in the active site participate directly in the reaction and frequently form hydrogen

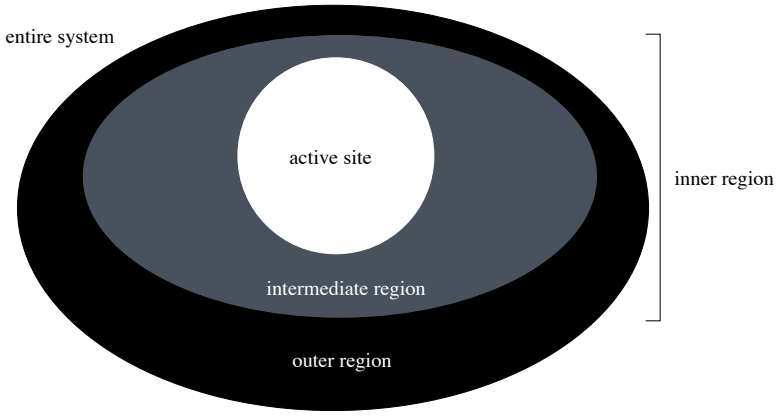


Figure 3.1: Definition of regions in biomolecular systems as defined from the QM/MM and MM/GSBP approaches.

or covalent bonds with the substrate. Therefore, the active site (including the substrate) is modeled with an accurate QM Hamiltonian in the QM/MM ansatz. The intermediate region provides a scaffold for the geometrical arrangement of the residues in the active site and shapes the electrostatic potential in the active site. In both approaches, an atomistic force field is considered sufficient to model these effects. Although the outer region can be important for the overall function of an enzyme, it influences the catalytic reaction steps mainly through its electrostatic potential. The outer region is thus represented by an efficient but even more approximate boundary potential in the boundary potential ansatz. This three-layer concept is only suitable for well-localized chemical events. However, this is exactly the field of application of QM/MM methods.
The convincing results obtained with both dual-layer approaches indicate that these levels of theory are adequate to describe the influence of the respective region on the reaction energetics. It thus seems natural to combine both approaches. In the resulting three-layer method, the active site is described at the QM level, the intermediate region at the MM level, and the outer region by a boundary potential. From a QM/MM perspective, the MM region is separated into an MM and a boundary potential region. From the perspective of classical MM simulations with boundary potentials, the inner region is separated into a QM and an MM region. The relationship of the three approaches is summarized in table 3.1.

For a specific three-layer method, choices regarding boundary potential and QM method have to be made. As already mentioned in chapter 1, the GSBP was selected due to its convincing theoretical construction and

Table 3.1: Levels of theory used for the three regions in a biomolecular system with different multi-level methods.

	QM/MM	MM with BP[a]	QM/MM/BP[a]
active site	QM	MM	QM
intermediate region	MM	MM	MM
outer region	MM[b]	BP[a]	BP[a]

[a] boundary potential

[b] frequently constrained

its performance in previous applications.[55,58–64] The GSBP was specifically designed to increase the efficiency of MD simulations, and therefore, the QM/MM/GSBP method will also be mainly applied in MD simulations. Several thousands of gradient and energy evaluations are necessary in those calculations so that currently only semiempirical QM methods are applicable.

Previously, the GSBP method was adapted to the hybrid QM/MM framework exclusively in combination with the semiempirical SCC-DFTB Hamiltonian for the QM region by Cui and co-workers.[56] In the light of the success of reaction-specific parameterizations of semiempirical methods based on the neglect of diatomic differential overlap (NDDO) approximation in QM/MM simulations[85–89] and the widespread use of NDDO-based QM/MM methods in general,[24] it was considered desirable to adapt and implement the GSBP as an efficient means to treat long range electrostatics in NDDO-based QM/MM simulations. This is further substantiated by recent findings that NDDO-based methods are more reliable for certain properties and systems,[90] and that the use of SCC-DFTB may be problematic for specific systems.[91] Accordingly, the GSBP was adapted and implemented for NDDO-based QM/MM approaches in this work.

As illustrated in section 2.2, the use of the GSBP is connected with a significant overhead. Initially, the solvent-shielded static field and the matrix representation of the reaction field Green's function have to be calculated. Computation of the reaction field matrix implies solving several hundred linearized PB equations, and is therefore rather demanding. Furthermore, the accuracy of the GSBP and the costs of its overhead strongly depend on the choice of parameters that are inherent to the GSBP and the finite-difference solution of the PB equation. A systematic determination of the best parameters for the GSBP has not been pursued up to date.

Hence, a set of parameters that provide the accuracy that is necessary to

mimic the effect of the outer region at optimal computational costs is determined. Based on these parameters, the overhead and the savings related to the GSBP are quantified and the minimum system size is established for which the GSBP is more efficient than standard approaches using non-truncated Coulombic electrostatics. Moreover, improved algorithms are presented that decrease the costs for computation of the reaction field matrix significantly.

3.2 Implementation for NDDO-based QM/MM Methods

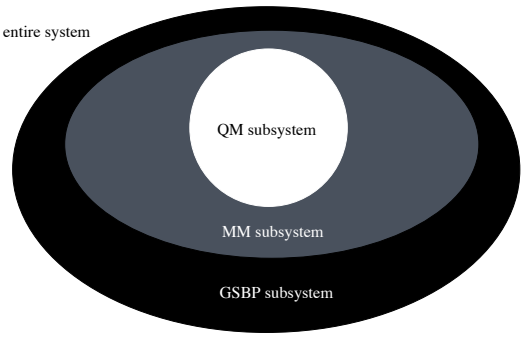


Figure 3.2: Definition of QM/MM/GSBP system partitioning.

Extension of the GSBP to QM/MM methods necessitates further subdivision of the inner region into QM and MM regions. Consequently, the inner region charge distribution splits up into QM and MM charge distributions that interact separately with the static outer region field, ϕ_s^o, and the reaction field Green's function, G_{rf}. Equation 2.22 has to be modified as follows to account for these changes:

$$\begin{aligned}\Delta W_{elec} &= \sum_{A\in MM} q_A\phi_s^o(\mathbf{r}_A) + \int d\mathbf{r}\rho^{QM}(\mathbf{r})\phi_s^o(\mathbf{r}) \\ &+\frac{1}{2}\sum_{mn} Q_m^{QM} M_{mn} Q_n^{QM} + \sum_{mn} Q_m^{QM} M_{mn} Q_n^{MM,cs} \\ &+\frac{1}{2}\sum_{mn} Q_m^{MM} M_{mn} Q_n^{MM}. \end{aligned} \tag{3.1}$$

The main issue that arises when introducing QM atoms into the GSBP framework is the representation of the QM charge distribution in the terms that describe the interaction with the outer region field and the reaction field. As NDDO-based semiempirical QM methods use only a minimum set of relatively tight basis functions, the QM charge distribution is well represented by a set of Mulliken charges.[92] Now, the QM-dependent terms in equation 3.1 can be calculated in close analogy to the MM terms,

$$\int d\mathbf{r}\rho^{QM}(\mathbf{r})\phi_s^o(\mathbf{r}) = \sum_{A\in QM} q_A^{Mull}\phi_s^o(\mathbf{r}_A) \tag{3.2}$$

and

$$Q_n^{QM} = \int d\mathbf{r}\rho^{QM}(\mathbf{r})b_n(\mathbf{r}) = \sum_{A\in QM} q_A^{Mull}b_n(\mathbf{r}_A). \tag{3.3}$$

Here, q_A^{Mull} denotes the Mulliken charges representing the QM charge distribution, and Q_n^{QM} denotes the multipole moments of the QM charge distribution. Still, one is facing two technical difficulties. First, electrostatic interactions at the QM-MM boundary need to be treated with special care to avoid overpolarization of the QM electron density. Thus, the QM electron density does not interact with the full MM charge distribution but with a modified one. In this work, the GSBP is implemented for use in combination with the charge-shift scheme (see subsection 2.1.3),[76] and therefore, the QM charge distribution interacts with the reaction field potential that is induced by all MM charges after applying the charge-shift scheme to the M^1 atoms (MM^{cs}) (fourth term in equation 3.1), with

$$Q_n^{MM,cs} = \sum_{A\in MM^{cs}} q_A b_n(\mathbf{r}_A). \tag{3.4}$$

Second, using electronic embedding,[83] the QM wave function interacts with all MM point charges and the PDC. Hence, the GSBP contributions have to be accommodated at the level of the SCF iterations during optimization of the wave function by additional terms in the Fock matrix.

$$
\begin{aligned}
F_{\mu\nu}^{GSBP} &= \frac{\partial \Delta W_{elec}}{\partial P_{\mu\nu}} \\
&= \frac{\partial}{\partial P_{\mu\nu}} \int d\mathbf{r} \rho_{QM}(\mathbf{r}) \phi_s^o(\mathbf{r}) \\
&\quad + \frac{1}{2} \frac{\partial}{\partial P_{\mu\nu}} \sum_{mn} Q_m^{QM} M_{mn} Q_n^{QM} + \frac{\partial}{\partial P_{\mu\nu}} \sum_{mn} Q_m^{QM} M_{mn} Q_n^{MM,cs} \\
&= \int d\mathbf{r} \left[\frac{\partial}{\partial P_{\mu\nu}} \rho_{QM}(\mathbf{r}) \right] \phi_s^o(\mathbf{r}) \\
&\quad + \frac{1}{2} \sum_{mn} \left[\frac{\partial}{\partial P_{\mu\nu}} Q_m^{QM} \right] M_{mn} Q_n^{QM} + \frac{1}{2} \sum_{mn} Q_m^{QM} M_{mn} \left[\frac{\partial}{\partial P_{\mu\nu}} Q_n^{QM} \right] \\
&\quad + \sum_{mn} \left[\frac{\partial}{\partial P_{\mu\nu}} Q_m^{QM} \right] M_{mn} Q_n^{MM,cs} \qquad (3.5)
\end{aligned}
$$

Here, F and P are the SCF Fock and density matrices in the atomic orbital (AO) basis set representation, respectively. Using Mulliken charges to represent the QM charge density, the derivatives of the QM charge density and the QM multipole moments take the following form:

$$
\frac{\partial}{\partial P_{\mu\nu}} \rho_{QM}(\mathbf{r}) = \sum_A^{QM} \delta(\mathbf{r} - \mathbf{r}_A) \frac{\partial}{\partial P_{\mu\nu}} q_A^{Mull}, \qquad (3.6)
$$

with

$$
\begin{aligned}
\frac{\partial}{\partial P_{\mu\nu}} q_A^{Mull} &= \frac{\partial}{\partial P_{\mu\nu}} \left[Z_A - \frac{1}{2} \sum_{\alpha \in A} \sum_{\beta}^{AO} P_{\alpha\beta} S_{\alpha\beta} - \frac{1}{2} \sum_{\alpha}^{AO} \sum_{\beta \in A} P_{\alpha\beta} S_{\alpha\beta} \right] \\
&= -\frac{1}{2} S_{\mu\nu} \delta(\mathbf{r}_\mu - \mathbf{r}_A) - \frac{1}{2} S_{\mu\nu} \delta(\mathbf{r}_\nu - \mathbf{r}_A) \qquad (3.7)
\end{aligned}
$$

where $\mathbf{r}_\mu$ is the coordinate of the atom to which the AO basis function μ is attached. The derivatives of the QM multipole moments can be expressed based on the derivatives of the Mulliken charges.

$$
\begin{aligned}
\frac{\partial}{\partial P_{\mu\nu}} Q_n^{QM} &= \sum_A^{QM} \left[\frac{\partial}{\partial P_{\mu\nu}} q_A^{Mull} \right] b_n(\mathbf{r}_A) \\
&= \sum_A^{QM} \left[-\frac{1}{2} S_{\mu\nu} \delta(\mathbf{r}_\mu - \mathbf{r}_A) - \frac{1}{2} S_{\mu\nu} \delta(\mathbf{r}_\nu - \mathbf{r}_A) \right] b_n(\mathbf{r}_A) \\
&= -\frac{1}{2} S_{\mu\nu} \left[b_n(\mathbf{r}_\mu) + b_n(\mathbf{r}_\nu) \right]. \qquad (3.8)
\end{aligned}
$$

Employing these expressions, an explicit form of equation 3.5 can be given.

$$\begin{aligned} F_{\mu\nu}^{GSBP} &= \sum_{A}^{QM}\left[-\frac{1}{2}S_{\mu\nu}\delta(\mathbf{r}_\mu-\mathbf{r}_A)-\frac{1}{2}S_{\mu\nu}\delta(\mathbf{r}_\nu-\mathbf{r}_A)\right]\phi_s^o(\mathbf{r}_A) \\ &\quad -\frac{1}{4}S_{\mu\nu}\sum_{mn}\left[b_m(\mathbf{r}_\mu)+b_m(\mathbf{r}_\nu)\right]M_{mn}Q_n^{QM} \\ &\quad -\frac{1}{4}S_{\mu\nu}\sum_{mn}Q_m^{QM}M_{mn}\left[b_n(\mathbf{r}_\mu)+b_n(\mathbf{r}_\nu)\right] \\ &\quad -\frac{1}{2}S_{\mu\nu}\sum_{mn}\left[b_m(\mathbf{r}_\mu)+b_m(\mathbf{r}_\nu)\right]M_{mn}Q_n^{MM,cs} \qquad (3.9) \\ &= -\frac{1}{2}S_{\mu\nu}\left[\phi_s^o(\mathbf{r}_\mu)+\phi_{rf}^o(\mathbf{r}_\nu)\right] \qquad (3.10) \\ &\quad -\frac{1}{2}S_{\mu\nu}\sum_{mn}b_m(\mathbf{r}_\mu)M_{mn}Q_n^{QM}-\frac{1}{2}S_{\mu\nu}\sum_{mn}b_m(\mathbf{r}_\nu)M_{mn}Q_n^{QM} \\ &\quad -\frac{1}{2}S_{\mu\nu}\sum_{mn}b_m(\mathbf{r}_\mu)M_{mn}Q_n^{MM,cs}-\frac{1}{2}S_{\mu\nu}\sum_{mn}b_m(\mathbf{r}_\nu)M_{mn}Q_n^{MM,cs} \end{aligned}$$

Expansion of the multipole moments in terms of the Mulliken charges allows additional contraction of terms.

$$\begin{aligned} F_{\mu\nu}^{GSBP} &= -\frac{1}{2}S_{\mu\nu}\left[\phi_s^o(\mathbf{r}_\mu)+\sum_{mn}b_m(\mathbf{r}_\mu)M_{mn}Q_n^{MM,cs}\right] \\ &\quad -\frac{1}{2}S_{\mu\nu}\left[\phi_s^o(\mathbf{r}_\nu)+\sum_{mn}b_m(\mathbf{r}_\nu)M_{mn}Q_n^{MM,cs}\right] \\ &\quad -\frac{1}{2}S_{\mu\nu}\sum_{A}^{QM}q_A^{Mull}\sum_{mn}b_m(\mathbf{r}_\mu)M_{mn}b_n(\mathbf{r}_A) \\ &\quad -\frac{1}{2}S_{\mu\nu}\sum_{A}^{QM}q_A^{Mull}\sum_{mn}b_m(\mathbf{r}_\nu)M_{mn}b_n(\mathbf{r}_A) \qquad (3.11) \\ &= -\frac{1}{2}S_{\mu\nu}\left[\Omega_C+\Omega_D\right] \\ &\quad -\frac{1}{2}S_{\mu\nu}\sum_{A}^{QM}q_A^{Mull}\left[\Gamma_{CA}+\Gamma_{DA}\right];\ \mu\in C,\ \nu\in D \qquad (3.12) \end{aligned}$$

Here, the AO basis functions μ and ν are attached to the QM atoms C and D, respectively. The atom-dependent matrices Ω_C and Γ_{CA} are defined as

$$\Omega_C = \phi_s^o(\mathbf{r}_C)+\sum_{mn}b_m(\mathbf{r}_C)M_{mn}Q_n^{MM,cs} \qquad (3.13)$$

and

$$\Gamma_{CA} = \sum_{mn} b_m(\mathbf{r}_C) M_{mn} b_n(\mathbf{r}_A). \tag{3.14}$$

These matrices are first and second order quantities that are constant during the SCF procedure.

Moreover, the GSBP also affects the atomic forces. Its contribution to the analytic gradient can be evaluated by taking the first derivative of the GSBP contribution to the PMF with respect to the atomic coordinates. In case of a QM atom, the analytic derivative takes the following form

$$\begin{aligned} \frac{\partial}{\partial \mathbf{r}_A} \Delta W_{elec} &= q_A^{Mull} \frac{\partial}{\partial \mathbf{r}_A} \phi_s^o(\mathbf{r}_A) + \sum_{B \in QM} \frac{\partial q_B^{Mull}}{\partial \mathbf{r}_A} \phi_s^o(\mathbf{r}_B) \\ &\quad + \sum_{mn} \left[\frac{\partial}{\partial \mathbf{r}_A} Q_m^{QM} \right] M_{mn} \left[Q_n^{QM} + Q_n^{MM,cs} \right], \end{aligned} \tag{3.15}$$

where the derivatives of the QM multipole moments are calculated as

$$\frac{\partial}{\partial \mathbf{r}_A} Q_m^{QM} = q_A^{Mull} \frac{\partial}{\partial \mathbf{r}_A} b_m(\mathbf{r}_A) + \sum_{B \in QM} \frac{\partial q_B^{Mull}}{\partial \mathbf{r}_A} b_m(\mathbf{r}_B). \tag{3.16}$$

In contrast to a previous implementation of the GSBP for hybrid QM/MM approaches,[56] it was found to be necessary for NDDO-based QM methods to include the contribution from coupled Mulliken charge derivatives, $\frac{\partial q_A^{Mull}}{\partial \mathbf{r}_B}$, to compute accurate gradients of the QM atoms. Using only one-center Mulliken charge derivatives, $\frac{\partial q_A^{Mull}}{\partial \mathbf{r}_A}$, a mean absolute deviation (MAD) of the components of the QM gradient in the range of 10^{-3} a.u. was observed (compared with finite-difference reference values). Incorporating the contribution from the coupled Mulliken charge derivatives reduces the MAD to the order of 10^{-5} a.u. which is sufficiently accurate.
The derivative of the Mulliken charge of QM atom A with respect to its atomic coordinates is

$$\frac{\partial}{\partial \mathbf{r}_A} q_A^{Mull} = \frac{\partial}{\partial \mathbf{r}_A} \left[Z_A - \frac{1}{2} \sum_{\alpha \in A} \sum_{\beta}^{AO} P_{\alpha\beta} S_{\alpha\beta} - \frac{1}{2} \sum_{\alpha}^{AO} \sum_{\beta \in A} P_{\alpha\beta} S_{\alpha\beta} \right] \tag{3.17}$$

The zero differential overlap (ZDO) approximation simplifies this equation significantly since the overlap matrix, $S_{\alpha\beta}$, reduces to a unit matrix, and

therefore, the derivative of the overlap matrix is zero for all elements.

$$\frac{\partial}{\partial \mathbf{r}_A} q_A^{Mull} = -\frac{1}{2}\sum_{\alpha\in A}\sum_{\beta}^{AO} S_{\alpha\beta}\frac{\partial}{\partial \mathbf{r}_A}P_{\alpha\beta} - \frac{1}{2}\sum_{\alpha}^{AO}\sum_{\beta\in A} S_{\alpha\beta}\frac{\partial}{\partial \mathbf{r}_A}P_{\alpha\beta} \quad (3.18)$$

$$= -\sum_{\alpha\in A}\frac{\partial}{\partial \mathbf{r}_A}P_{\alpha\alpha} \quad (3.19)$$

$$(3.20)$$

Although Mulliken charge derivatives take a very simple form, their computation is complicated, since the coupled-perturbed SCF (CPSCF) equations have to be solved to calculate the derivatives of the SCF density matrix.[93,94] In case of an MM atom, the evaluation of the derivative of the GSBP contribution to the PMF is less demanding,

$$\frac{\partial}{\partial \mathbf{r}_A}\Delta W_{elec} = q_A\frac{\partial}{\partial \mathbf{r}_A}\phi^o_{rf}(\mathbf{r}_A) + q_A\sum_{mn}\left[\frac{\partial}{\partial \mathbf{r}_A}b_n(\mathbf{r}_A)\right] M_{mn}\left[Q_n^{QM} + Q_n^{MM}\right]. \quad (3.21)$$

3.3 Computation of the Reaction Field Matrix

Although the reaction field matrix is formally the matrix representation of the reaction field Green's function, the computation of this matrix follows a different approach that exploits the fact that its mnth element corresponds to the interaction of b_m with the reaction field induced by b_n.[54]

$$M_{mn} = \int d\mathbf{r} b_m(\mathbf{r})\phi_{rf}(\mathbf{r}; b_n(\mathbf{r})) \quad (3.22)$$

To calculate $\phi_{rf}(\mathbf{r}; b_n(\mathbf{r}))$, it is necessary to solve the PB equation with the dielectric boundary defined by the macromolecule, and the charge distribution defined by b_n in the inner region and set to zero in the outer region, for vacuum and solvent conditions. Since a standard GSBP calculation employs about 400 basis functions,[54] computation of the reaction field matrix implies solving the PB equation about 800 times. This procedure is computationally expensive and dominates the GSBP-related overhead. Therefore, three approaches to accelerate computation of the reaction field matrix are presented in this section.

3.3.1 Coarsening of the Inner Region

In finite-difference solutions of the PB equation, the boundary values are commonly set using the Debye-Hückel expression,[77]

$$\phi_i = \sum_j \frac{q_j e^{-\kappa r_{ij}}}{\epsilon r_{ij}}, \tag{3.23}$$

that implies summation over all point charges q_j for each boundary point ϕ_i. With a continuous charge distribution in the inner region, determination of the boundary values becomes computationally expensive. In the original GSBP work, a focusing procedure[95] is used to reduce these computational costs. In this procedure, the PB equation is first solved for a rough outer grid (grid I) with large spatial extent. Subsequently, a fine inner grid (grid II) focusing on the inner region with boundary values defined by grid I is used to calculate an accurate electrostatic potential. However, even when using such a focusing procedure, determination of the boundary values of grid I still has a significant share of the computational costs. Since the boundary points of grid I are far from the inner region and the "charge" in the outer region is zero, a less accurate representation of the basis function in the inner region is expected to be sufficient. Therefore, the "coarsening of the inner region" (CIR) approximation is introduced which utilizes a very rough grid (grid III) to represent the "charge" distribution that is only used to determine the boundary values of grid I. The mesh size of grid III is the product of the new CIR factor and the mesh size of grid I, *i.e.*, a CIR factor of 1.0 corresponds to a standard focusing procedure.

3.3.2 Linear Interpolation

In view of the large distance between the boundary points of grid I and the inner region, it is evident that the boundary values are slowly varying. Therefore, a simple interpolation scheme is introduced that reduces the number of explicitly determined boundary values significantly. On the edges, every second and on the faces every fourth boundary value is calculated using the Debye-Hückel expression. The remaining boundary values are determined by linear interpolation from the adjacent four or two boundary points. This scheme is illustrated in figure 3.3. For an example grid with 100^3 points, the linear interpolation scheme reduces the number of explicitly determined boundary values from 58,416 to 14,802.

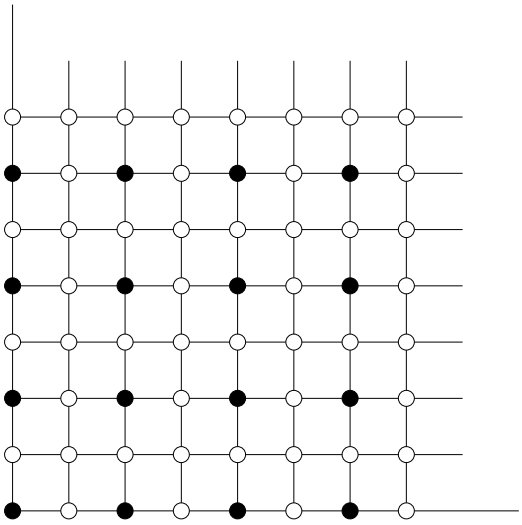


Figure 3.3: Interpolation scheme used to define boundary values in finite-difference solutions of the PB equation. Black spheres represent boundary points that are set using the Debye-Hückel expression. White spheres represent boundary points that are set by interpolation.

3.3.3 Modified Stripping

In a finite-difference solution to the PB equation with zero salt conditions, the potential at a particular grid point, ϕ_0, is updated using the potential at the six nearest neighbors, the dielectric constants at the midpoints, and the charge assigned to that grid point (see equation 2.26). This procedure implies 13 additions, 7 multiplications, and 1 division per grid point. Honig *et al.* demonstrated[80] that the number of mathematical operations can be reduced significantly for most grid points. For a point with zero charge that is surrounded by a uniform dielectric constant, equation 2.26 simplifies to

$$\phi_0 = \frac{1}{6}\sum_{i=1}^{6}\phi_i. \tag{3.24}$$

Updating these points requires only 6 additions and 1 multiplication. This procedure is termed "stripping" because the points are updated separately.[80] As a continuous charge distribution is used in the computation of the reaction field matrix, there are no points without charge in the inner region. Therefore, a "modified stripping" approach is applied and the zero charge condition is dropped: for all points surrounded by a uniform dielectric constant with

arbitrary charge, equation 2.26 is simplified as follows:

$$\phi_0 = \frac{1}{6}\sum_{i=1}^{6}\phi_i + \frac{2\pi q_0}{3h\epsilon} \tag{3.25}$$

Although three additional operations per grid point (one addition, one multiplication, and one division) are necessary compared to the standard stripping approach, modified stripping offers computational savings since it is applicable to a significantly larger number of grid points in the computation of the reaction field matrix.

3.4 Computational Details

The GSBP was implemented in a developmental version of the modular program package ChemShell.[76] The energy and gradient evaluations for the QM part were performed with the MNDO2004 program that was modified locally to account for the GSBP contribution. The AM1 method was chosen as the QM Hamiltonian.[96] The SCF convergence criterion was 10^{-8} eV. For the MM part, the DL_POLY[97] code was employed to run the CHARMM22 force field.[98] The PB equation was solved with the new ChemShell PB module and a convergence criterion of $2 \cdot 10^{-5}$ a.u. was employed for the maximum absolute change in every grid point. If not stated otherwise, the dielectric constants of the macromolecule, ϵ_m, and the solvent, ϵ_s, were set to 1 and 80, respectively. Van der Waals radii from the CHARMM22 force field were used to define the dielectric boundary. All calculations for which timings are reported were performed serially on 2.6 GHz AMD Opteron machines with 16 Gigabyte of memory.

3.5 Optimization of Parameters

The accuracy and the efficiency of the GSBP strongly depend on the values that are chosen for its inherent parameters. In this section, optimal mesh sizes of the inner and outer grid are determined. In addition, the accuracy of the approximations introduced in section 3.3 is assessed, and an optimum CIR factor is chosen.

3.5.1 Static Outer Region Field

The reliability of the static solvent-shielded outer region field, ϕ_s^o, to mimic the electrostatic potential in the inner region is judged by comparison to the

exact Coulombic potential. In vacuum environment, *i.e.*, with $\epsilon_m = \epsilon_s = 1$, the electrostatic interaction between the inner and outer region have to be identical when using the electrostatic potentials from solution of the PB equation and from the Coulomb expression. A model system consisting of a threonine molecule solvated in a TIP3P water ball with 30 Å radius and 4,175 water molecules was set up for this study. By means of classical molecular dynamics (MD) simulation, 10 different configurations of this model system were generated. For each configuration, the center of the inner region was taken to be the C_α carbon of threonine. All molecules with any atom within 18 Å from the center were assigned to the inner region. Depending on the configuration, the inner region contained between 2,858 and 2,978 atoms. As the electrostatic interaction energy varies with the size of the inner region, the average of the absolute percentage deviation in the electrostatic interaction energy is considered as a criterion for assessing the accuracy.
On average, the vacuum electrostatic interaction between inner and outer region (U^{io}_{elec}) was -3096.2 ± 243.5 kcal/mol. The average absolute percentage deviation was calculated for all combinations of outer grid mesh sizes of 1.0, 1.25, 1.5, 1.75, 2.0, and 2.5 Å, and inner grid mesh sizes of 0.25, 0.4, 0.6, and 0.8 Å. The results given in table 3.2 indicate that the interaction energy calculated from the PB electrostatic potential is very accurate. All mesh size combinations provide average deviations $< 0.3\%$ if the inner grid spacing is ≤ 0.6 Å, indicating that 0.6 Å is a reasonable choice for the inner grid spacing. For the outer grid, no reliable correlation was found between mesh size and accuracy.

To ensure that the accuracy is not euphemized by cancellation of errors, the reliability of the electrostatic forces in the inner region was also assessed. For this purpose, the MAD of the electrostatic force components of all atoms inside spherical active regions with radii of 15 and 17 Å were computed for each configuration. In table 3.3, the average of the MADs is given for all mesh size combinations. For both active regions, the accuracy of the electrostatic forces seems to be rather independent of the mesh sizes. Within 15 Å of the center of the inner region, computation of the electrostatic forces based on the potential from the PB equation is quite accurate with average MADs around $4 \cdot 10^{-5}$ to $8 \cdot 10^{-5}$ a.u. The average deviation increases by a factor of $4 - 5$ if the radius of the active region is extended to 17 Å. For each active region, there is only a very weak correlation between accuracy and mesh size. However, the accuracy strongly depends on the size of the active region. In figure 3.4, the average MAD for one mesh size combination (0.6 and 1.75 Å) is plotted as a function of the radius of the active region. This figure shows that the accuracy is very high for radii of up to 16 Å, then the deviation increases strongly. This behavior is identical for all mesh size

Table 3.2: Average absolute percentage deviation [%] of the electrostatic interaction between inner and outer region computed from the PB electrostatic potential with different mesh sizes of the inner and outer grid. Reference data was taken from calculations with the exact Coulombic potential.

	Inner grid size [Å]			
Outer grid size [Å]	0.25	0.40	0.60	0.80
1.00	0.13	0.14	0.22	0.55
1.25	0.16	0.28	0.26	0.61
1.50	0.25	0.28	0.28	0.62
1.75	0.21	0.28	0.28	0.64
2.00	0.23	0.16	0.24	0.61
2.50	0.25	0.27	0.27	0.60

combinations.

Keeping in mind the size of the inner region (18 Å, see above), one can conclude that a grid-based PB potential is not adequate to represent the details of the electrostatic potential of the outer region in close proximity to the outer region. Although these inaccuracies are likely to have only an insignificant effect on the region of interest if the size of the inner region is adequate, it is recommended to keep all atoms in the outer $2-3$ Å layer of the inner region fixed, since such a frozen layer will increase the reliability of the GSBP. A mesh size of 0.6 Å for the inner grid seems to provide an ideal trade-off between accuracy and computational costs. The results are not clear-cut concerning the outer grid mesh size where the accuracy seems to be rather independent of the mesh size. This indicates that the electrostatic potential is only slowly varying at the boundary of the inner grid. To be on the safe side, an outer grid spacing of 1.75 Å is selected.

3.5.2 Reaction Field Matrix

In section 3.3, three approaches to accelerate computation of the reaction field matrix were presented. While the modified stripping technique provides speed-up without loss of accuracy, the CIR and the linear interpolation approaches are approximations to define the boundary values more efficiently. Therefore, the computational savings and the associated loss of accuracy of these methods have to be analyzed.
For this assessment, one configuration of the model system with an inner

Table 3.3: Average mean absolute deviation [10^{-4} a.u.] of the electrostatic forces computed from the Poisson-Boltzmann (PB) electrostatic potential for different mesh sizes of the inner and outer grid used for solving the PB equation. Active regions with radii of 15 and 17 Å were chosen. Results from calculations with the exact Coulombic potential were used as reference data.

	Inner grid size [Å]			
Outer grid size [Å]	0.25	0.40	0.60	0.80
15 Å active region				
1.00	0.63	0.34	0.52	0.72
1.25	0.52	0.64	0.58	0.58
1.50	0.77	0.68	0.60	0.59
1.75	0.76	0.79	0.76	0.75
2.00	0.74	0.39	0.55	0.76
2.50	0.86	0.66	0.59	0.61
17 Å active region				
1.00	2.13	1.98	2.36	2.85
1.25	2.02	2.27	2.42	2.72
1.50	2.28	2.31	2.43	2.73
1.75	2.27	2.42	2.58	2.87
2.00	2.26	2.03	2.39	2.88
2.50	2.37	2.28	2.43	2.75

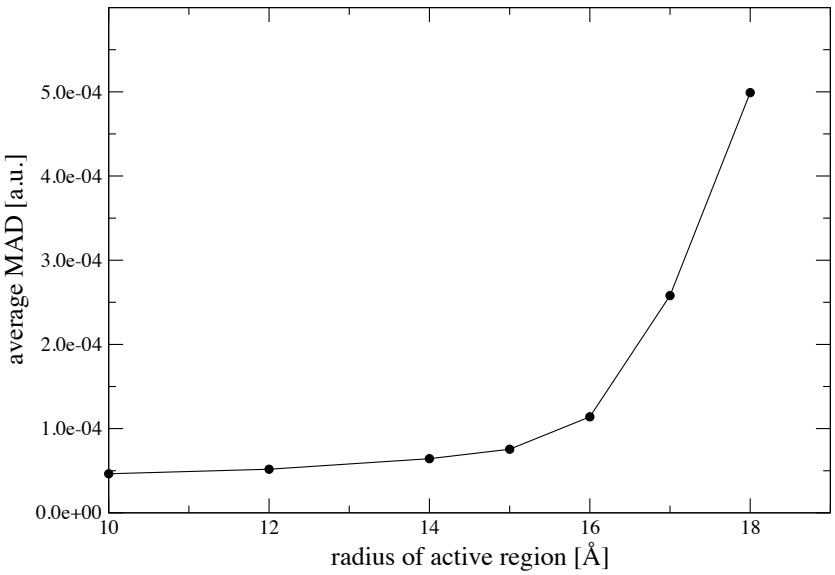


Figure 3.4: Average mean absolute deviation (MAD) of the electrostatic forces of all atoms inside the active region as a function of the radius of the active region. The forces were computed with the GSBP and compared to reference data from calculations that use the exact Coulombic potential. Mesh sizes of 0.6 and 1.75 Å were used for the inner and outer grid, respectively. The radius of the inner region is 18 Å (see text).

Table 3.4: Accuracy and computational costs of the reaction field matrix calculation using different approximations to define the boundary values (see text).

Boundary[a]	CIR[b]	ΔW^{ii}_{elec} [kcal/mol]	Dev.[c] [kcal/mol]	Time [h]	Rel. time [%]
DH	1.0	-120.27	3.73	8.22	100.00
DH	1.5	-120.07	3.93	4.83	58.81
DH	2.0	-120.10	3.91	4.01	48.78
DH	2.5	-120.08	3.93	3.79	46.08
DH	3.0	-119.21	4.80	3.56	43.28
DHLI	1.0	-120.26	3.74	4.53	55.11
DHLI	1.5	-120.06	3.95	3.68	44.80
DHLI	2.0	-120.08	3.92	3.54	43.05
DHLI	2.5	-120.06	3.94	3.38	41.19
DHLI no MS[d]	2.5	-120.06	3.94	3.71	45.17
DHLI	3.0	-119.20	4.81	3.33	40.54
ZERO	-	-116.38	7.62	3.34	40.65

[a] DH: Debye-Hückel, DHLI: Debye-Hückel with linear interpolation, ZERO: all boundary values are set to zero

[b] coarsening of the inner region factor

[c] dev. = deviation = calc. − ref. Reference values were obtained by solving the Poisson-Boltzmann equation without a basis set representation.

[d] no modified stripping

region of 2,978 atoms was selected. Spherical harmonics with the first 20 multipole moments ($L = 0 - 19$), *i.e.*, 400 basis functions, were used to represent the charge distribution. The previously determined best mesh size combination of 0.6 and 1.75 Å for the inner and outer grid was employed. The accuracy of the reaction field matrix was assessed by comparing the GSBP results for the solvation free energy of the inner region (ΔW^{ii}_{elec}) to the results of a finite-difference solution of the PB equation without a basis set representation, *i.e.*, in the complete basis set limit.

The accuracy and the costs for computation of the reaction field matrix were tested for CIR factor values of 1.0, 1.5, 2.0, 2.5, and 3.0 in combination with the standard Debye-Hückel (DH) method and the DH expression with linear interpolation (DHLI). The results are given in table 3.4. The combination of a CIR factor of 1.0 with DH boundary values corresponds to the standard GSBP method that reproduced the free solvation energy very well. With

the selected basis set, about 97 % of the free solvation energy are recovered. These results are certainly satisfying and support the finding of Im *et al.* that the solvation free energy is sufficiently converged with a basis set of this size.[54] If the CIR factor is increased to 1.5, 2.0 or 2.5, the deviation increases by only 0.2 kcal/mol from 3.73 kcal/mol to 3.93 kcal/mol. At the same time computational costs are reduced by 54 % from 8.22 h to 3.79 h. The DHLI method proves to be similarly efficient. With a CIR factor of 1.0 and DHLI boundary values, the deviation increases by only 0.1 kcal/mol relative to DH boundary values and the computational costs are reduced by 45 % to 4.53 h. Unfortunately, these two methods cannot be combined without loss of efficiency. The DHLI method in combination with a CIR factor of 2.5 yields a deviation of 3.94 kcal/mol (*i.e.*, 0.21 kcal/mol higher relative to the standard GSBP method), but the computational costs are merely reduced from 3.79 h (DH with a CIR factor of 2.5) to 3.38 h (DHLI with a CIR factor of 2.5). Considering the computation time for zero boundary values, it is understandable that the CIR and the DHLI method cannot be combined without loss of efficiency. Using zero boundary values, the relative computation time drops to 40.65 %. Hence, in a standard reaction field matrix computation, about 60 % of the computation time is used to define the boundary values. As either method, DHLI or CIR, reduces the computational costs for this step to only a fraction, combining the two methods gives only marginal extra savings. Overall, the combination of DHLI with a CIR factor of 2.5 reduces the computational costs by about 60 % with minimal loss of accuracy. Finally, the results also show that the computation time for this calculation increases by 10% without modified stripping (table 3.4).

In summary, it was found that the GSBP yields reliable results for the electrostatic potential and the free solvation energy at moderate computational costs using a recommended parameter set with an inner grid spacing of 0.6 Å, an outer grid spacing of 1.75 Å, a CIR factor of 2.5 and DHLI boundary values.

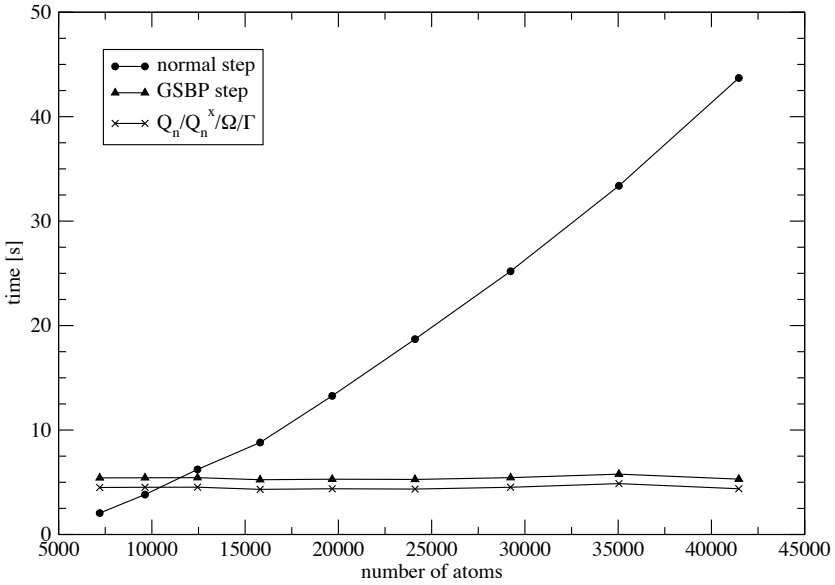


Figure 3.5: Computation times [s] for a single MD step using non-truncated Coulombic (normal step) or GSBP electrostatics (GSBP step) as a function of the system size. Furthermore, the computation times for the GSBP-related terms ($Q_n/Q_n^x/\Omega/\Gamma$) are plotted.

Table 3.5: Computation times related to non-truncated Coulombic electrostatics and the GSBP approach for different system sizes. Single step computation times are average values from a sample of 100 steps.

		Computation time [s]							
Radius [Å]	Atoms	Overhead	Normal step	GSBP step	QM saving	MM saving	$Q_n/Q_n^x/\Omega/\Gamma^a$	Saving	Steps[b]
25.0	7205	12166.1	2.0	5.4	-0.2	1.3	4.5	-3.4	-
27.5	9632	12104.2	3.8	5.4	-0.1	3.0	4.5	-1.6	-
30.0	12449	12243.9	6.2	5.4	0.0	5.3	4.5	0.8	15461
32.5	15806	12538.2	8.8	5.2	0.1	7.8	4.3	3.6	3503
35.0	19670	12399.3	13.3	5.3	0.3	12.1	4.4	8.0	1555
37.5	24110	12590.8	18.7	5.3	0.5	17.3	4.3	13.4	937
40.0	29234	12511.0	25.2	5.4	0.7	23.6	4.5	19.8	633
42.5	35042	12761.3	33.4	5.8	0.9	31.6	4.9	27.6	462
45.0	41468	12697.4	43.7	5.3	1.1	41.6	4.4	38.4	331

[a] computation of additional terms related to the GSBP

[b] number of MD steps necessary to compensate for the GSBP overhead

Concluding this section, it seems worthwhile to reiterate the reference that was used to assess the performance of the GSBP. It was confirmed that the GSBP provides an accurate representation of the electrostatic potential that arises from the fixed outer region point charges and the PDC. This does not necessarily imply that a biomolecular simulation with the GSBP will be realistic in a chemical sense. Whether application of the GSBP is reasonable, and which choice is appropriate for physical parameters like the size of the inner region or the dielectric constants, is highly system-specific and beyond the scope of this study.

3.6 GSBP Efficiency

As the application of the GSBP is linked with a significant overhead, it is of interest to quantify the computational costs and savings related to the GSBP. In this section, the efficiency of the GSBP is documented for model systems of different sizes that were generated by solving one threonine molecule in TIP3P water balls with radii increasing from 25 to 45 Å. As in the previous calculations, the inner region was centered on the C_α carbon of threonine and contains all molecules with any atom within 18 Å of the center. While the inner region consists of 2,738 atoms for all models, the overall system size increases from 7,205 atoms to 41,468 atoms with increasing radius.
A detailed analysis of the computation times for an MD simulation using either a standard approach with non-truncated Coulombic electrostatics or the GSBP is given in table 3.5 and illustrated in figure 3.5. This data provides interesting insights into the applicability and efficiency of the GSBP. First of all, the computation time for the GSBP overhead, *i.e.*, calculation of the reaction field matrix and the static field, is almost constant and increases only slightly from 3.4 h to 3.5 h when increasing the system size by a factor of 6. Also the computation time of a single GSBP MD step is almost constant at 5.4 s. For a standard MD step with full electrostatics in contrast, the computation time increases from 2.0 to 43.7 s. Accordingly, impressive savings per step can be achieved if the GSBP is used for extended systems.
However, the GSBP is not always more efficient than full electrostatics. For the two smallest systems, even a single MD step is computationally more expensive with the GSBP (in addition to the initial overhead). This can be attributed to two factors. First, with the GSBP several additional terms have to be computed for each step, such as the Ω and Γ matrices that allow interaction with the QM code, the multipole moments, Q_n, and their derivatives, Q_n^x (see equations 3.5-3.16). Especially the computation of all multipole moment derivatives for each degree of freedom is laborious and

increases the GSBP step time by roughly 4.5 s for all system sizes. Second, evaluation of the QM energy and gradient is computationally more expensive with the GSBP, since the QM part takes 0.3 s with the standard approach and 0.5 s with the GSBP. This can be traced back to the calculation of the SCF density derivatives that is not necessary in a pure QM/MM calculation. However, these factors are dominant only for small systems. With increasing system size, evaluation of the QM energy and gradient becomes more efficient in the GSBP, due to the fact that the calculation of the numerous one-electron integrals in the standard electronic embedding procedure becomes more expensive than the solution of the CPSCF equation for large systems with 12,000 atoms and more. Moreover, the introduction of coupled Mulliken charge derivatives (to ensure accurate gradients) increases the computational costs of the GSBP method only marginally. The computation time for the MM part remains constant at about 1.0 s when using the GSBP, providing the main contribution to the GSBP savings.
Overall, in the chosen example, one starts to see minor savings for a system with 12,500 atoms. Assuming an MD step size of 1 fs, the first 15 ps of simulation time are needed to compensate for the GSBP overhead, and afterwards the computation time per step is reduced by 13 %. Hence, in typical semiempirical QM/MM MD simulations, the breakeven point between the GSBP and Coulombic electrostatics without truncations appears at a system size of around 12,500 atoms. Significant savings are achieved for larger systems. In simulations of the 37.5 Å system with 24,110 atoms, only 937 steps are necessary to compensate for the GSBP overhead, and subsequently, the computation time per step decreases by more than 70 % from 18.7 to 5.3 s. For larger systems even more impressive savings are observed (table 3.5). Since in theoretical biochemistry one is frequently interested in QM/MM simulations of biomolecular systems with 25,000 atoms and more, the GSBP method offers an efficient approach to perform such simulations at a fraction of the computational costs compared to Coulombic electrostatics without truncation.

3.7 Conclusions

In this chapter, the implementation of the GSBP for QM/MM approaches using NDDO-based semiempirical QM methods is presented. Three methods to accelerate computation of the reaction field matrix were introduced: coarsening of the inner region, linear interpolation of Debye-Hückel boundary values, and modified stripping. It was found that a combination of these methods reduces the computational costs for assembling the reaction field

matrix by 60 % with only minimal loss of accuracy. Furthermore, the accuracy of the GSBP as a function of its inherent parameters was studied, and a set of parameter values that offer an ideal trade-off between accuracy and computational costs was defined. On the basis of theses values, the computational overhead and the savings of the GSBP were quantified in QM/MM MD simulations for model systems containing from around 7,000 to more than 40,000 atoms. The breakeven point where the savings in comparison to non-truncated Coulombic electrostatics roughly compensate for the overhead was determined at around 12,500 atoms. For larger systems, the GSBP showed an impressive performance. Compensation for the overhead was achieved in less than 1,000 MD steps, and subsequently, the computation time per step decreased by 70 % and more compared to non-truncated Coulombic electrostatics.

The GSBP is thus an efficient and accurate method to perform semiempirical QM/MM MD simulations on large biomolecular systems without neglecting or truncating long range electrostatics if the outer layer of the inner region is fixed. It is clearly desirable to achieve similar computational savings by applying the GSBP in combination with higher-level QM/MM methods. Work in this direction is presented in the next chapter.

Chapter 4

Solvated Macromolecule Boundary Potential

4.1 Motivation

Although semiempirical QM/MM MD simulations enable sufficient sampling of phase space, they suffer from the inaccuracies of the semiempirical QM Hamiltonian. For many systems, more accurate QM approaches such as DFT or correlated ab initio methods are necessary for reliable simulations. However, QM/MM MD simulations with such accurate QM methods are computationally very intense. Only very few examples of DFT/MM MD studies can be found in the literature and those studies used massive computational resources.[99,100] Instead, most QM/MM studies rely on investigations of the minimum energy path on the potential energy surface (PES) to explain reaction mechanisms or identify catalytically active residues. Despite the absence of configurational sampling, the PES studies permit qualitative or semi-quantitative conclusions, and have given valuable insight in the past. However, studying only a single or very few configurations is dangerous and can yield even qualitatively wrong results if non-representative configurations are selected.[101] Therefore, several schemes have been devised to approximate free energy differences for QM/MM approaches with DFT or ab initio QM methods.[14,21,102–110] The QM/MM-FEP method[82] is one of these approaches and has been explained in detail in section 2.4. Based on a PES scan, it provides an estimation of the finite-temperature effects due to the dynamics of the MM region. The validity of this approach is supported by successful applications[16,111–113] and comparisons to non-approximated free energy methods.[20,23]
These considerations motivated the development of a new general boundary

potential for QM/MM calculations that was designed to offer the following two new capabilities. First, it extends the QM/MM method to a general three-layer approach (valid for any QM treatment) that describes the outer solvent and macromolecule region by a boundary potential and thus enables an accurate treatment of long range electrostatic interactions including bulk solvent effects. Second, within the FEP framework, it allows application of the GSBP to sample the MM phase space more efficiently. Both options are available in combination with every QM/MM potential. Since this boundary potential mimics the electrostatic potential of the outer region of macromolecules in solution, it has been named solvated macromolecule boundary potential (SMBP).

4.2 Theory

The main issue of any QM/MM implementation of the GSBP is the representation of the continuous QM charge density. In previous implementations (see section 3.2),[56,114] the QM density was represented by Mulliken charges.[92] This choice has the two advantages that the working equations of the GSBP for MM methods can be easily extended to the QM/MM case, and that the interaction of the QM density with the boundary potential during the SCF procedure can be expressed in simple terms that have to be added to the Fock matrix. However, there are also two disadvantages. First, the GSBP has to be implemented for each QM program and method individually. Second, to compute accurate analytical gradients, it is necessary to calculate the derivative of the Mulliken charges which involves solution of the CPSCF equations. While the computational costs of this step are acceptable for semiempirical methods, they will increase significantly for higher level QM methods with larger AO basis sets.[114]

Hence, a different approach was devised to make use of the efficiency of the GSBP in QM/MM free energy calculations with accurate DFT or ab initio QM methods. The SMBP is developed to facilitate geometry optimizations (and single-point calculations) applying the same approximations as the GSBP. It is then possible to compute the reaction profile on the PES using a QM/MM/SMBP Hamiltonian, and to sample over MM phase space subsequently using the FEP approximation with the efficient GSBP. The details of this linkage will be explained in subsection 4.2.1.

The design of the SMBP was guided by the requirements that it should be conceptually similar to the GSBP, efficient in geometry optimizations, and applicable in QM/MM calculations with any kind of QM method.

We first consider the definition of the dielectric boundary: The core of the GSBP is the analytical expression for the electrostatic interaction with the outer region charge distribution that is shielded in a non-trivial way by the PDC. To find a closed-form expression for this potential, it is necessary to assume that the dielectric interface is fixed during the simulation.[54] Usually, in solutions of the PB equation the interface is defined by the superposition of the van der Waals envelope of the atoms. In the GSBP, a constant and smooth dielectric interface throughout dynamics simulations is ensured by extending the dielectric cavity region that encloses the inner region. For the sake of consistency, the same approach is used in the SMBP. The inner region is restricted to have a spherical shape with radius R_{inner} that comprises all inner region atoms. Since all atoms inside the sphere are modeled explicitly, the dielectric constant inside the inner region is set to 1. In the bulk solvent and the macromolecule region the dielectric constant is set to ϵ_s and ϵ_m, respectively. To secure that the shape of the interface is independent of the position of the active atoms, the radius of the inner region cavity is extended by ΔR. This value has to be chosen sufficiently large to avoid the van der Waals radius of any active atom touching the interface. The resulting shape of the dielectric interface in the SMBP is illustrated in figure 4.1. The previous evaluation of the GSBP showed that its accuracy deteriorates close to the boundary of the inner region (see subsection 3.5.1).[55,114] Therefore, it was found necessary to freeze the outer layer of the inner region, providing an "insulation" region with a thickness of 2-3 Å.[114] In figure 4.1, this is the area between the red and black line. Here, the atoms are described explicitly but their positions are fixed.

The construction of the SMBP is based on the same approximations as in the case of the GSBP. Outer macromolecule and bulk solvent regions are represented by fixed point charges and a PDC, respectively, so that the electrostatic contributions to the PMF consist of direct Coulombic interactions (U^{io}_{elec}) and the solvation free energy (ΔW^{solv}_{elec}). Again, the outer-outer contribution to the solvation free energy is constant and therefore neglected so that the SMBP takes the following form:

$$\Delta W^{SMBP}_{elec} = U^{io}_{elec} + \Delta W^{io}_{elec} + \Delta W^{ii}_{elec} \tag{4.1}$$

As in the GSBP, the electrostatic interactions of the inner region with the outer region charges (U^{io}_{elec}) and with the response of the PDC to the outer region charges (ΔW^{io}_{elec}) are combined for efficient computation (equation 2.13).

$$\Delta W^{SMBP}_{elec} = \int d\mathbf{r}\rho_i(\mathbf{r})\phi^o_s(\mathbf{r}) + \Delta W^{ii}_{elec} \tag{4.2}$$

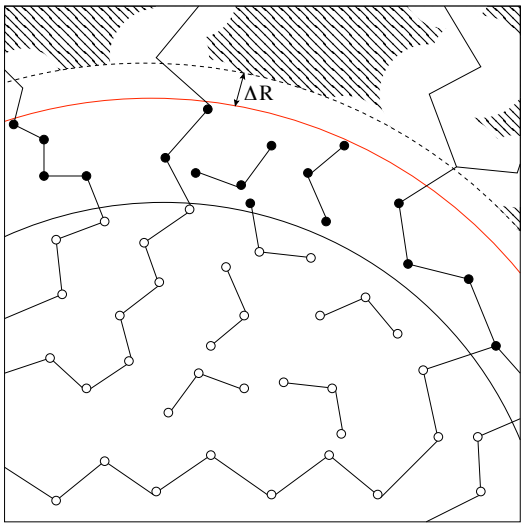


Figure 4.1: Definition of the constant dielectric interface in the SMBP and the GSBP. The extended cavity region is encircled by the dashed black line and the implicit solvent region is indicated by the hatched area. An "insulation" region of frozen explicit atoms (black circles) ensures (see text) that the dielectric interface is not touched by the van der Waals radius of any active explicit atom (white circles). The inner region and the active region are encircled by black and red lines, respectively.

In the GSBP, ΔW_{elec}^{ii} is approximated by a closed-form expression that makes use of a basis set representation of the inner region charge distribution and the reaction field Green's function (see equations 2.15 and 2.22). This approach is designed for MD simulations, but is computationally not efficient for geometry optimizations or single-point calculations. Using a standard sized basis set to represent the charge density, computation of the reaction field matrix corresponds to solving the PB equation about 800 times. Even with a large active region, geometry optimizations rarely take more than 800 steps to converge, and therefore, solving the PB equation after each step is more efficient. Since geometry optimizations are the field of application of the SMBP, a different ansatz is used and the individual contributions to the PMF are updated by solving the PB equation whenever needed. This is the main conceptual difference between SMBP and GSBP.
In the QM/MM/SMBP approach, the inner region charge distribution splits up into QM and MM charges densities, leading to a more complicated ex-

pression.

$$\begin{aligned}\Delta W_{elec} &= \frac{1}{2}\int d\mathbf{r}d\mathbf{r}'\left[\rho_{QM}(\mathbf{r})+\rho_{MM}(\mathbf{r})\right]G_{rf}(\mathbf{r},\mathbf{r}')\left[\rho_{QM}(\mathbf{r}')+\rho_{MM}(\mathbf{r}')\right] \\ &\quad + \int d\mathbf{r}\left[\rho_{QM}(\mathbf{r})+\rho_{MM}(\mathbf{r})\right]\phi_s^o(\mathbf{r}) \end{aligned} \tag{4.3}$$

This can also be formulated as the interaction with the individual potentials ϕ_{tot}^{QM} and ϕ_{tot}^{MM} that are experienced by the QM and MM charge densities, respectively:

$$\Delta W_{elec} = \int d\mathbf{r}\rho_{QM}(\mathbf{r})\phi_{tot}^{QM}(\mathbf{r}) + \int d\mathbf{r}\rho_{MM}\phi_{tot}^{MM}(\mathbf{r}) \tag{4.4}$$

with

$$\phi_{tot}^{QM}(\mathbf{r}) = \phi_s^o(\mathbf{r}) + \phi_{rf}^{MM}(\mathbf{r}) + \frac{1}{2}\phi_{rf}^{QM}(\mathbf{r}) \tag{4.5}$$

$$\phi_{tot}^{MM}(\mathbf{r}) = \phi_s^o(\mathbf{r}) + \frac{1}{2}\phi_{rf}^{MM}(\mathbf{r}) \tag{4.6}$$

Both inner region reaction field potentials, ϕ_{rf}^{QM} and ϕ_{rf}^{MM}, are computed by solving the PB equation in solution and vacuum with all charges set to zero except the explicit QM and MM charges, respectively (see equation 2.24). Since ϕ_{rf}^{QM} and ϕ_{rf}^{MM} depend on the inner region charge distributions, they have to be updated for each inner region configuration, *i.e.*, after each step in a geometry optimization.

Moreover, computation of the QM reaction field potential is exacerbated by the mutual dependence of the QM wave function and the QM reaction field potential via the QM charge density. To find a self consistent solution to the SCF and the PB equation at the same time, a doubly iterative self consistent reaction field (SCRF) scheme is employed.

In previous implementations of the GSBP for QM/MM methods, the interaction of the QM charge density was modeled by QM Mulliken charges interacting with the boundary potential. Although this leads to simple additional terms that have to be added to the Fock matrix, it also necessitates modifications to the QM programs.[56,114] In accordance with the modular philosophy of ChemShell, a different approach is used in the SMBP to describe the interaction of the QM charge density with the boundary potential. The boundary potential is projected onto a set of N virtual surface charges $\{q_i\}$ which are distributed uniformly on a sphere with radius $R_{inner} + \Delta R$ that defines the extended dielectric cavity (see figure 4.2).

$$\phi_{tot}^{QM}(\mathbf{r}) \approx \sum_i^N \frac{q_i}{|\mathbf{r}-\mathbf{r}_i|} \tag{4.7}$$

The values of the surface charges are optimized to reproduce ϕ_{tot}^{QM} at the position of the QM atoms by minimization of the penalty function $\tilde{F}$.

$$\tilde{F} = \sum_{j}^{QM} \left[\phi_{tot}^{QM}(\mathbf{r}_j) - \sum_{i}^{N} \frac{q_i}{|r_{ij}|} \right]^2 \tag{4.8}$$

The minimization of $\tilde{F}$ starts with all virtual surface charges set to zero. The charges are optimized with a conjugate gradient algorithm until ϕ_{tot}^{QM} is reproduced with a maximum absolute deviation of $2 \cdot 10^{-5}$ a.u. at the position of every QM atom. The QM wave function is optimized in the presence of the atomic charges of the inner MM region and the virtual surface charges. The surface charge projection approach has the advantage to enable the application of the SMBP in combination with every QM program that can handle external point charges.

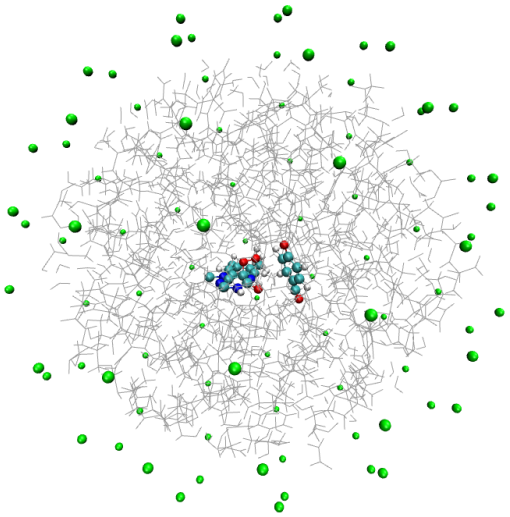


Figure 4.2: Distribution of virtual surface charges used to represent the SMBP (green balls) in the case of *p*-hydroxybenzoate hydroxylase (see subsection 4.4.2). The QM region and the explicit MM region are shown as ball-and-stick model and as grey lines, respectively.

For each new geometry, the MM reaction field potential ϕ_{rf}^{MM} is computed by neglecting all charges in the outer region and the QM region. Since the

MM charges are not polarizable and the potential is independent of the QM charge distribution, the MM reaction field does not have to be updated due to changes in the QM density. Subsequently, an SCRF procedure begins:

1. Initially, the QM reaction field potential ϕ_{rf}^{QM} is computed based on a guess for the atomic QM charges.
2. Then the total potential experienced by the QM atoms ϕ_{tot}^{QM} is assembled and projected onto a set of virtual surface charges $\{q_i\}$.
3. Next, the QM wave function is computed in the field of the inner MM region point charges and the virtual surface charges. After convergence, the QM ESP charges based on the new wave function are calculated.
4. With these new QM charges, the PB equation is solved again to update the QM reaction field potential.
5. Finally, the potential is checked for convergence. If the deviations in the QM reaction field potential are too large, the algorithm returns to step 2 and updates the wave function and the QM reaction field potential.
6. Upon convergence, the force contributions from the total potential are computed and added to the gradient.

These force contributions are the derivative of the electrostatic contribution to the PMF with respect to the atomic coordinates. For a QM atom this yields:

$$\begin{aligned}\frac{\partial}{\partial x_A}\Delta W_{elec} &= \frac{\partial}{\partial x_A}\frac{1}{2}\int d\mathbf{r}d\mathbf{r}'\rho_{QM}(\mathbf{r})G_{rf}(\mathbf{r},\mathbf{r}')\rho_{QM}(\mathbf{r}') \\ &\quad +\frac{\partial}{\partial x_A}\int d\mathbf{r}d\mathbf{r}'\rho_{QM}(\mathbf{r})G_{rf}(\mathbf{r},\mathbf{r}')\rho_{MM}(\mathbf{r}') \\ &\quad +\frac{\partial}{\partial x_A}\int d\mathbf{r}\rho_{QM}(\mathbf{r})\phi_s^o(\mathbf{r}) \qquad (4.9) \\ &= \int d\mathbf{r}\left[\frac{\partial}{\partial x_A}\rho_{QM}(\mathbf{r})\right]\phi_{tot}^{grad}(\mathbf{r}) \qquad (4.10)\end{aligned}$$

with

$$\phi_{tot}^{grad}(\mathbf{r}) = \phi_s^o(\mathbf{r}) + \phi_{rf}^{MM}(\mathbf{r}) + \phi_{rf}^{QM}(\mathbf{r}) \qquad (4.11)$$

Projection of the total gradient potential $\phi_{tot}^{QM,grad}$ onto a set of K gradient surface charges $\{q_p\}$ leads to the following approximation:

$$\frac{\partial}{\partial x_A}\Delta W_{elec} \approx \int d\mathbf{r}\left[\frac{\partial}{\partial x_A}\rho_{QM}(\mathbf{r})\right]\left[\sum_p^K \frac{q_p}{|\mathbf{r}_p-\mathbf{r}|}\right] \qquad (4.12)$$

These terms are computed and added to the QM gradient automatically by every standard quantum chemistry code if the QM gradient calculation is performed in the presence of a set of point charges that encompasses the point charges of the inner MM atoms and the gradient surface charges. For SCF wave functions, solution of the CPSCF equations is not necessary for gradient computations, since all terms involving derivatives of variationally optimized orbital coefficients are zero. In the QM/MM/SMBP approach, however, the Fock matrix in the gradient calculation is not strictly diagonal because the virtual surface charges are different in the energy and the gradient calculations (see equations 4.5 and 4.11). Thus the QM wave function is not converged in the field of the gradient surface charges. However, the QM contribution to the total gradient potential is very small, and the differences between virtual surface charges for energy and gradient calculations are therefore almost zero. By comparison to finite-difference gradient calculations it was found that all terms involving orbital coefficient derivatives can be neglected in geometry optimizations with standard convergence criteria (see table 4.1). If the dielectric constant of the solvent region is 1, *i.e.*, in case of a calculation in vacuo, all reaction field contributions are zero and the analytical gradient is exact within the QM/MM/SMBP approximation.

Table 4.1: Mean absolute (MAD) and maximum absolute deviations (MAX) of the QM gradient components [10^{-4} a.u.] of glycine in water (see subsection 4.4.1). Finite-difference gradients were used as reference.

Configuration	MAD	MAX
1	0.44	3.21
2	2.60	4.76
3	0.21	1.03
4	0.29	1.25
5	0.26	1.64

In summary, the potential projection approach offers a two-fold advantage: First, the SMBP can be used in combination with every quantum chemistry code, and second, solution of the CPSCF equation can be avoided for all practical purposes. Conceptually similar SCRF procedures have been used previously to combine pure QM[115] and hybrid QM/MM approaches[116] with implicit solvation models. The method presented in this work extends upon these approaches and employs a combination of SCRF procedure and virtual

surface charges to compute and represent a boundary potential that mimics not only the implicit solvent but also the outer macromolecule region.
For an MM atom, the derivative takes a similar form:

$$\frac{\partial}{\partial x_A}\Delta W_{elec} = \int d\mathbf{r} \left[\frac{\partial}{\partial x_A}\rho_{MM}(\mathbf{r})\right] \phi_{tot}^{grad}(\mathbf{r}) \tag{4.13}$$

As the MM charges are constant, the derivative of the MM charge distribution is just the derivative of the function that is used to distribute the MM charges onto the grid employed for solving the PB equation.[117]

4.2.1 QM/MM/GSBP-FEP Approach

The QM/MM-FEP approach (see section 2.4) is based on a discretization of the reaction into windows which are characterized by the value of a reaction coordinate ξ. The SMBP allows computation of the potential energy profile and the molecular and electronic structures of the discrete windows with the same approximations as in the GSBP. The outer region solvent molecules are represented by a PDC and the outer macromolecule by fixed point charges. The explicit atoms do not interact directly with all outer region charges but only with the potential that is induced by these charges in interaction with the PDC. This potential is computed as the finite-difference solution to the PB equation and is saved on a grid which enables massive computational savings. Therefore, the geometries, QM densities, and ESP charges that result from (constrained) geometry optimizations with the QM/MM/SMBP method can be used for sampling the free energy difference over the MM phase space with the GSBP. The concept of the combined use of QM/MM/SMBP and QM/MM/GSBP is illustrated in figure 4.3.
At this point it seems adequate to highlight the complementary nature of the approximations in QM/MM-FEP and in the GSBP. The QM/MM-FEP ansatz reduces the problem of configurational sampling with a QM/MM Hamiltonian to a sampling over MM phase space with a classical MM method. The GSBP enhances efficiency of classical MM simulations by representing the outer part of the system by a boundary potential. Hence, these two approaches complement each other and may be combined without loss of efficiency. Also in the GSBP, the QM density is represented by the ESP charges, leading to simple expressions for the QM multipole moments and the interaction with the static outer region field.

$$Q_n^{QM} = \sum_{A \in QM} q_A^{ESP} b_n(\mathbf{r}_A) \tag{4.14}$$

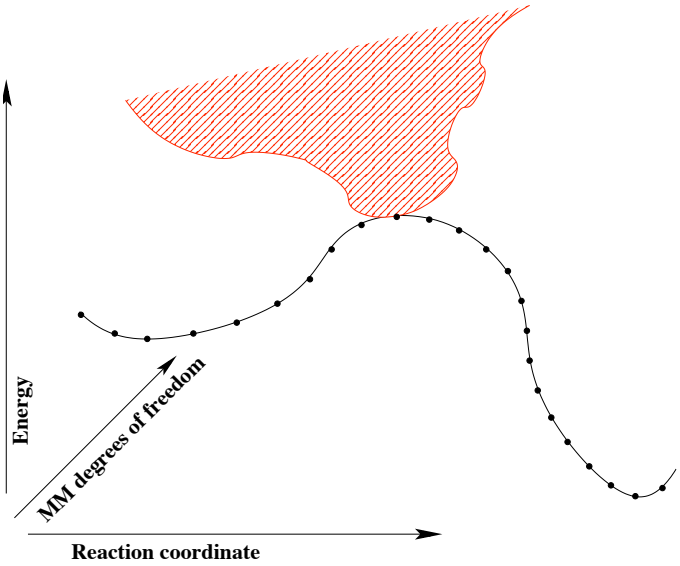


Figure 4.3: Illustration of the QM/MM/GSBP-FEP approach. The potential energy profile (black line) is calculated at the QM/MM/SMBP level of theory. The discrete windows are represented by black spheres. Sampling over MM phase space (red hashed area) is performed at the QM/MM/GSBP level.

$$\int d\mathbf{r}\rho^{QM}(\mathbf{r})\phi_s^o(\mathbf{r}) = \sum_{A\in QM} q_A^{ESP}\phi_s^o(\mathbf{r}_A) \tag{4.15}$$

Since the values and positions of the ESP charges are different for windows i and $i+1$, the GSBP will contribute to the QM/MM energy difference that is sampled in equation 2.36. As the QM atoms are fixed, computation of the QM gradient in interaction with the GSBP is not necessary. The MM gradient is calculated in analogy to other QM/MM/GSBP implementations.[54,56]

4.3 Computational Details

The SMBP was implemented in a developmental version of the modular program package ChemShell.[76,118] The energy and gradient evaluations for the QM part were performed with the MNDO[119] and Turbomole 5.7.1 programs.[120] For the MM part, the DL_POLY[97] code was employed to run the CHARMM22 force field in all calculations.[98] Stationary points were optimized in hybrid delocalized internal coordinates using the HDLCOpt optimizer.[121] The PB equation was solved with the ChemShell PB module and a maximum absolute change in every grid point of $2 \cdot 10^{-5}$ a.u. was

employed as convergence criterion. Third order B-splines were used to interpolate between the grid points.[122] The definition of the dielectric boundary was based on van der Waals radii from the CHARMM22 force field. All MD simulations were performed unter NVT conditions at a temperature of 300 K which was controlled by a Nosé-Hoover chain thermostat.[123–126] The mass of deuterium was assigned to all hydrogen atoms and free water molecules were kept rigid with SHAKE constraints.[127] A time step of 1 fs was used. The QM reaction field potential was considered converged when the root-mean-squared deviation drops below $2 \cdot 10^{-5}$ a.u. In the first iteration of the SCRF procedure, all QM atoms were assumed to be neutral.

4.4 Results

In this section, the performance of the SMBP is evaluated using three test cases: the proton transfer reaction in solvated glycine, the hydroxylation reaction in *p*-hydroxybenzoate hydroxylase (PHBH), and the spin state energy gaps in cytochrome P450cam. Glycine surrounded by explicit water molecules is a highly flexible and polar system which makes it a challenging test case: the reaction energy of the intramolecular proton transfer is sensitive to the description of the solvent, and many solvation models incorrectly predict the neutral form to be more stable than the zwitterionic form.[128–131] The hydroxylation reaction in the catalytic cycle of PHBH has been much studied theoretically[11,12,22,132–134] and has become a prototypical test system for benchmarking theoretical treatments of enzymatic reactions.[135] The relative spin state energies of cytochrome P450cam[136,137] provide another, rather different test case: the pentacoordinated ferric complex is addressed whose spin state energies are strongly affected by the protein environment.[138] Taken together, three diverse systems were studied to evaluate the accuracy and range of applicability of the SMBP.
Previous studies indicate that it may sometimes be important to allow fluctuations in the number of solvent molecules in approaches based on inner regions of fixed size.[139,140] This should not be problematic in the present test calculations which address localized events at the center of the inner region (using a fixed number of solvent molecules).

4.4.1 Glycine in Water

The glycine/water model system was set up using a commonly applied protocol of solvation and equilibration steps by means of classical MD simulations with the CHARMM program.[141] The glycine molecule was solvated in a

TIP3P water ball with 30 Å radius, all water molecules with an oxygen atom within 2.8 Å of any glycine atom were deleted, and the system was equilibrated. These steps were repeated until the number of water molecules was stable, leading to a total system size of 12,769 atoms with 4,253 TIP3P water molecules. Finally, the system was equilibrated by means of a 500 ps classical MD simulation, and five configurations were selected after 340, 380, 420, 460, and 500 ps. For each configuration, the inner region was centered on the C_α carbon of the glycine and defined to encompass all water molecules with any atom within 18 Å of the center. In all subsequent QM/MM geometry optimizations, the glycine molecule and all water molecules with any atom within 14 Å of the center were allowed to move, while all other water molecules were frozen. The radius of the extended dielectric cavity was set to 21 Å, and a set of 90 virtual surface charges was used to represent the boundary potential in the QM calculations. The glycine molecule was described quantum mechanically with the AM1 Hamiltonian,[96] and the water molecules were treated by the force field or the SMBP. The details of this setup are summarized in table 4.2.

Table 4.2: Details of the configurations of the glycine in water system.

Configuration	ps of MD	Active region	Inner region	Outer region
1	340	1402	2890	9879
2	380	1405	2905	9864
3	420	1396	2902	9867
4	460	1357	2863	9906
5	500	1378	2827	9942

In vacuum environment, the electrostatic potential of the SMBP has to be identical to the exact potential from Coulombic electrostatics. Therefore, the accuracy of the QM/MM/SMBP approach is again evaluated in vacuo by direct comparison to standard QM/MM calculations. To allow the use of finely spaced grids in the finite-difference solution of the PB equation also for large biomolecules, the focusing approach is employed.[95] The PB equation is first solved with a coarse outer grid that covers the full biomolecule. Then the PB equation is solved again with a fine inner grid that focuses onto the inner region. The boundary values of the inner grid are set by interpolation from the outer grid. The spacing of the two grids are the most important parameters of the SMBP. Hence, the accuracy of the SMBP was evaluated

for all mesh size combinations of 0.15, 0.25, 0.4, 0.6, and 0.8 Å for the inner grid and 0.80, 1.25, 1.50, 1.75, and 2.50 Å for the outer grid.

Table 4.3: Mean absolute deviations (MAD) [10^{-4} a.u.] of the electrostatic forces computed with the SMBP for configuration 1 of the glycine/water system. Different mesh size combinations were used. Reference values were computed with the exact Coulombic potential.

	Inner grid size [Å]				
Outer grid size [Å]	0.15	0.25	0.40	0.60	0.80
MAD - QM atoms					
0.80	0.31	0.31	0.31	0.32	0.31
1.25	0.31	0.31	0.31	0.31	0.31
1.50	0.35	0.32	0.33	0.33	0.33
1.75	0.33	0.32	0.32	0.32	0.32
2.50	0.33	0.34	0.34	0.34	0.33
MAD - atoms within 16 Å					
0.80	0.18	0.10	0.15	0.20	0.26
1.25	0.21	0.15	0.19	0.23	0.28
1.50	0.28	0.21	0.26	0.29	0.34
1.75	0.26	0.19	0.24	0.27	0.32
2.50	0.29	0.23	0.28	0.30	0.35
MAD - atoms within 20 Å					
0.80	0.23	0.27	0.57	1.02	1.42
1.25	0.28	0.32	0.61	1.04	1.44
1.50	0.37	0.38	0.67	1.09	1.48
1.75	0.34	0.36	0.65	1.08	1.46
2.50	0.38	0.39	0.68	1.10	1.48

Tables 4.3 and 4.4 show the mean absolute (MAD) and maximum absolute deviations (MAX) of the components of the electrostatic gradient for configuration 1. Similar deviations were observed for the other configurations (see tables 4.5 and 4.6). Although only 90 virtual surface charges are used to represent the electrostatic potential that is induced by almost 10,000 atoms in the outer region, the electrostatic gradient at the position of the QM atoms is reproduced with high accuracy. For all mesh size combinations, the MAD and MAX values are on the order of $0.3 \cdot 10^{-4}$ a.u. and $1.6 \cdot 10^{-4}$ a.u., respectively. Moreover, the accuracy seems to be independent of the grid spacing

Table 4.4: Average maximum absolute deviations (MAX) [10^{-4} a.u.] of the electrostatic forces computed with the SMBP for configuration 1 of the glycine/water system. Different mesh size combinations were used. Reference values were computed with the exact Coulombic potential.

	Inner grid size [Å]				
Outer grid size [Å]	0.15	0.25	0.40	0.60	0.80
MAX - QM atoms					
0.80	1.53	1.64	1.57	1.55	1.60
1.25	1.62	1.69	1.65	1.65	1.67
1.50	1.62	1.66	1.64	1.64	1.66
1.75	1.59	1.65	1.61	1.61	1.64
2.50	1.57	1.63	1.60	1.60	1.62
MAX - atoms within 16 Å					
0.80	1.53	1.64	3.72	7.16	9.08
1.25	1.67	1.90	3.68	7.08	8.89
1.50	3.22	2.90	3.91	7.30	9.45
1.75	2.32	2.44	3.73	7.14	9.02
2.50	3.15	3.06	3.88	7.19	9.51
MAX - atoms within 20 Å					
0.80	2.41	5.43	14.79	35.67	35.35
1.25	2.91	5.57	14.59	35.46	35.27
1.50	4.93	5.79	14.53	35.42	34.71
1.75	3.60	5.64	14.52	35.39	35.03
2.50	4.67	5.51	14.25	35.11	34.81

Table 4.5: Mean absolute deviations (MAD) [10^{-4} a.u.] of the electrostatic forces computed with the SMBP averaged over five configurations of the glycine/water system. Different mesh size combinations were used. Reference values were computed with the exact Coulombic potential.

	Inner grid size [Å]				
Outer grid size [Å]	0.15	0.25	0.40	0.60	0.80
MAD - QM atoms					
0.80	0.23	0.22	0.22	0.22	0.22
1.25	0.25	0.25	0.25	0.24	0.24
1.50	0.30	0.29	0.29	0.29	0.28
1.75	0.28	0.27	0.27	0.27	0.26
2.50	0.29	0.29	0.29	0.28	0.28
MAD - atoms within 16 Å					
0.80	0.14	0.12	0.14	0.19	0.26
1.25	0.20	0.20	0.20	0.24	0.30
1.50	0.29	0.26	0.28	0.31	0.36
1.75	0.25	0.24	0.25	0.28	0.33
2.50	0.30	0.27	0.29	0.32	0.37
MAD - atoms within 20 Å					
0.80	0.21	0.29	0.56	0.99	1.41
1.25	0.27	0.37	0.62	1.04	1.44
1.50	0.38	0.43	0.68	1.09	1.48
1.75	0.33	0.41	0.65	1.07	1.46
2.50	0.39	0.45	0.69	1.10	1.49

within the chosen limits. Both findings suggest that the static outer region potential varies only slowly and has no detailed structure in the QM region. Considering all atoms within 16 Å of the center, the SMBP reproduces the gradient of the electrostatic potential with high accuracy if the spacing of the inner grid is ≤ 0.4 Å. Under these conditions the MAX values are below $4 \cdot 10^{-4}$ a.u. The spacing of the outer grid does not influence the accuracy unless very fine inner grids are used. As one approaches the boundary separating inner and outer region, the static outer region potential naturally becomes stronger and more complex. Nevertheless, its details are captured with sufficient accuracy also at the position of all inner region atoms which have a distance of up to 20 Å to the center due to the residue-based selection criterion that was employed to define the inner region. With an inner

Table 4.6: Average maximum absolute deviations (MAX) [10^{-4} a.u.] of the electrostatic forces computed with the SMBP averaged over five configurations of the glycine/water system. Different mesh size combinations were used. Reference values were computed with the exact Coulombic potential.

	Inner grid size [Å]				
Outer grid size [Å]	0.15	0.25	0.40	0.60	0.80
MAX - QM atoms					
0.80	1.11	1.10	1.10	1.12	1.12
1.25	1.22	1.28	1.27	1.27	1.28
1.50	1.51	1.49	1.52	1.53	1.51
1.75	1.34	1.34	1.35	1.35	1.35
2.50	1.46	1.46	1.48	1.48	1.47
MAX - atoms within 16 Å					
0.80	1.39	1.72	3.14	6.25	9.84
1.25	1.83	2.10	3.31	6.37	10.02
1.50	2.55	2.26	3.04	6.28	10.02
1.75	2.15	2.23	3.33	6.37	10.06
2.50	2.41	2.46	3.19	6.34	10.06
MAX - atoms within 20 Å					
0.80	2.58	5.89	13.77	25.93	36.16
1.25	3.03	5.90	13.72	25.79	36.04
1.50	4.49	5.96	13.74	25.80	35.88
1.75	3.53	5.87	13.71	25.75	35.98
2.50	4.34	5.97	13.68	25.85	35.93

grid spacing of ≤ 0.25 Å, the MAD values do not exceed $0.4 \cdot 10^{-4}$ a.u. and maximum deviations are around $5.5 \cdot 10^{-4}$ a.u.

The accuracy of the SMBP depends strongly on the radial position of the atoms since the electrostatic potential is more complex at the boundary. Figure 4.4 illustrates this point and shows that the increase of the MAD and MAX values in proximity to the boundary is strongly affected by the mesh size of the inner grid. The deviations increase only slowly with grid spacings of 0.15 or 0.25 Å, and much more rapidly for coarser inner grids.
Overall, the accuracy provided by the SMBP is sufficient for QM/MM geometry optimizations where the default convergence criterion is a maximum gradient component of $4.5 \cdot 10^{-4}$ a.u.[121] Therefore, an inner grid spacing of

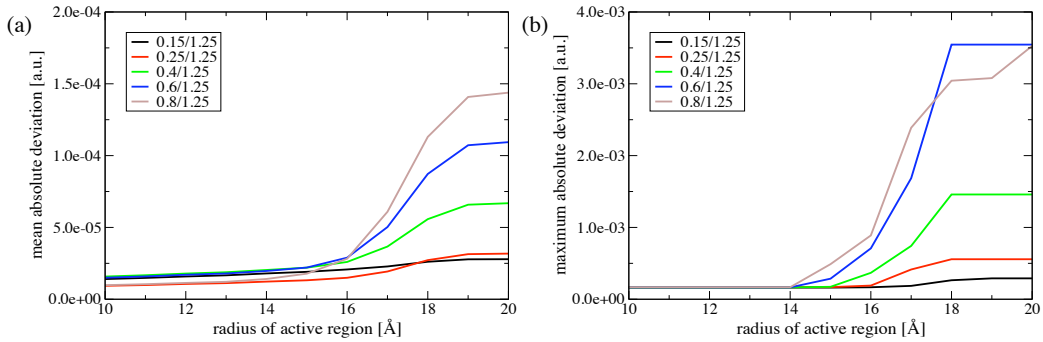


Figure 4.4: Mean absolute deviation (a) and maximum absolute deviation (b) of the electrostatic forces of all atoms inside the active region relative to the exact QM/MM values. Results are shown for different mesh sizes of the inner grid and plotted as a function of the radius of the active region. An outer grid size of 1.25 Å is used and all calculations were performed on configuration 1 of the glycine/water test system. The radius of the inner region was 18 Å (see text).

0.25 Å excels as best choice providing high accuracy at tolerable computational costs. Using a finer grid spacing of 0.15 Å seems to yield only marginal improvements but raises computational demands significantly. The mesh size of the outer grid has no observable influence on the accuracy. Since computational costs are only slightly affected by the outer grid spacing, a rather fine outer grid with a mesh size of 1.25 Å was selected in combination with an inner grid spacing of 0.25 Å in all calculations in this chapter.

Representation of the boundary potential by a small set of point charges in the QM calculations is one of the main approximations connected with the SMBP. The accuracy converges rapidly with respect to the number of point charges as illustrated in figure 4.5. The MAD and MAX deviations of the gradient components are around $0.3 \cdot 10^{-4}$ and $1.7 \cdot 10^{-4}$ a.u., respectively, if the number of point charges is greater than 20. The residual error is not caused by the point charge representation but results from the limited accuracy of the boundary potential which is computed from finite-difference solution of the PB equation. Similar deviations are encountered also for the MM atoms within 16 Å of the center that interact directly with the boundary potential without a charge representation (see tables 4.3 and 4.4). Hence, higher accuracies can only be achieved with finer mesh sizes and not with a higher number of virtual surface charges. A set of 90 point charges was employed for all calculations reported in this chapter.

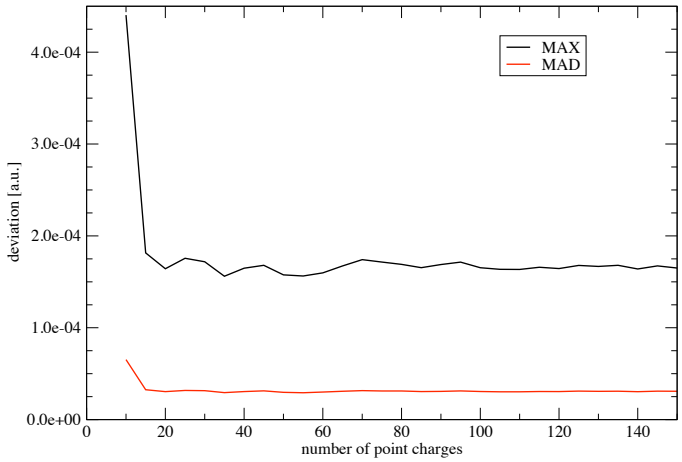


Figure 4.5: Mean absolute (MAD) and maximum absolute deviations (MAX) of the QM gradient components relative to the exact values from full QM/MM calculations. The results are plotted as a function of the number of point charges that are used to represent the boundary potential in the QM calculations. All calculations were performed on configuration 1 of the glycine/water test system.

Using electronic embedding, computation of QM/MM energies and gradients necessitates evaluation of numerous one-electron integrals and their derivatives with respect to the position of the MM atoms. Therefore, computation of these terms constitutes a significant share of the total computational costs of the QM calculation. When applying the SMBP, the numerous outer MM atoms are replaced by a small set of point charges that reproduces the electrostatic potential in the QM region. In vacuum environment, the analytical QM/MM/SMBP gradient is exact and the additional costs of the SCRF procedure can be avoided. Hence, application of SMBP in vacuo offers a reduction of computational costs for QM/MM geometry optimizations. As in standard QM/MM calculations, the bulk solvent is then modeled by fixed water molecules which contribute to the static outer region field (ϕ_s^o in equation 4.2). Table 4.7 shows the computation times for the QM part of single-point QM/MM and QM/MM/SMBP energy and gradient evaluations. Timings were performed for three QM methods: the semiempirical AM1 method, the pure density functional BLYP,[142,143] and the hybrid density functional B3LYP.[144] Two different basis sets were employed in the DFT calculations: the small SVP[145] and the larger TZVPP[146] basis sets. If the QM calculation is not dominated by the two-electron part, *i.e.*, if semiempirical methods,

Table 4.7: Computation time for a single QM energy and gradient evaluation. All timings were computed serially on 2.6 GHz AMD Opteron machines with 16 GB of memory.

		Computation Time [s]		
Basis	QM Method	QM/MM	QM/MM/SMBP	Saving [%]
	AM1	0.5	0.2	-57
SVP	BLYP	76.8	30.5	-60
	B3LYP	87.0	39.5	-54
TZVPP	BLYP	303.1	126.1	-58
	B3LYP	542.5	370.5	-31

pure functionals, or small basis sets are employed, application of the SMBP can reduce computational costs by up to 60 %. Even for the hybrid B3LYP functional with a larger basis set, computation time is reduced by about 30 %. Computational savings strongly depend on the size of the QM region, the inner region and the outer region, and on the QM method employed, and can vary significantly for different systems.

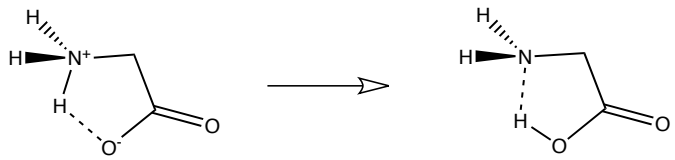


Figure 4.6: Intramolecular proton transfer reaction in glycine.

The reaction and activation energies for the intramolecular proton transfer process in glycine (figure 4.6) were computed using the standard QM/MM and the new QM/MM/SMBP Hamiltonian for the five different configurations. The results in table 4.8 show little agreement of QM/MM and QM/MM/SMBP results for the individual configurations. For configurations 3 and 5 deviations of the QM/MM/SMBP results from the QM/MM values are on the order of 1 kcal/mol. Higher deviations of reaction and activation energies up to 6 kcal/mol are encountered for the other configurations. These strong discrepancies can be attributed to the high flexibility and polarity of the system. A closer inspection of the reactant structures revealed that a small number of water molecules at the boundary of the active region adopt

a different orientation in the QM/MM/SMBP optimized structures. Due to the hydrogen bonding network, some of these modifications get relayed to the center of the water sphere and modify the hydrogen bonding situation in close proximity to the QM region. For this reason, geometry optimizations of the starting structures with the QM/MM and QM/MM/SMBP approach lead to different local minima. Since the relative energies depend on the solvation of the polar groups of the reactant and product state, the reaction energies vary significantly when starting from different local minima. However, when computations of the reaction profiles are initiated from the same local minimum, *i.e.*, by using QM/MM optimized geometries as input structures for QM/MM/SMBP geometry optimizations, both methods provide virtually identical results (see table 4.8). This is not practical in applications where QM/MM/SMBP should be used for geometry optimizations, but shows that QM/MM/SMBP can reproduce QM/MM results accurately. In summary, for systems with a large number of close-lying local minima that have significantly different characteristics, geometry optimizations using QM/MM and QM/MM/SMBP can yield deviating results because of convergence to different local minima.

The mean values of the reaction and activation energies for the five configurations considered differ by less than 1 kcal/mol between QM/MM and QM/MM/SMBP. Moreover, the mean values from QM/MM/SMBP calculations lie within the error bars of the QM/MM mean values (corresponding to a confidence level of 68 %) while the standard deviations within the individual data sets range from 3-6 kcal/mol. One may expect in general that the mean values of interest from QM/MM and QM/MM/SMBP optimizations will tend to approach each other for a sufficiently large number of configurations.

Table 4.8: Reaction and activation energies for the proton transfer reaction in solvated glycine.

	Reaction Energies [kcal/mol]			Activation Energies [kcal/mol]		
Configuration	QM/MM	QM/MM/SMBP	QM/MM/SMBP(opt)[a]	QM/MM	QM/MM/SMBP	QM/MM/SMBP(opt)[a]
1	6.84	8.42	6.88	30.01	30.09	30.01
2	9.54	15.61	9.61	32.28	36.43	32.31
3	6.51	7.68	6.61	26.82	27.10	26.84
4	7.27	1.39	7.28	31.18	27.92	31.14
5	13.57	14.58	13.53	32.66	33.35	32.62
mean value	8.75	9.54	8.78	30.73	31.20	30.58
std. dev. of data[b]	2.94	5.77	2.91	2.69	4.46	2.34
std. dev. of mean[c]	1.32	2.58	1.31	1.20	1.99	1.04

[a] starting from QM/MM optimized structures

[b] standard deviation of individual energy values

[c] standard deviation of the mean value (68 % confidence limit)

4.4.2 *p*-Hydroxybenzoate Hydroxylase

The setup for PHBH was based on a system that has been used in previous QM/MM studies of PHBH.[11,12,22] It was generated by solvating the enzyme (394 amino acids) containing the flavin-adenine hydroperoxo cofactor (FADHOOH), the dianionic *p*-hydroxybenzoate substrate (pOHB), and 294 crystallographic water molecules in a 90 Å water box. The system was equilibrated with gradually decreasing harmonic restraints on the non-water atoms, followed by an MD run with harmonic restraints acting only on the FADHOOH and pOHB. In the resulting structure that served as starting point for this setup, all water molecules outside 11 Å from any protein atom were discarded.[11,12,22]
Due to a change of force field from GROMOS (previously) to CHARMM (this study), the system was re-equilibrated for 500 ps with constraints on the cofactor, substrate, and all water molecules outside 2.9 Å from any protein atom. Two configurations were selected from this MD run after 460 and 500 ps that were used as starting structures to locate the stationary points of the hydroxylation reaction. The QM region consisted of pOHB and the isoalloxazine part of FADHOOH up to the first methylene unit of the ribityl side chain that was saturated with a hydrogen link atom. The semiempirical AM1 Hamiltonian was employed to describe the QM part. The inner region was centered on the initial position of the distal oxygen atom of the hydroperoxo group of FADHOOH. All charge groups with any atom within 18.5 Å of the center belonged to the inner region and were modeled explicitly. All charge groups with any atom within 16 Å of the center belonged to the active region and were allowed to move. The radius of the extended dielectric cavity was set to 22.5 Å, and a set of 90 virtual surface charges was used to represent the boundary potential in the QM calculations.

In the hydroxylation step of the PHBH catalytic cycle, the OH unit of the hydroperoxo group of FADHOOH is transferred to the *meta* carbon atom of pOHB (see figure 4.7). To compute the potential energy profile and split the reaction into discrete windows for the FEP calculations, a reaction coordinate was defined:

$$\zeta = d(O_d - O_p) - d(C_m - O_d) \tag{4.16}$$

Here, O_d and O_p designate the distal and proximal oxygen atoms of the hydroperoxo unit of FADHOOH, respectively. C_m is the *meta* carbon atom of pOHB. Starting from the two initial structures, the stationary points of this reaction were located using the QM/MM and the QM/MM/SMBP Hamiltonian. Both methods yield similar results (table 4.9). For configuration 1, QM/MM and QM/MM/SMBP geometry optimizations lead to slightly dif-

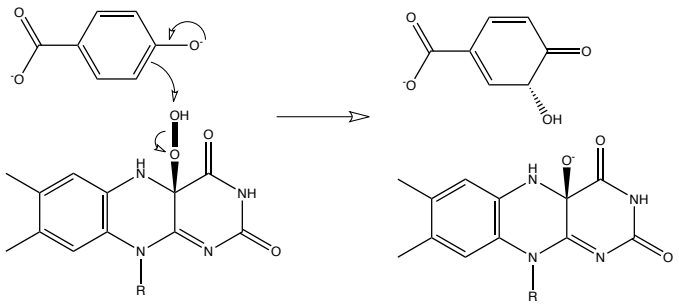


Figure 4.7: Hydroxylation reaction catalyzed by *p*-hydroxybenzoate hydroxylase. R denotes the ribityl side chain of the hydroperoxo flavine-adenine cofactor.

ferent local minima as indicated by a root-mean-square (RMS) deviation (of the active atoms) of 5.4 pm. However, the reaction and activation energies deviate by only 1.2 and 0.9 kcal/mol, respectively. These differences are in the same range as the differences between the two configurations on the pure QM/MM level. For configuration 2, both Hamiltonians lead to the same local minimum with a RMS deviation of 0.8 pm. Hence, the reaction and activation energies differ by only 0.4 and 0.1 kcal/mol, respectively. The plot of the potential energy profile in figure 4.8 illustrates this impressive agreement.

Table 4.9: Potential and free energies of activation and reaction of the hydroxylation reaction in *p*-hydroxybenzoate hydroxylase [kcal/mol].

Configuration	Hamiltonian	ΔE	$\Delta E^{\ddagger}$	ΔA	$\Delta A^{\ddagger}$
1	QM/MM	-47.44	22.42	-50.38	21.27
	QM/MM/BP[a]	-48.65	21.56	-50.47	20.40
2	QM/MM	-48.90	22.00	-51.28	19.49
	QM/MM/BP[a]	-49.33	21.95	-52.84	19.17

[a] Outer macromolecule region is represented by SMBP in geometry optimizations and by GSBP in FEP calculations.

The optimized geometries of the discrete windows along the reaction coor-

dinate were used as input structures to sample over the MM phase space in the framework of FEP. For each window, the QM atoms were fixed and the QM charge density was approximated by constant ESP charges. For the structures that were optimized with the QM/MM/SMBP Hamiltonian, the GSBP could be applied to accelerate the MD steps. To avoid mobile water molecules or flexible residues approaching the boundary separating inner and outer region, the size of the active region was reduced in the QM/MM/GSBP calculations. Here, all atoms within 15 Å of the center were allowed to move. Moreover a spherical restraint with a radius of 17 Å and a force constant of 0.004 a.u. was applied to all active atoms to avoid that any mobile residue leaves the inner region. For both Hamiltonians, the molecular structure in each window was equilibrated for 10 ps followed by an FEP production run of 10 ps. The resulting MD data was coarse-grained and subjected to a standard set of statistical tests to ensure lack of trend and correlation.[23] If necessary, data points at the beginning of the production run were discarded (at most 4 ps so that production runs lasted at least 6 ps for each window).

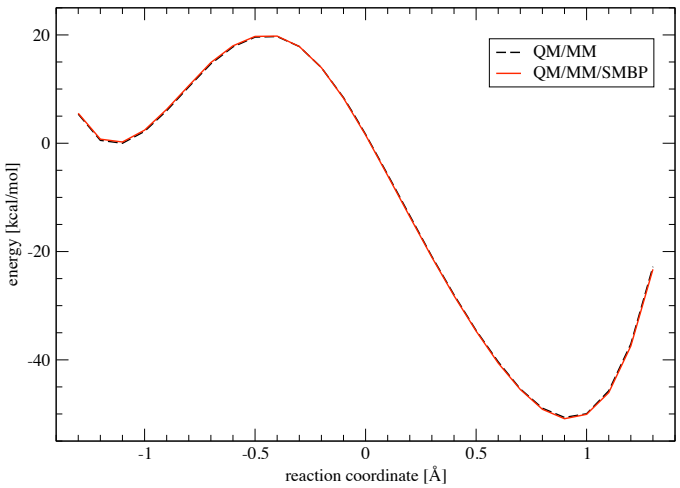


Figure 4.8: Potential energy profile of the OH transfer reaction in *p*-hydroxybenzoate hydroxylase computed with QM/MM and QM/MM/SMBP independently (configuration 2). The QM atoms are described by the AM1 method and the MM atoms by the CHARMM force field.

For configuration 1, the free energies of activation and reaction deviate by 0.9 and 0.1 kcal/mol, respectively (table 4.9). The results for configuration 2 are similar with deviations of 0.3 and 1.6 kcal/mol, respectively. Figure 4.9 illustrates that the free energy profiles computed with and without GSBP

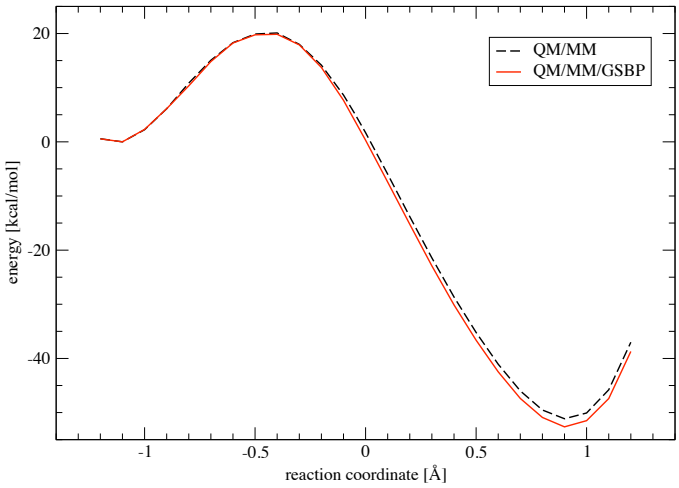


Figure 4.9: Free energy profile of the OH transfer reaction in *p*-hydroxybenzoate hydroxylase computed with QM/MM and QM/MM/GSBP independently (configuration 2, excluding entropic QM contributions). The QM atoms are described by the AM1 method and the MM atoms by the CHARMM force field.

Table 4.10: Computation time per MD step [s] in FEP simulations of the *p*-hydroxybenzoate hydroxylase model system. All values are averaged over 200 MD steps and were computed serially on 2.6 GHz AMD Opteron machines with 16 GB of memory.

Module	QM/MM	QM/MM/GSBP
MM energy+gradient	69.2	2.6
QM energy+gradient	0.4	0.0
FEP	46.6	2.2
GSBP	-	0.1
total	116.2	4.9

agree well for all stages of the reaction. In view of the other approximations that are necessary for QM/MM-FEP simulations, the deviations caused by the GSBP seem small and tolerable. Application of the GSBP offers massive computational savings. In the PHBH system, the computational time for a single MD step of the FEP calculation is reduced by 95 % from 116.2 s to only 4.9 s (table 4.10). Even when taking the GSBP overhead into account, the computational costs of QM/MM/GSBP-FEP calculations are

roughly one order of magnitude smaller than those of standard QM/MM-FEP calculations. In modeling enzymatic reactions, the combination of SMBP and GSBP for geometry optimizations and FEP calculations thus offers good accuracy combined with high efficiency.

4.4.3 Cytochrome P450cam

A recent QM/MM study addressed the steric and electrostatic factors that affect the geometrical and electronic structure of the pentacoordinated cytochrome P450cam complexes.[138] Among other properties, the energies of the doublet and quartet states relative to the sextet state were found to be strongly influenced by the protein environment. Therefore, the spin state energy gaps of the ferric complex were chosen as a challenging protein test system to see if the SMBP is able to reproduce such subtle energy differences.

The system setup was based on the x-ray structure of the ferric complex (PDB code 1DZ4)[147] that was subjected to a standard solvation and relaxation protocol, followed by a protonation procedure that led to a final system with a total charge of -9e. The calculations started from the two structures selected previously from a classical MD simulation after 31 and 93 ps.[138] For the QM calculations, three different density functionals were used that combine Becke's B88 exchange functional[142] with a varying fraction of Hartree-Fock (HF) exchange and the Lee-Yang-Parr (LYP) correlation functional:[143] BLYP, B3-LYP,[144] and BHLYP.[148] A 6-31+G* basis was used for all atoms coordinated to the iron and the 6-31G basis set for the remaining ligand atoms. The iron atom was described by a Wachters all-electron basis set with additional sets of diffuse d and polarizing f functions.[149–151]

The QM region included the iron-porphyrin system of the heme unit and the sulfur atom of the coordinating Cys357 (with a hydrogen link atom attached to sulfur). The atoms of all residues with any atom within 4 Å of the heme-Cys357 complex or the camphor substrate were allowed to move, all other atoms were frozen. In the QM/MM/SMBP calculations, the inner region was centered on the initial position of the iron atom with an extended dielectric cavity radius of 29 Å. All residues with any atom within 20 Å of the iron atom belonged to the inner region and were described explicitly. The influence of all other residues was mimicked by the boundary potential which was represented by 90 virtual surface charges in the QM calculations.

The initial structures were first optimized in the sextet state with the QM/MM and the QM/MM/SMBP Hamiltonian, respectively. The two resulting geometries were subsequently re-optimized in the quartet and doublet state.

Table 4.11: Sextet-quartet and sextet-doublet energy gaps in cytochrome P450cam [kcal/mol] computed with a standard QM/MM and the approximated QM/MM/SMBP Hamiltonian for two configurations and three different density functionals (BLYP, B3LYP, BHLYP).

			BLYP			B3LYP			BHLYP		
Snapshot	Hamiltonian	Gap	QM/MM	QM	MM	QM/MM	QM	MM	QM/MM	QM	MM
31	QM/MM	E(^4A)-E(^6A)	-9.50	-11.31	1.81	0.83	-0.71	1.54	16.49	14.40	2.09
		E(^2A)-E(^6A)	-9.98	-9.74	-0.24	6.94	7.11	-0.18	28.80	28.92	-0.12
	QM/MM/SMBP	E(^4A)-E(^6A)	-9.63	-10.89	1.26	0.82	-0.74	1.56	16.50	14.42	2.08
		E(^2A)-E(^6A)	-10.19	-9.80	-0.39	7.10	7.00	0.10	28.70	28.34	0.36
93	QM/MM	E(^4A)-E(^6A)	-9.52	-11.00	1.48	0.73	-0.98	1.72	16.40	13.69	2.72
		E(^2A)-E(^6A)	-10.66	-10.49	-0.17	6.33	6.75	-0.43	28.74	28.96	-0.22
	QM/MM/SMBP	E(^4A)-E(^6A)	-9.54	-10.62	1.08	0.83	-0.83	1.67	16.34	13.69	2.65
		E(^2A)-E(^6A)	-10.70	-10.32	-0.39	6.44	6.36	0.08	28.66	29.00	-0.35

Table 4.12: Root-mean-square (RMS) deviations of the QM/MM and QM/MM/SMBP optimized structures of cytochrome P450cam in Å. Frozen atoms are not considered.

Snapshot	State	BLYP	B3LYP	BHLYP
31	^{6}A	0.002	0.063	0.000
	^{4}A	0.007	0.061	0.000
	^{2}A	0.002	0.064	0.009
93	^{6}A	0.013	0.010	0.001
	^{4}A	0.012	0.005	0.011
	^{2}A	0.010	0.006	0.014

In the original study,[138] the B3LYP functional was found to provide the most realistic description of the spin state splittings with the correct sextet ground state although the doublet-sextet gap is overestimated. The objective of this section is not the absolute quality of the QM/MM results but the ability of the QM/MM/SMBP approach to reproduce the QM/MM results. The values for the spin state energy gaps are given in table 4.11. For eight out of twelve energy gaps, the QM/MM/SMBP results are within 0.1 kcal/mol of the full QM/MM results. The maximum absolute deviation is 0.21 kcal/mol, and the absolute deviations of the individual QM and MM components rarely exceed 0.4 kcal/mol. Moreover, subtle effects are reproduced very well: Using the B3LYP functional, the QM calculation (in the field of the protein point charges) favors the quartet state over the sextet state by 0.71 kcal/mol in configuration 31. This preference is over-compensated by the MM contribution that favors the sextet state by 1.54 kcal/mol; leading to a QM/MM energy difference of 0.83 kcal/mol. In the QM/MM/SMBP calculations both components are reproduced almost exactly and sum up to an energy gap of 0.82 kcal/mol in favor of the sextet state. The results for configuration 93 are similar. This data shows that the SMBP reproduces the electrostatic effects of the protein environment onto the QM and MM regions accurately, implying that geometry optimizations with the QM/MM/SMBP approach lead to highly similar local minima on the potential energy surface compared to standard QM/MM optimizations. This is corroborated by a direct comparison of the QM/MM and QM/MM/SMBP optimized geometries in table 4.12: The RMS deviations (of the active atoms) are usually around 1 pm or less. Only for the B3LYP optimized structures of configuration 31, there are larger RMS deviations of about 6 pm. These can be traced back to a 30° rotation of a methyl group attached to the porphine ring in a hydrophobic

environment. However, the corresponding spin state gaps are not affected by this peripheral conformational change and match almost perfectly.
In summary, optimizations with the QM/MM/SMBP Hamiltonian lead to biomolecular structures that are either almost identical to those from full QM/MM optimizations, or they represent nearby local minima which are as representative for the molecular and electronic structure of the biomolecule as those resulting from standard QM/MM optimizations.

4.5 Conclusion

In this chapter, a general boundary potential (SMBP) for hybrid QM/MM calculations is introduced that complements the previously implemented boundary potential (GSBP). Both the SMBP and the GSBP extend the QM/MM approach to a three-layer model in which the outer solvent molecules and outer macromolecule region are represented by a boundary potential. Therefore, both account for the effect of bulk solvent and treat long range electrostatic interactions accurately and efficiently. In both cases, the reaction field potential in the inner region needs to be computed by finite-difference solution of the Poisson-Boltzmann equation (describing the bulk solvent as a dielectric continuum). In the GSBP scheme, this inner reaction field potential is expressed by its Green's function and represented by a reaction field matrix that is determined once and for all at the beginning of a simulation and is then used to calculate the corresponding electrostatic interactions with the inner region charge density. In the SMBP scheme, the inner reaction field potential is computed on-the-fly as needed, and the interactions with the QM density are handled by a self consistent reaction field procedure and a set of virtual surface charges that represent the SMBP in the QM calculations.
The GSBP performs best in MD simulations where the initial overhead for constructing the reaction field matrix (typically about 800 Poisson-Boltzmann calculations) is quickly overcompensated by the gains in each of the many steps during the MD simulation. The SMBP targets single-point calculations and geometry optimizations with a limited number of steps where the on-the-fly approach is most efficient. Since the approximations in the GSBP and SMBP treatments are very similar by design, and compatible with each other, the electronic and molecular structures resulting from QM/MM/SMBP geometry optimizations can be used as starting points for sampling over MM phase space using the QM/MM/GSBP Hamiltonian in the FEP framework. Free energy calculations on the PHBH enzyme show that this reduces the computational costs of the FEP calculations by one order of magnitude. The combined use of the SMBP and GSBP for computing potential energy pro-

files and subsequent sampling, respectively, thus provides an attractive and efficient strategy to perform free energy QM/MM calculations.
The GSBP implementation at the QM/MM level requires modifications of the underlying QM code, and corresponding work has been reported up to now only for semiempirical QM methods.[56,114] By contrast, because of its representation in terms of virtual surface charges, the SMBP can be used with any standard QM code that can handle external point charges, thus facilitating ab initio QM/MM/SMBP and DFT/MM/SMBP geometry optimizations in the context of three-layer QM/MM/continuum models. Another practical advantage of the SMBP is that it also offers significant speed-ups for standard two-layer QM/MM calculations: thousands of MM charges are replaced by a small set of virtual surface charges (with little overhead since no SCRF procedure is required in this case) whose electrostatic interactions with the inner region are easily computed (with overall savings typically by a factor of 2).
The accuracy of the SMBP has been evaluated by comparing the results from QM/MM/SMBP calculations to those from standard QM/MM calculations for three diverse test systems: Glycine in water turned out to be problematic for the SMBP. Due to the high flexibility of the polar solvent many close-lying minima with different hydrogen bond patterns and different relative energies exist, and as a consequence, geometry optimizations by QM/MM and QM/MM/SMBP normally follow a different course and yield different local minima (unless starting from a given QM/MM minimum which is retained by QM/MM/SMBP). The individual reaction and activation energies for proton transfer in solvated glycine thus differ appreciably between QM/MM and QM/MM/SMBP, while the mean values for a small sample of five configurations are much closer to each other (within 1 kcal/mol). The two enzymatic test systems are more rigid. They are treated by the SMBP with impressive accuracy. Geometry optimizations by QM/MM and QM/MM/SMBP normally follow the same course and lead to essentially identical structures, and relative energies differ on average by less than 1 kcal/mol. The magnitude of these deviations is comparable to the spread of results that naturally occurs for different initial configurations. Finally, in the case of PHBH, the combined use of the SMBP and GSBP leads to free energy profiles and barriers that are essentially the same as those from full QM/MM calculations. The results presented in this chapter show that these boundary potentials enable an efficient and accurate description of enzymes at the QM/MM level.

Chapter 5

Application of the SMBP to Study Long Range Electrostatic and Bulk Solvent Effects

5.1 Motivation

One of the motivations for the development of boundary potentials for QM/MM calculations is to capture the influence of long range electrostatics on enzymatic reactions. In this context, one may differentiate between the contributions from bulk solvent and from the electrostatic potential of the outer macromolecule region (EPOM). By construction, the SMBP and the GSBP facilitate an independent study of both effects by appropriate choice of the dielectric constant of the PDC. Hence, in this chapter, both boundary potentials are applied to investigate and quantify the significance of these two contributions in QM/MM calculations. Two enzymatic reactions with different characteristics are studied: The intramolecular Claisen rearrangement in chorismate mutase (CM) is associated with little charge transfer. The hydroxylation reaction in PHBH, by contrast, corresponds to a formal "OH^+" transfer and thus involves significant charge transfer. Studying these two different systems will provide insight into the importance of bulk solvent and the EPOM. This will be valuable when applying SMBP and GSBP to other enzymatic systems or when developing new approaches to model long range electrostatics.
Prior to studying both contributions, the accuracy of SMBP and GSBP for these two systems is validated, and optimal values for their inherent parameters for typical enzymatic systems are determined. Although values for these parameters were proposed in the preceding two chapters, the values were

either optimized for a small model system (GSBP, chapter 3) or for accuracy tests (SMBP, chapter 4) but not for efficient applications. In this chapter, both boundary potentials are validated for two typical enzymatic systems as commonly studied with QM/MM methods. The protocol for determining optimal parameters and the resulting values are expected to be transferable to other enzymatic systems.

5.2 Chorismate Mutase: Validation and Parameter Determination

CM catalyzes the intramolecular Claisen rearrangement from chorismate to prephenate (figure 5.1). A recent review provides a summary of theoretical work on the reaction mechanism and the origin of catalysis in CM.[135] The Claisen rearrangement is a pericyclic reaction without significant charge transfer, and one would thus expect only minor effects of the EPOM and bulk solvent on this enzymatic reaction.

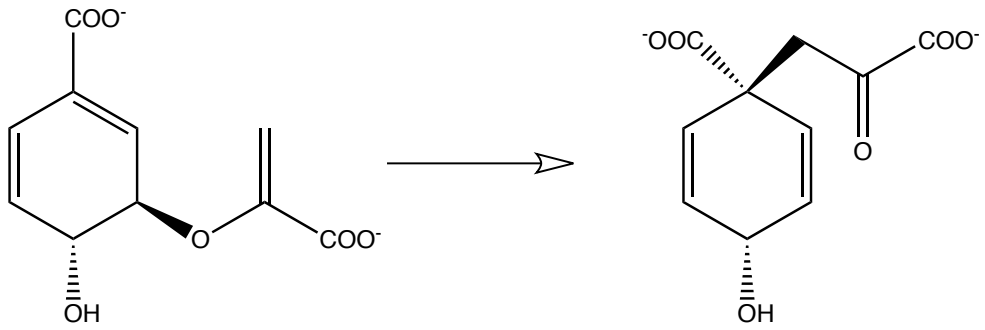


Figure 5.1: Intramolecular Claisen rearrangement catalyzed by chorismate mutase.

The setup for CM was based on a system that has been used in previous work.[152] Initial coordinates were taken from the crystallographic structure of *Bacillus subtilis* CM (PDB code 2CHT) with a bound transition state analog (TSA). Only the first of four trimers in the asymmetric unit was retained and the TSA between chains A and B was transformed into a chorismate molecule; the other TSAs were removed. The system was solvated with a 30 Å water sphere and then subjected to a 200 ps QM/MM MD simulation using self consistent charge-density functional tight binding (SCC-DFTB)[57] as QM method. The setup for this study started from the resulting structure

corresponding to snapshot 1 in the previous study.[152] To generate ten initial configurations for the QM/MM calculations that are independent of the previous study, snapshots were taken every 10 ps after 100 ps of extra equilibration. The QM region consisted only of the chorismate molecule which was modeled by the semiempirical AM1 Hamiltonian.[96] All protein atoms were assigned to the MM region. For each configuration, the inner region was centered on the initial position of the C_1 atom (following IUPAC nomenclature) with an extended dielectric cavity radius of 21.0 Å. All atoms within 19.0 Å of the center were assigned to the inner region and modeled explicitly. Due to the inaccuracies of the SMBP and the GSBP at the boundary,[114,153] the inner region was further subdivided into an active inner region and a frozen inner region. All atoms within 17.0 Å of the center belong to the active region so that it is surrounded by an "insulation" region of 2 Å. A spherical restraint with a radius of 17.0 Å and a force constant of 50 kcal/(mol $\cdot$ Å^2) was applied to the oxygen atoms of all active water molecules in MD simulations with the QM/MM/GSBP method. The reaction is described by means of a reaction coordinate (RC) defined as the difference of the lengths of the breaking C-O and the forming C-C bonds. Potential energy profiles of the reaction are computed by constraining the RC to values from -2.5 to 2.5 Å in steps of 0.1 Å.

The electrostatic potentials that constitute the SMBP (see equations 4.5 and 4.6) are obtained as grid-based solutions of the PB equation. As in previous applications, a focusing approach is applied to allow usage of fine grids for the inner region.[95] The PB equation is first solved for a coarse outer grid that covers the full system and then for a fine inner grid that focuses on the inner region. The boundary values of the inner grid are set by interpolation from the outer grid. Therefore, the mesh sizes of the grids are the main parameters that determine the accuracy as well as the efficiency of the SMBP. They have to be chosen carefully. Hence, one of the objectives of this study is the development of a transferable protocol to estimate adequate mesh sizes based only on fast single-point energy and gradient calculations.

The accuracy of the SMBP is evaluated by direct comparison to standard QM/MM results with both results obtained in vacuo. Table 5.2 provides the mean absolute (MAD) and maximum absolute deviations (MAX) of the gradient components for mesh size combinations of 0.15, 0.25, 0.4, 0.6, and 0.8 Å for the inner grid and 0.8, 1.25, 1.5, 2.0, and 2.5 Å for the outer grid. Here, all atoms within 18 Å of the center were considered to account for fluctuations of the active atoms that may occur later in the simulations. MAD and MAX results show no significant dependence on the mesh size of the outer grid but a strong dependence on the mesh size of the inner grid. For all mesh size combinations, the MAD values are between 10^{-5} and $5 \cdot 10^{-5}$

a.u. The default convergence criterion for QM/MM geometry optimizations is a maximum absolute gradient component of less than $4.5 \cdot 10^{-4}$ a.u.[121] In this perspective, the MAD results can be deemed accurate. However, they do not allow determination of the optimal mesh size. The MAX deviations are more helpful for this task. For inner mesh sizes of ≤ 0.25 Å, MAX deviations are below the convergence criterion so that this mesh size excels as a safe choice at acceptable computational costs. Since neither the accuracy nor the computational demands depend strongly on the mesh size of the outer grid, a relatively fine outer grid with a mesh size of 1.25 Å was selected in combination with an inner grid spacing of 0.25 Å for all calculations (unless noted otherwise).

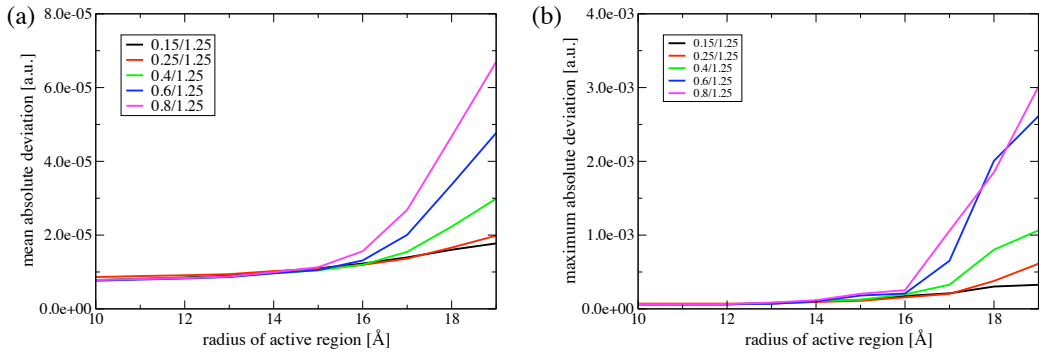


Figure 5.2: Mean absolute deviations (a) and maximum absolute deviations (b) of the electrostatic forces at all atoms inside the active region relative to the exact QM/MM values for the chorismate mutase (CM) test system. Results are shown for different mesh sizes of the inner grid and are plotted as a function of the radius of the active region. An outer grid size of 1.25 Å is used. All calculations were performed on configuration 1 of the CM system. The radius of the inner region was 19 Å (see text).

Detailed examination shows that the accuracy of the SMBP strongly depends on the radial position of the atoms. This point is illustrated in figure 5.2 which shows the MAD and MAX values for different inner grid mesh sizes as a function of the radial position. In the center of the inner region, the electrostatic potential varies only slowly and is described accurately by all tested mesh sizes. At the boundary, however, the electrostatic potential becomes more complex and the deviation increases significantly for mesh sizes > 0.4 Å. These results are very similar to those of the glycine/water test system in subsection 4.4.1.

Fine inner mesh sizes appear to be necessary to model the electrostatic potential at the boundary. Since the chemical process occurs in the center of the inner region, however, it is possible that moderate deviations at the boundary are tolerable. To test this hypothesis, reaction energies and activation energies were computed for grid size combinations ranging from 0.25/1.25 Å to 1.8/3.5 Å. The results are documented in table 5.1 and show that in comparison to full Coulombic electrostatics, the MAD and MAX deviations of the potential energy differences are below 0.3 and 0.8 kcal/mol, respectively, if the inner mesh size is ≤ 0.8 Å. Such an inner grid mesh size corresponds to a MAX deviation of the gradient components in the range of $2 \cdot 10^{-3}$ a.u. (see table 5.2). These findings lead to the conclusion that the SMBP will provide results of high accuracy if the mesh size is chosen such that the MAX values do not exceed $4.5 \cdot 10^{-4}$ a.u. anywhere in the inner region. The SMBP will still give results of adequate accuracy with coarser mesh sizes if the MAX values do not exceed $4.5 \cdot 10^{-4}$ a.u. for atoms more than 3 Å away from the boundary and if the MAX values do not exceed $2 \cdot 10^{-3}$ a.u. for atoms more than 1 Å away from the boundary.

Table 5.1: Reaction energies (ΔE) and activation energies ($\Delta E^{\ddagger}$) in chorismate mutase [kcal/mol] computed with the SMBP for different mesh size combinations.

	mesh size combination [Å]															
	0.25/1.25		0.4/1.25		0.6/1.25		0.8/1.25		0.8/2.5		1.2/2.5		1.5/2.5		1.8/3.5	
configuration	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$
1	-19.3	33.1	-19.2	33.1	-19.2	33.1	-19.2	33.2	-19.3	33.1	-19.3	33.1	-19.2	33.2	-19.5	32.8
2	-17.2	34.8	-17.1	34.8	-17.2	34.8	-17.2	34.8	-17.2	34.8	-17.2	34.8	-17.2	34.8	-17.2	34.8
3	-17.1	40.4	-17.1	40.4	-17.0	40.5	-17.1	40.4	-17.1	40.4	-17.4	40.3	-17.0	40.4	-17.1	40.5
4	-17.7	38.3	-17.6	38.3	-17.7	38.3	-17.6	38.4	-17.6	38.4	-19.4	38.3	-19.9	36.0	-18.3	37.7
5	-21.8	32.5	-21.8	32.4	-22.5	31.7	-21.8	32.4	-21.9	32.4	-21.8	32.5	-21.7	32.5	-21.9	32.4
6	-17.8	36.3	-17.8	36.4	-17.9	36.5	-17.8	36.5	-17.8	36.5	-17.8	36.5	-18.1	36.4	-17.9	36.5
7	-19.5	35.4	-19.8	35.1	-19.7	35.2	-19.7	35.1	-19.6	35.1	-19.7	35.1	-19.7	35.1	-19.7	35.1
8	-19.3	35.7	-19.2	35.7	-19.1	35.8	-19.2	35.7	-19.2	35.7	-19.2	35.7	-19.2	35.7	-19.3	35.7
9	-16.1	34.5	-16.9	34.5	-16.9	34.4	-16.9	34.5	-16.9	34.5	-16.9	34.4	-16.9	34.5	-16.8	34.5
10	-20.7	30.6	-20.6	30.7	-20.6	30.7	-20.9	30.6	-20.8	30.7	-21.6	29.7	-20.7	30.7	-20.5	30.8
mean value	-18.6	35.2	-18.7	35.1	-18.8	35.1	-18.7	35.1	-18.7	35.1	-19.0	35.0	-19.0	34.9	-18.8	35.1
std. dev. of data[a]	1.8	2.8	1.7	2.8	1.8	2.9	1.7	2.9	1.7	2.8	1.7	3.0	1.7	2.6	1.6	2.8
std. dev. of mean[b]	0.6	0.9	0.5	0.9	0.6	0.9	0.5	0.9	0.5	0.9	0.5	0.9	0.5	0.8	0.5	0.9
MAD[c]	0.2	0.1	0.3	0.1	0.3	0.2	0.3	0.2	0.3	0.1	0.5	0.2	0.5	0.3	0.2	0.2

[a] standard deviation of individual energy values

[b] standard deviation of the mean value (68% confidence limit)

[c] mean absolute deviation relative to full Coulombic electrostatics

Table 5.2: Mean absolute (MAD) and maximum absolute (MAX) deviations [10^{-4} a.u.] of the electrostatic forces computed with the SMBP for the chorismate mutase system and averaged over 10 configurations (relative to QM/MM results with full Coulombic electrostatics). Different mesh size combinations were used.

	inner grid size [Å]				
outer grid size [Å]	0.15	0.25	0.40	0.60	0.80
MAD - atoms within 18 Å					
0.80	0.11	0.12	0.19	0.31	0.45
1.25	0.16	0.16	0.23	0.34	0.48
1.50	0.17	0.17	0.23	0.35	0.48
2.00	0.15	0.16	0.22	0.34	0.47
2.50	0.18	0.18	0.24	0.36	0.49
MAX - atoms within 18 Å					
0.80	2.19	3.67	8.46	16.29	20.99
1.25	3.15	4.01	8.58	16.40	21.12
1.50	3.37	4.04	8.59	16.44	21.07
2.00	2.94	4.01	8.50	16.37	21.09
2.50	3.26	4.22	8.58	16.37	21.17

In the GSBP, another important parameter has to be determined: the size of the basis set which models the inner region charge distribution. The chosen orthonormal basis functions are based on spherical harmonics so that the size of the basis set is determined by the order of the highest multipole moment that is included. Table 5.3 gives the MAD and MAX deviations for basis sets of increasing size determined by maximum multipole moments from order $L = 1$ to $L = 19$. Moreover, it provides the fraction of inaccurate gradient components whose deviation is larger than the standard convergence criterion (see above). SMBP values serve as reference, since the SMBP represents the basis set limit of the GSBP. In previous applications of the GSBP, multipole moments up to order $L = 19$ were usually included. The MAX results confirm this choice and show that with $L = 19$ MAX deviations on the order of $4.5 \cdot 10^{-4}$ a.u. are observed which is sufficiently accurate. The fractions of inaccurate gradient components indicate that the reaction field potential converges in this system at $L = 17$. The residual error can be attributed to technical differences of SMBP and GSBP. It seems interesting to check whether looser criteria can also yield results of adequate accuracy. Table 5.4 shows that the basis set dependent error converges for $L = 10$ if only those

atoms are considered that are at least 3 Å away from the boundary. The impression that $L = 10$ provides sufficient accuracy is further supported by the fact that the MAX deviation for all atoms at least 1 Å away from the boundary is $1.3 \cdot 10^{-3}$ a.u. (table 5.3), and therefore, below the criterion of $2 \cdot 10^{-3}$ a.u. that was used to determine the mesh sizes. Hence, $L = 19$ emerges as the safe and $L = 10$ as the efficient choice. In section 5.3, both values are tested, and it will be shown that they describe bulk solvent effects in CM with similar accuracy.

Table 5.3: Mean absolute (MAD) and maximum absolute (MAX) deviations [10^{-4} a.u.] and the fraction of inaccurate gradient components (x_{grad}) [%] of the electrostatic forces of the chorismate mutase system computed with the GSBP with different basis set sizes (relative to SMBP results, see text). All values are averaged over 10 configurations and computed with a mesh size combination of 0.25/1.25 Å.

L[a]	MAD	MAX	x_{grad} [b]
1	4.00	69.77	26.47
2	3.19	60.99	21.03
3	2.39	50.28	15.60
4	1.74	39.47	11.05
5	1.22	30.00	6.76
6	0.95	27.21	4.67
7	0.76	22.59	3.16
8	0.61	18.56	1.97
9	0.50	14.48	1.13
10	0.42	13.31	0.70
11	0.36	10.97	0.40
12	0.32	8.75	0.25
13	0.29	7.34	0.12
14	0.27	5.95	0.07
15	0.25	5.26	0.05
16	0.24	5.03	0.04
17	0.23	4.72	0.03
18	0.22	4.55	0.03
19	0.22	4.53	0.03

[a] highest order multipole moment

[b] fraction of gradient components with a deviation $> 4.5 \cdot 10^{-4}$a.u. [%]

Table 5.4: Mean absolute (MAD) and maximum absolute (MAX) deviations [10^{-4} a.u.] and the fraction of inaccurate gradient components (x_{grad}) [%] of the electrostatic forces of the chorismate mutase test system computed with the GSBP with different basis set sizes (relative to the SMBP results, see text). Only atoms that are at least 3 Å away from the boundary are considered. All values are averaged over 10 configurations with a mesh size of 0.25/1.25 Å.

L[a]	MAD	MAX	x_{grad} [b]
1	2.93	49.96	19.87
2	2.16	40.17	13.83
3	1.47	31.16	8.42
4	0.99	22.00	4.41
5	0.64	14.91	1.68
6	0.49	11.44	0.83
7	0.38	8.13	0.39
8	0.32	6.82	0.12
9	0.27	5.47	0.06
10	0.25	4.73	0.04
11	0.23	4.60	0.04
12	0.22	4.47	0.04
13	0.22	4.42	0.04
14	0.21	4.42	0.04
15	0.21	4.42	0.04
16	0.21	4.41	0.04
17	0.21	4.41	0.04
18	0.20	4.41	0.04
19	0.20	4.43	0.04

[a] highest order multipole moment

[b] fraction of gradient components with a deviation $> 4.5 \cdot 10^{-4}$a.u. [%]

5.3 Chorismate Mutase: Results

Now the SMBP is applied to study the effects of the EPOM and bulk solvent on the enzymatic reaction in CM. Table 5.5 gives the reaction energies and activation energies that were computed with different methods to describe these effects. In each case, the energy differences were averaged over ten configurations. The standard deviation of any given mean value defines a confidence interval corresponding to a confidence limit of 68 %. Statistically significant are differences between mean values that are larger than the confidence interval of the reference value.
Standard QM/MM calculations with full Coulombic electrostatics (Coulomb in table 5.5) give reaction and activation energies of -18.8 and 35.1 kcal/mol, respectively. In the standard QM/MM approach using electronic embedding, the EPOM is computed without approximation but the effect of bulk solvent is neglected. Compared to the experimental value of 12.7 kcal/mol,[154] the computed value for the activation energy is strongly overestimated because of the use of the AM1 Hamiltonian in the QM region.[135] Despite this shortcoming, the AM1 Hamiltonian is expected to be adequate for studying the relative effects caused by long range electrostatics, which should be captured at this level in a realistic manner. For two of the ten configurations, the conclusions drawn from AM1 results were confirmed by more accurate DFT calculations using the B3LYP functional[142–144] in combination with a 6-31G* basis (see below).

Table 5.5: Reaction energies (ΔE) and activation energies ($\Delta E^{\ddagger}$) of the Claisen rearrangement in chorismate mutase [kcal/mol] computed with different treatments of long range electrostatics.

	Coulomb		SMBP(vac)[a]		NOR[b]		SMBP(solv)[c]		SMBP(solv,app)[d]	
configuration	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$
1	-20.0	33.0	-19.3	33.1	-19.9	33.4	-20.1	32.6	-19.1	32.8
2	-17.3	34.8	-17.2	34.8	-11.5	41.6	-18.3	33.1	-17.2	34.3
3	-17.5	40.1	-17.1	40.4	-18.9	37.2	-16.8	39.5	-16.5	40.2
4	-17.7	38.2	-17.7	38.3	-14.9	36.5	-17.3	37.5	-17.3	37.8
5	-21.9	32.4	-21.8	32.5	-21.8	33.1	-21.9	32.4	-21.9	32.7
6	-18.0	36.2	-17.8	36.3	-17.1	36.6	-17.7	35.4	-17.7	35.4
7	-19.6	35.3	-19.5	35.4	-19.3	35.3	-19.9	34.7	-19.6	35.1
8	-19.3	35.8	-19.3	35.7	-19.2	37.2	-19.4	34.9	-19.0	36.0
9	-16.3	34.4	-16.1	34.5	-17.4	34.7	-17.9	33.5	-17.2	34.0
10	-20.7	30.6	-20.7	30.6	-20.6	31.3	-21.0	30.4	-20.0	31.5
mean value	-18.8	35.1	-18.6	35.2	-18.1	35.7	-19.0	34.4	-18.5	35.0
standard deviation of data[e]	1.8	2.8	1.8	2.8	3.0	2.9	1.7	2.6	1.7	2.6
standard deviation of mean[f]	0.6	0.9	0.6	0.9	0.9	0.9	0.5	0.8	0.5	0.8
MAD[g]	-	-	0.2	0.1	1.3	1.5	-	-	0.5	0.6

[a] $\epsilon = 1$

[b] neglect of outer region

[c] $\epsilon = 80$, [d] $\epsilon = 80, \phi_{rf}^{QM} = 0$

[e] standard deviation of individual energy values

[f] standard deviation of the mean value (68% confidence limit)

[g] mean absolute deviation relative to full Coulombic electrostatics. For SMBP(solv,app), SMBP(solv) values are used as reference.

Applying the SMBP under vacuum conditions (SMBP(vac)), the effect of the EPOM is approximated and bulk solvent effects are neglected. The SMBP describes the EPOM accurately and reproduces the standard QM/MM results with very small deviations. The mean values for the reaction and activation energy differ by only 0.2 and 0.1 kcal/mol, respectively. Also the results for the individual configurations are very similar as indicated by MAD values of similar size.
Going one step further, one could completely neglect the electrostatic influence from the outer solvent and macromolecule region (neglect of outer region, NOR). This simplest approximation should lead to significant deviations from the standard QM/MM results, if the energetics are sensitive to long range electrostatics. With the NOR approximation, the mean reaction and activation energies increase by 0.7 and 0.6 kcal/mol, respectively (table 5.5). These differences are at the boundary of the confidence interval of the standard QM/MM results and thus cannot be considered significant. However, the MAD values of 1.3 and 1.5 kcal/mol for reaction and activation energies show that the results deviate significantly for the individual configurations. With less fortuitous error cancellation, significant discrepancies of more than 1 kcal/mol are therefore possible when applying the NOR approximation.
The effect of bulk solvent on this reaction was studied using the SMBP with a dielectric constant of 80 for the outer solvent region (SMBP(solv)). The inclusion of bulk solvent lowers the reaction energy only marginally by 0.2 kcal/mol. A stronger change is observed for the activation energy which is reduced by 0.7 kcal/mol to a value of 34.4 kcal/mol. However, both values lie still within the statistical error bars of the standard QM/MM results so that the effect of bulk solvent has to be deemed insignificant for this reaction.
For configurations 3 and 8, the effects of the EPOM and bulk solvent were checked by DFT/MM calculations using B3LYP/6-31G* for the QM region. The results are summarized in table 5.6. Focusing on the computed barriers, one first notes that the standard DFT/MM treatment with Coulombic electrostatics yields values (13.0 - 15.9 kcal/mol) close to experiment (see above). These barriers are lowered slightly when applying the SMBP under vacuum conditions (by 0.0 - 0.4 kcal/mol) and more so when using the NOR approximation (by 0.2 - 1.1 kcal/mol). The barrier lowering due to bulk solvent is again small (0.2 - 0.4 kcal/mol). These DFT results are fully consistent with the AM1/MM results (table 5.5).

Since the QM region is usually located far away from the dielectric boundary, one can assume that the QM contribution to the reaction field potential is small and can be neglected. This is the SMBP(solv,app) approximation

Table 5.6: Reaction energies (ΔE) and activation energies ($\Delta E^{\ddagger}$) in chorismate mutase [kcal/mol] computed at the QM(B3LYP/6-31G*)/MM level of theory with different treatments of long range electrostatics (see table 5.5 for notation).

	Coulomb		SMBP(vac)		NOR[a]		SMBP(solv)	
configuration	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$
3	-14.0	15.9	-14.8	15.5	-14.3	14.8	-13.5	15.7
8	-16.2	13.0	-16.2	13.0	-17.1	12.8	-16.3	12.6

[a] neglect of outer region

which neglects the ϕ_{rf}^{QM} term in equation 4.5 and thus avoids the SCRF procedure, thereby accelerating the SMBP calculations. This approach yields reaction and activation energies of -18.5 and 35.0 kcal/mol, respectively. These values are within the statistical error bars of the standard QM/MM and the SMBP(solv) results. Hence, this approximation is valid for this reaction.
Finally, it was checked whether the effect of the EPOM and bulk solvent is more pronounced for the dynamical behavior of this system, *i.e.*, when computing free energy differences. The QM/MM-FEP method was applied to compute the free energies of reaction and activation at T = 300 K (table 5.7). Along the RC, the reaction was split up into discrete windows. For each of these, ESP charges for the QM atoms were derived by fitting to the electrostatic potential at the 200 MM atoms closest to the QM region. For each window, the active region was equilibrated for 10 ps with the QM region held fixed and the QM-MM electrostatic interactions modeled classically using the ESP charges. Subsequently, energy differences were sampled for 10 ps and the data were subjected to statistical tests for lack of correlation and trend.[22] If necessary, data were discarded to obtain a series of values without trend. Although the effective sampling length was determined by the statistical tests, it was ensured that at least 5 ps of data were retained.

In the QM/MM-FEP calculations, the GSBP was applied to sample the MM phase space efficiently.[153] Again, the accuracy of the QM/MM/GSBP-FEP ansatz was validated against standard QM/MM-FEP results using vacuum conditions. However, since QM/MM-FEP calculations are computationally intense, free energy profiles without GSBP were obtained only for two configurations. The results are documented in table 5.8 and show that QM/MM/GSBP reproduces QM/MM results in FEP calculations well, with

Table 5.7: Reaction free energies (ΔA) and activation free energies ($\Delta A^{\ddagger}$) in chorismate mutase [kcal/mol] at T = 300 K computed with different treatments of long range electrostatics.

	GSBP(vac)		GSBP(solv)[a]		GSBP(solv,fast)[b]		NOR[c]	
configuration	ΔA	$\Delta A^{\ddagger}$	ΔA	$\Delta A^{\ddagger}$	ΔA	$\Delta A^{\ddagger}$	ΔA	$\Delta A^{\ddagger}$
1	-19.0	31.1	-17.6	31.1	-17.0	31.6	-16.9	31.7
2	-17.1	32.6	-16.2	33.1	-15.7	33.1	-14.5	36.8
3	-16.0	37.8	-14.6	36.2	-15.6	36.6	-16.3	33.9
4	-18.0	35.1	-15.2	35.7	-15.9	34.6	-15.8	33.7
5	-18.3	33.5	-19.2	32.2	-18.3	32.7	-18.9	32.2
6	-17.0	34.8	-17.0	32.4	-16.6	33.4	-15.8	34.6
7	-18.6	33.2	-18.1	32.5	-18.5	32.1	-19.1	31.7
8	-18.1	33.2	-16.6	33.4	-17.3	32.9	-16.2	33.1
9	-17.5	32.2	-16.6	32.3	-17.3	31.7	-16.2	31.9
10	-19.7	30.4	-18.4	30.2	-18.0	31.1	-17.8	31.9
mean value	-17.9	33.4	-16.9	32.9	-17.0	33.0	-16.8	33.1
std. dev. of data[d]	1.1	2.1	1.4	1.8	1.1	1.6	1.4	1.6
std. dev. of mean[e]	0.3	0.7	0.5	0.6	0.3	0.5	0.5	0.5

[a] grid size: 0.25/1.25 Å, 400 basis functions (L=19)

[b] grid size: 0.6/2.0 Å, 121 basis functions (L=10)

[c] neglect of outer region

[d] standard deviation of individual energy values

[e] standard deviation of the mean value (68% confidence limit)

deviations of 0.0 - 0.2 kcal/mol in the reaction free energies (ΔA) and 0.3 - 0.5 kcal/mol in the activation free energies ($\Delta A^{\ddagger}$). Now the changes in the calculated mean values when going from potential energies (table 5.5) to free energies (table 5.7) are considered. Using the GSBP in vacuum, the reaction energy increases by 0.7 kcal/mol to -17.9 kcal/mol while the activation energy decreases by 1.8 kcal/mol to 33.4 kcal/mol. Four terms contribute to the difference between free and potential energy: the zero point vibrational energy (ZPE, ΔE_{ZPE}), the thermal contribution to the internal energy (ΔU_{th}), the QM entropy contribution ($-T\Delta S_{QM}$), and the entropy of the QM-MM interactions ($-T\Delta S_{QM-MM}$) (see subsection 2.4). In the QM/MM-FEP approach, $-T\Delta S_{QM-MM}$ is assumed to be the difference of free and potential QM-MM interaction energies and neglect other contributions. Table 5.9

shows that the total effect on the barrier is dominated by the ZPE. The entropic QM contribution from the harmonic approximation (0.8 kcal/mol) cancels the entropic QM-MM contribution from the sampling (-0.8 kcal/mol) so that entropy does not contribute significantly to finite-temperature effects on the activation energy. This result contradicts the experimental observation of an entropic contribution of $(-T\Delta S)_{exp} = 2.7$ kcal/mol.[154] The negative change of entropy in the transition state has been ascribed to the loss of conformational flexibility due to the partial formation of two covalent bonds.[152] Since the degrees of freedom of the QM region are held fixed during QM/MM-FEP sampling, this phenomenon cannot be captured in the free energy difference of QM-MM interactions. The entropic QM contribution based on the harmonic approximation shows the correct trend but underestimates the magnitude. Therefore, one may tentatively assume that a significant contribution to the negative change in entropy in the activation energy of CM comes from the degrees of freedom that involve coupled motions of QM and MM atoms which are not sampled in the QM/MM-FEP ansatz. The deviation of QM/MM/GSBP-FEP results from experiment would then arise from the QM/MM-FEP ansatz itself, not from the approximations in the boundary potentials. This view is supported by the results of previous semiempirical QM/MM umbrella sampling simulations (QM = SCC-DFTB) that do not restrict the flexibility of the QM region and reproduce the entropic contribution to the barrier with good accuracy.[152]

Table 5.8: Reaction free energies (ΔA) and activation free energies ($\Delta A^{\ddagger}$) in chorismate mutase [kcal/mol] computed with standard QM/MM and with a combination of SMBP and GSBP to mimic the outer macromolecule region in vacuum (QM/MM/GSBP(vac)).

	QM/MM		QM/MM/GSBP(vac)	
configuration	ΔA	$\Delta A^{\ddagger}$	ΔA	$\Delta A^{\ddagger}$
3	-16.05	37.36	-16.03	37.85
8	-17.90	33.44	-18.08	33.18

The effects of bulk solvent on the free energy differences were studied using a dielectric constant of 80 for the outer solvent region (GSBP(solv)). The activation free energy decreases upon such inclusion of bulk solvent by 0.5 kcal/mol and is thus still within the confidence interval of the vacuum result. The reaction energy, however, increases significantly by 1.0 kcal/mol to -16.9

Table 5.9: Contributions [kcal/mol] to the differences between free and potential energies in chorismate mutase using the QM/MM/GSBP-FEP method (see text). All values are averaged over ten configurations.

	reaction	activation
ΔE_{ZPE}	0.3	-1.5
ΔU_{th}	-0.1	-0.3
$-T\Delta S_{QM}$	0.3	0.8
$-T\Delta S_{QM-MM}$	0.2	-0.8
Total	0.7	-1.8

kcal/mol. These results are reproduced quite accurately if coarser mesh sizes of 0.6/2.0 Å and a smaller basis set ($L = 10$) are used (GSBP(solv,fast)). However, simply neglecting the electrostatic effect of the outer region (NOR) yields results of very similar accuracy. This can be explained by the shielding effect of the outer solvent which is mimicked most simply by neglecting the outer region charges. The quantitative accuracy is probably fortuitous given that the results for the potential energy differences deviate significantly from the SMBP(solv) results (table 5.5).
Based on these results, one may draw the following conclusions for CM: The SMBP and the GSBP reproduce long range electrostatic interactions with high accuracy when using fine grids with mesh sizes of 0.25/1.25 Å as well as coarser grids with mesh sizes of 0.6/2.0 Å. It should be emphasized, however, that the influence of the EPOM on this reaction is limited. This can be attributed to the "neutral" character of the intramolecular Claisen rearrangement in CM. Neglecting the electrostatic influence of the outer region does not have much effect on the mean potential energy differences although the results for the individual configurations differ significantly. Similarly, bulk solvent effects do not affect the potential energy differences. A statistically significant influence of bulk solvent on the order of 1 kcal/mol was only observed for the reaction free energy.

5.4 p-Hydroxybenzoate Hydroxylase

The setup of the PHBH test system was very similar to the one used previously in section 4.4.2 and differed only in two aspects. First, in this study five configurations were selected from the 500 ps MD run after 420, 440, 460, 480, and 500 ps to reduce the dependency of the results from the initial

geometries. While two configurations were enough to prove that the SMBP works in principle (section 4.4.2), a larger number of configurations is necessary to study the effect of the EPOM and bulk solvent. Second, a higher level of QM theory was applied to show that the SMBP and the protocol for parameter determination work also for ab initio and DFT QM methods. Although physical reasoning and previous results (section 5.3) suggest that the performance of the SMBP is independent of the QM method, a test is preferable. Hence, the B3LYP functional[142–144] in combination with a 6-31G* basis was used to model the QM atoms.
The same RC was employed as in section 4.4.2 and the potential energy profile was scanned for RC values between -1.7 Å and 1.7 Å in steps of 0.1 Å. The same protocol as in the CM system was applied to validate the boundary potentials for PHBH with a DFT QM method, and the optimal mesh and basis set sizes were determined. The outcome of these statistical tests is analyzed only briefly since the PHBH results with a DFT QM method are analogous to the CM results with a semiempirical QM method.

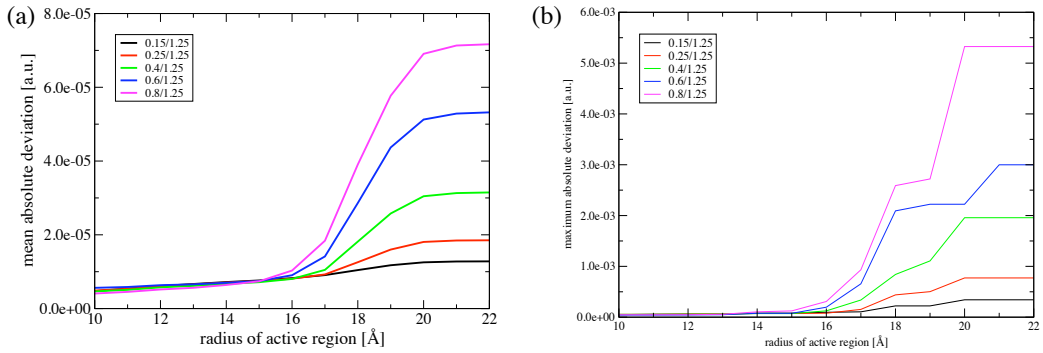


Figure 5.3: Mean absolute deviations (a) and maximum absolute deviations (b) of the electrostatic forces at all atoms inside the active region relative to the exact QM/MM values for the *p*-hydroxybenzoate hydroxylase (PHBH) test system. Results are shown for different mesh sizes of the inner grid and plotted as a function of the radius of the active region. An outer grid size of 1.25 Å is used. All calculations were performed on configuration 1 of the PHBH system. The radius of the inner region was 18.5 Å (see text).

This holds for the deviations of the gradient components (table 5.10) as well as the relationship of accuracy and radial position: deviations are very small in the center of the inner region and increase significantly only for coarse mesh sizes and only at the boundary separating inner and outer region (fig-

ure 5.3). Therefore, a mesh size combination of 0.25/1.25 Å excels again as a safe choice that reproduces full Coulombic electrostatics very accurately in all parts of the inner region. However, also coarser mesh sizes reproduce the electrostatic potential accurately everywhere except in close proximity to the boundary. As in CM, potential energy differences are less sensitive to the mesh sizes. Computation of potential energy differences with mesh size combinations ranging from 0.25/1.25 Å to 1.8/3.5 Å yield accurate results with MAX deviations that do not exceed 0.3 kcal/mol for all mesh sizes (table 5.12). In GSBP calculations, basis sets of similar size as in CM are adequate for PHBH (table 5.11).

Table 5.10: Mean absolute (MAD) and maximum absolute (MAX) deviations [10^{-4} a.u.] of the electrostatic forces of the *p*-hydroxybenzoate hydroxylase system computed with the SMBP and averaged over 5 configurations (relative to QM/MM results with full Coulombic electrostatics). Different mesh size combinations were used.

	inner grid size [Å]				
outer grid size [Å]	0.15	0.25	0.40	0.60	0.80
MAD - atoms within 18 Å					
0.80	0.09	0.11	0.17	0.27	0.39
1.25	0.11	0.13	0.19	0.29	0.40
1.50	0.14	0.15	0.20	0.29	0.41
2.00	0.14	0.15	0.20	0.29	0.41
2.50	0.15	0.16	0.22	0.30	0.42
MAX - atoms within 18 Å					
0.80	2.25	5.19	9.83	20.94	31.63
1.25	2.12	5.27	9.88	21.08	31.58
1.50	2.97	5.22	9.93	21.25	31.73
2.00	2.54	5.30	9.90	21.12	31.64
2.50	2.83	5.36	9.92	21.25	31.68

Table 5.11: Mean absolute (MAD) and maximum absolute (MAX) deviations [10^{-4} a.u.] and the fraction of inaccurate gradient components (x_{grad}) [%] of the electrostatic forces of the *p*-hydroxybenzoate hydroxylase system computed with the GSBP with different basis set sizes (relative to the SMBP results, see text). All values are averaged over 5 configurations with a mesh size of 0.25/1.25 Å.

L[a]	MAD	MAX	x_{grad} [b]
1	1.63	22.52	8.69
2	1.05	15.50	3.11
3	0.82	15.28	1.67
4	0.72	12.05	1.15
5	0.67	11.83	0.77
6	0.64	11.82	0.63
7	0.59	11.78	0.47
8	0.57	11.79	0.41
9	0.56	11.77	0.40
10	0.56	11.77	0.40
11	0.56	11.79	0.41
12	0.56	11.77	0.40
13	0.55	11.77	0.38
14	0.55	11.76	0.38
15	0.55	11.77	0.38
16	0.55	11.78	0.39
17	0.55	11.76	0.39
18	0.55	11.74	0.39
19	0.55	11.65	0.39

[a] highest order multipole moment

[b] fraction of gradient components with a deviation $> 4.5 \cdot 10^{-4}$a.u. [%]

Table 5.12: Reaction energies (ΔE) and activation energies ($\Delta E^{\ddagger}$) of the hydroxylation reaction catalyzed by *p*-hydroxybenzoate hydroxylase [kcal/mol] computed with the SMBP for different mesh size combinations.

	mesh size combination [Å]															
	0.25/1.25		0.4/1.25		0.6/1.25		0.8/1.25		0.8/2.5		1.2/2.5		1.5/2.5		1.8/3.5	
configuration[a]	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$
1	-55.0	30.5	-54.9	30.6	-54.8	30.6	-54.9	30.5	-54.9	30.6	-54.9	30.6	-54.9	30.7	-55.0	30.8
3	-52.3	33.1	-52.2	33.0	-52.2	33.4	-52.4	33.2	-52.6	32.8	-52.4	33.1	-52.4	33.0	-52.3	33.0
4	-51.1	29.4	-50.9	29.6	-51.0	29.4	-51.0	29.5	-51.1	29.4	-50.9	29.7	-51.1	29.5	-50.9	29.6
5	-46.8	33.1	-46.6	33.1	-46.6	33.0	-46.9	32.9	-46.8	32.9	-46.7	33.0	-46.7	32.9	-46.8	33.1
mean value	-51.3	31.5	-51.2	31.6	-51.1	31.6	-51.3	31.5	-51.3	31.4	-51.2	31.6	-51.3	31.5	-51.3	31.6
std. dev. of data[b]	3.4	1.9	3.4	1.7	3.4	1.9	3.3	1.8	3.4	1.7	3.4	1.7	3.4	1.7	3.4	1.7
std. dev. of mean[c]	1.7	0.9	1.7	0.9	1.7	0.9	1.7	0.9	1.7	0.8	1.7	0.9	1.7	0.9	1.7	0.9
MAD[d]	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.0	0.1	0.1
MAX[e]	0.1	0.1	0.2	0.1	0.2	0.4	0.1	0.2	0.3	0.2	0.1	0.2	0.1	0.1	0.1	0.2

[a] Configuration 2 was discarded since a continuous energy profile could not be obtained (see text).

[b] standard deviation of individual energy values

[c] standard deviation of the mean value (68% confidence limit)

[d] mean absolute deviation relative to full Coulombic electrostatics

[e] maximum absolute deviation relative to full Coulombic electrostatics

These results permit the conclusion that SMBP and GSBP behave very similarly with respect to their accuracy and its dependence on the parameters for these two enzymes despite the different nature of the two reactions and the different QM methods used (AM1 and B3LYP, respectively). Moreover, very similar results and trends were observed in the glycine/water test system (see subsection 4.4.1). This substantiates the expectation that not only the validation protocol but also the resulting parameters are transferable to other enzymatic systems. A mesh size combination of 0.25/1.25 Å with multipole moments up to order $L = 19$ is suggested as a safe choice for accurate calculations. As default options, a mesh size combination of 0.6/1.25 Å with multipole moments up to order $L = 10$ is recommended for efficient application of SMBP and GSBP. Since even coarser mesh sizes reproduce energy differences with only marginal deviations, these values will yield accurate results at reduced computational costs for general enzymatic systems. A decrease of the order of the highest multipole moment from $L = 19$ to $L = 10$ corresponds to a decrease of the number of basis functions by about 70 % from 400 to 121. The computation time for the reaction field matrix depends directly on the number of basis functions and thus decreases by about 70 % as well (see section 3.3). Moreover, the most time consuming part of the GSBP gradient is the computation of the derivatives of the basis functions with respect to the position of all atoms in the inner region (see equation 3.21). The computation time for this step also decreases significantly upon reduction of the basis set size.

A mesh size combination of 0.25/1.25 Å was applied to compute the potential energy differences of the hydroxylation step in PHBH with the SMBP in solution and in vacuum, with full Coulombic electrostatics, and with the simple NOR approach. It was necessary to discard configuration 2, since continuous energy profiles could not be obtained for this configuration despite many attempts. The results in table 5.13 show that the SMBP reproduces standard QM/MM results with high accuracy. The mean values of the reaction and activation energy deviate by only 0.2 and 0.0 kcal/mol. MAD values of similar magnitude show that the results for the individual configurations also agree well. In contrast to CM, the EPOM has a significant effect on the energetics of the hydroxylation reaction in PHBH. If the electrostatic influence of the outer region is simply neglected (NOR), the mean value of the reaction energy changes by more than 6 kcal/mol from -25.4 to -31.5 kcal/mol. The activation energy is reduced by more than 2 kcal/mol from 9.7 to 7.6 kcal/mol. MAD values of 6.1 and 3.0 kcal/mol for reaction and activation energies show that the deviations are even larger for the individual configurations.

Table 5.13: Reaction energies (ΔE) and activation energies ($\Delta E^{\ddagger}$) of the hydroxylation reaction in *p*-hydroxybenzoate hydroxylase [kcal/mol] computed with different treatments of long range electrostatics.

	Coulomb		SMBP(vac)[a]		NOR[b]		SMBP(solv)[c]		SMBP(solv,app)[d]	
configuration	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$	ΔE	$\Delta E^{\ddagger}$
1	-27.0	9.5	-26.9	9.4	-28.9	11.4	-28.5	8.0	-29.7	7.9
3	-21.2	11.4	-20.9	11.4	-35.5	4.6	-21.8	10.7	-30.6	7.9
4	-28.1	7.9	-27.9	7.9	-32.1	6.6	-29.8	5.9	-31.3	6.1
5	-25.3	10.1	-25.2	10.2	-29.5	7.9	-25.2	9.0	-28.9	8.1
mean value	-25.4	9.7	-25.2	9.7	-31.5	7.6	-26.3	8.4	-30.1	7.5
std. dev. of data[e]	3.0	1.5	3.1	1.5	3.0	2.9	3.6	2.0	1.0	1.0
std. dev. of mean[f]	1.5	0.7	1.5	0.7	1.5	1.4	1.8	1.0	0.5	0.5
MAD[g]	-	-	0.2	0.0	6.1	3.0	-	-	3.8	1.0

[a] $\epsilon = 1$

[b] neglect of outer region

[c] $\epsilon = 80$, [d] $\epsilon = 80, \phi_{rf}^{QM} = 0$

[e] standard deviation of individual energy values

[f] standard deviation of the mean value (68% confidence limit)

[g] mean absolute deviation relative to full Coulombic electrostatics. For SMBP(solv,app), SMBP(solv) values are used as reference.

A more detailed analysis reveals that two different effects are responsible for the observed differences in the reaction energies. In configurations 1 and 3, the hydrogen bonding networks connecting the QM and MM region are different after geometry optimization with full Coulombic electrostatics and with the NOR approximation. By contrast, in configurations 4 and 5 the hydrogen bonding networks between the QM and MM region are very similar for both electrostatic treatments. In these two configurations, the differences of the reaction energies are dominated by the differences of the electrostatic QM-MM interaction energies as shown in table 5.14. Here, the electrostatic QM-MM interaction energy includes direct QM-MM interactions as well as the polarization effect of the MM point charges. These results suggest that with the NOR approximation the MM atoms are more flexible to adapt to the electrostatic potential of the QM region since they do not feel the EPOM. This effect becomes more pronounced when the QM region becomes more polar. During the hydroxylation reaction in PHBH, the less polar peroxide group separates into more polar alcohol and alcoholate groups which form hydrogen bonds with the neighboring MM residues. Therefore, the difference of the QM-MM interaction energies of product and reactant is greater with the NOR approximation compared to full Coulombic electrostatics. The NOR approximation favors the more polar product state and thus shifts the reaction energies to more negative values. In conclusion, two consequences of neglecting the EPOM are possible: geometry optimizations may either lead to structures that already differ in the hydrogen bonding network, or changes in the QM-MM interaction energies may bias reaction energetics significantly even when the hydrogen bonding network is similar. These results underline that the EPOM can have a significant influence on potential energy differences. They cast doubt on the quantitative accuracy of QM/MM and pure QM studies that neglect the EPOM.

Table 5.14: Contributions to the QM/MM reaction energies [kcal/mol] of the hydroxylation reaction in *p*-hydroxybenzoate hydroxylase computed with full Coulombic electrostatics and with the NOR (neglect of outer region) approximation.

	configuration 1			configuration 3			configuration 4			configuration 5		
contribution	Coulomb	NOR	dev.[a]	Coulomb	NOR	dev.[a]	Coulomb	NOR	dev.[a]	Coulomb	NOR	dev.[a]
QM_{pol} [b]	-18.4	-23.8	-5.4	-6.2	-28.5	-22.3	-21.3	-25.1	-3.8	-20.4	-23.8	-3.4
QM_{iso} [c]	-49.0	-55.2	-6.2	-48.8	-43.9	4.9	-48.0	-49.3	-1.3	-48.8	-49.2	-0.4
MM	-8.7	-5.2	3.4	-15.0	-6.8	8.3	-6.8	-7.0	-0.2	-4.7	-5.7	-0.9
QM-MM[d]	30.7	31.5	0.8	42.6	15.4	-27.2	26.8	24.2	-2.6	28.5	25.4	-3.0
QM/MM[e]	-27.0	-29.0	-2.0	-21.2	-35.3	-14.0	-28.0	-32.1	-4.1	-25.1	-29.5	-4.4

[a] dev. = NOR - Coulomb

[b] polarized QM energy including electrostatic QM-MM interactions

[c] energy of isolated QM region with QM/MM geometry

[d] electrostatic QM-MM interaction energy: QM-MM = QM_{pol} - QM_{iso}

[e] full QM/MM energy: QM/MM = QM_{pol} + MM

In PHBH, one also observes a significant solvent effect on the reaction. Modeling bulk solvent with the SMBP, the activation energy decreases by more than 1 kcal/mol from 9.7 to 8.4 kcal/mol. This agrees qualitatively with chemical reasoning: the bulk solvent stabilizes the charged OH^+ species and lowers the energy of the transition state. Given the distinct charge transfer in this reaction, the effect of bulk solvent that is observed in PHBH should be in the upper range of what can be expected in enzymatic reactions. Even stronger effects seem possible if the inner region is chosen to be rather small and the charge transfer thus occurs closer to the dielectric boundary.
The significant solvent effect renders this reaction an interesting test for the SMBP(solv,app) method which neglects the QM contribution to the reaction field potential. A satisfying agreement with SMBP(solv) results can only be observed for configuration 1. For the other configurations, deviations reach up to 10 kcal/mol. The mean value for the reaction energy is -30.1 kcal/mol and therefore far outside the confidence interval of the SMBP(solv) mean value. In the case of PHBH, the QM contribution to the reaction field potential thus has a significant influence on the energetics of the hydroxylation reaction. Hence, the SMBP(solv,app) method is not a valid approximation in the PHBH system.

5.5 Conclusion

In this chapter, the electrostatic effect of the outer macromolecule region and of bulk solvent was evaluated for two enzymatic systems. The SMBP and GSBP were applied to distinguish the effects of bulk solvent and the EPOM by appropriate choice of the dielectric constant of the PDC. Both boundary potentials introduce approximations to describe electrostatic interactions with the outer macromolecule region more efficiently. Therefore, the accuracy of SMBP and GSBP was evaluated for both enzymatic test systems, and a protocol for validation and determination of adequate values for its inherent parameters was presented. This protocol was applied to generate a set of optimal parameters that is transferable to general enzymatic systems. SMBP and GSBP were found to describe the EPOM with high accuracy. Typically, deviations of mean values on the order of 0.1 to 0.2 kcal/mol are observed for both enzymes. Deviations for individual configurations may be slightly higher but rarely exceed 0.3 kcal/mol.
Two enzymatic reactions with rather different characteristics were used to study the effect of the EPOM and bulk solvent. The Claisen rearrangement in CM is a pericyclic reaction without much charge transfer. For this kind

of reaction, the electrostatic influence of the outer macromolecule region on the reaction energetics is not significant when considering the mean values of all 10 configurations. However, deviations on the order of 1.5 kcal/mol are observed for individual configurations so that neglect of long range electrostatics can be detrimental in the absence of adequate sampling. Bulk solvent effects on the reaction energetics in CM are found to be small.
The hydroxylation reaction in PHBH, in contrast, is associated with a stronger charge transfer since the reaction formally corresponds to an OH^+ transfer. In consequence, the EPOM has a strong influence on the reaction energetics, and its neglect causes errors of several kcal/mol due to a systematic overstabilization of the more polar product state (arising from the higher flexibility of the MM residues without the EPOM). Moreover, bulk solvent stabilizes the transition state and reduces the reaction barrier by about 1 kcal/mol in PHBH.
Depending on the charge transfer characteristics of the chemical process, the EPOM and bulk solvent can thus have a significant effect on the energetics of enzymatic reactions. Among these two contributions, the EPOM is clearly more important. SMBP and GSBP offer a convenient way to evaluate both contributions accurately and efficiently in QM/MM calculations.

Chapter 6

Conclusion

The main goal of this thesis was to merge the concept of boundary potentials from classical simulations with the hybrid quantum mechanical/molecular mechanical (QM/MM) ansatz to create a three-layer method. This new method treats long range electrostatic interactions accurately and efficiently, and reduces the computational costs compared to standard QM/MM approaches.

First, the generalized solvent boundary potential (GSBP) was adapted for semiempirical QM/MM methods and implemented into the modular QM/MM software ChemShell. Application of the GSBP is connected with a large overhead that is dominated by numerical solutions of the Poisson-Boltzmann (PB) equation for continuous charge distributions. Three algorithmic improvements were introduced to reduce the computation time of this step: coarsening of the inner region, linear interpolation of Debye-Hückel boundary values, and modified stripping. It was shown that these approximations reduce the computational costs of the overhead by 60% and introduce only marginal errors.

Moreover, the efficiency of the resulting QM/MM/GSBP method was tested on a model system whose size could be increased systematically to determine the breakeven point. The GSBP reduces computational costs for systems with more than 12,500 atoms. For smaller systems, the additional terms that need to be computed for the GSBP even increase the computation time for a single energy and gradient evaluation. For system with a size that is common in QM/MM applications (*i.e.*, around 25,000 atoms), the GSBP reduces the computational costs of a single molecular dynamics (MD) step by about 70 %. Although the overhead is significant, the savings of only about 1,000 steps in an MD simulation balance the costs of the overhead. Even more impressive savings were observed for larger systems.

The main problem of the QM/MM/GSBP method with semiempirical QM

methods is the limited accuracy of the QM component. QM/MM studies mostly apply more accurate density functional theory (DFT) or ab initio methods for the QM region. Due to the computational costs of these methods, configurational sampling is not possible and those studies are limited to the calculation of minimum energy paths on the potential energy surface (PES). The GSBP, however, is by construction only efficient in MD simulations. Therefore, the solvated macromolecule boundary potential (SMBP) was developed that was designed with three objectives: efficiency in geometry optimizations, applicability with any QM/MM method, and conceptual similarity with the GSBP. These targets were met in a modular implementation by combining a self consistent reaction field (SCRF) procedure with a set of virtual surface charges to represent the boundary potential in the QM calculations. The outer macromolecule is represented by a boundary potential obtained from solution of the PB equation. Bulk solvent is treated as a polarizable dielectric continuum. The resulting SMBP method offers three new possibilities: (1) The SMBP can be used to model the effect of bulk solvent in QM/MM calculations. (2) Within the scheme of QM/MM-free energy perturbation (QM/MM-FEP), finite-temperature effects due to the dynamics of the MM region can be estimated based on a scan of the potential energy profile. Using molecular and electronic structures from QM/MM/SMBP geometry optimizations as input for QM/MM-FEP, the efficient GSBP can be employed in the sampling step of the QM/MM-FEP approach. This reduces the computational costs of this step by one order of magnitude. (3) The SMBP also offers significant speed-ups to QM/MM calculations if the outer solvent molecules are neglected. The time-consuming SCRF procedure can then be skipped and thousands of MM point charges are replaced by a small set of virtual surface charges in the QM calculation. This reduces the computational costs of a single QM/MM energy and gradient evaluation by about 50 %.

Both boundary potentials were tested for a broad range of systems. The accuracy of the GSBP was checked in a model system consisting of threonine solvated in a water sphere. In MD simulations of *p*-hydroxybenzoate hydroxylase (PHBH) and chorismate mutase (CM), the GSBP reproduced the electrostatic potential with high accuracy and only small deviations were observed from full QM/MM calculations, with deviations of computed free energy differences usually below 1 kcal/mol. The SMBP was found to reproduce the electrostatic potential with similarly high accuracy. It was applied to study spin state energy gaps in cytochrome P450cam and the PESs of the proton transfer reaction in solvated glycine, the hydroxylation reaction in PHBH, and the Claisen rearrangement in CM. The glycine system was problematic for the SMBP. Geometry optimizations with QM/MM and QM/MM/SMBP

Hamiltonians lead to different close-lying local minima. Due to the high flexibility and polarity of this system, small deviations in the molecular structure give rise to significant deviations of the reaction and activation energies for the individual configurations, although reasonable agreement was obtained for the mean value of the five configurations studied. For the enzymatic systems, the SMBP yielded high accuracy, and potential energy differences computed with the SMBP deviated typically by less than 0.3 kcal/mol from standard QM/MM results.

The accuracy and efficiency of GSBP and SMBP depend on the values that are chosen for their intrinsic parameters. Therefore, a transferable protocol was developed and applied to determine optimal values for these parameters that relies exclusively on fast single-point calculations. Since very similar results were obtained for the three different systems CM, PHBH, and glycine in water, it is expected that the resulting recommended parameter set is transferable to general enzymatic systems.

The effects of long range electrostatics have two sources: the electrostatic potential of the outer macromolecule region (EPOM) and of bulk solvent. Since both contributions can be distinguished with SMBP and GSBP, they are applied to study the effect of the EPOM and bulk solvent on enzymatic reactions. Two reactions with rather different characteristics were studied: While little charge transfer is associated with the intramolecular Claisen rearrangement in CM, the hydroxylation reaction in PHBH is of opposite character. In CM, neither the EPOM nor bulk solvent have a significant influence on the reaction energetics. Due to the stronger charge transfer character in PHBH, the EPOM and bulk solvent influence the reaction energetics significantly, and deviations of several kcal/mol were observed upon their neglect. Moreover, neglect of the EPOM was shown to result in a systematic overstabilization of the more polar state.

To summarize, two boundary potentials for application with hybrid QM/MM methods were presented. The GSBP was adapted for the QM/MM ansatz and the SMBP was newly developed. The GSBP targets MD simulations while the SMBP is efficient in geometry optimizations and single-point calculations. Together both boundary potentials cover the full range of standard applications of QM/MM methods. One major objective of extending QM/MM to a three-layer method is the reduction of computational costs. This was achieved through the boundary potentials presented in this thesis. For systems with a size typical for QM/MM applications, the GSBP reduces the computation time for single MD step in QM/MM simulations with semiempirical QM methods by about 70 %. When the GSBP is used in the context of QM/MM-FEP simulations, even more impressive savings are possible and computation times are reduced by up to 90 %. By means

of algorithmic improvements and the realization that smaller basis sets are sufficient to reproduce the effect of bulk solvent, the overhead of the GSBP was reduced by more than 85 %. If the SMBP is not used to model bulk solvent but to lower the computational costs of standard QM/MM calculations in vacuo, the computation time of a single energy and gradient evaluation is reduced by about 50 %.
The computational effort that is saved may now be invested to use more accurate QM or MM methods. The SMBP can be applied with correlated ab initio QM methods without modifications. For both boundary potentials a combination with polarizable force fields is another logical step forward toward a more realistic description of enzymatic processes.

Bibliography

[1] McCammon, J. A.; Gelin, B. R.; Karplus, M. *Nature* **1977**, *267*, 585–590.

[2] Snow, C. D.; Nguyen, H.; Pande, V. S.; Gruebele, M. *Nature* **2002**, *420*, 102–106.

[3] Karplus, M.; McCammon, J. A. *Nat. Struct. Mol. Biol.* **2002**, *9*, 646–652.

[4] van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. *Angew. Chem. Int. Ed.* **2006**, *45*, 4060–4092.

[5] Ponder, J. W.; Case, D. A. *Adv. Prot. Chem.* **2003**, *66*, 27–85.

[6] Warshel, A.; Kato, M.; Pisiliakov, A. V. *J. Chem. Theory Comput.* **2007**, *3*, 2034–2045.

[7] MacKerell Jr., A. D. *J. Comput. Chem.* **2004**, *25*, 1584–1604.

[8] Warshel, A.; Levitt, M. *J. Mol. Biol.* **1976**, *103*, 227–249.

[9] Field, M. J.; Bash, P. A.; Karplus, M. *J. Comp. Chem.* **1990**, *11*, 700–733.

[10] Singh, U. C.; Kollman, P. A. *J. Comp. Chem.* **1986**, *7*, 718–730.

[11] Claeyssens, F.; Harvey, J. N.; Manby, F. R.; Mata, R. A.; Mulholland, A. J.; Ranaghan, K. E.; Schütz, M.; Thiel, S.; Thiel, W.; Werner, H.-J. *Angew. Chem.* **2006**, *45*, 6856–6859.

[12] Mata, R. A.; Werner, H.-J.; Thiel, S.; Thiel, W. *J. Chem. Phys.* **2008**, *128*, 025104/1–8.

[13] Dieterich, J. M.; Werner, H.-J.; Mata, R. A.; Metz, S.; Thiel, W. *J. Chem. Phys.* **2010**, *132*, 035101/1–10.

[14] Hu, H.; Yang, W. *Annu. Rev. Phys. Chem.* **2008**, *59*, 573–601.

[15] Zhang, Y.; Liu, H.; Yang, W. *J. Chem. Phys.* **2000**, *112*, 3483–3492.

[16] Cisneros, G. A.; Liu, H.; Zhang, Y.; Yang, W. *J. Am. Chem. Soc.* **2003**, *125*, 10384–10393.

[17] Ridder, L.; Rietjens, I. M. C. M.; Vervoort, J.; Mulholland, A. J. *J. Am. Chem. Soc.* **2002**, *124*, 9926–9936.

[18] Kaminski, G. A.; Jorgensen, W. L. *J. Phys. Chem. B* **1998**, *102*, 1787–1796.

[19] Acevedo, O.; Jorgensen, W. L.; Evanseck, J. D. *J. Chem. Theory Comput.* **2007**, *3*, 132–138.

[20] Rod, T. H.; Ryde, U. *J. Chem. Theory Comput.* **2005**, *1*, 1240–1251.

[21] Strajbl, M.; Hong, G.; Warshel, A. *J. Phys. Chem. B* **2002**, *106*, 13333–13343.

[22] Senn, H. M.; Thiel, S.; Thiel, W. *J. Chem. Theory Comput.* **2005**, *1*, 494–505.

[23] Kästner, J.; Senn, H. M.; Thiel, S.; Otte, N.; Thiel, W. *J. Chem. Theory Comput.* **2006**, *2*, 452–461.

[24] Senn, H. M.; Thiel, W. *Top. Curr. Chem.* **2007**, *268*, 173–290.

[25] Warshel, A.; Papazyan, A. *Curr. Opin. Struct. Biol.* **1998**, *2*, 211–217.

[26] Sagui, C.; Darden, T. A. *Annu. Rev. Biophys. Struct.* **1999**, *28*, 155–179.

[27] Davis, M. E.; McCammon, J. A. *Chem. Rev.* **1990**, *90*, 509–521.

[28] Garcia-Viloca, M.; Gao, J.; Karplus, M.; Truhlar, D. G. *Science* **2004**, *303*, 186–195.

[29] Ewald, P. *Ann. Phys.* **1921**, *369*, 253–287.

[30] Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

[31] Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

[32] Nam, K.; Gao, J.; York, D. M. *J. Chem. Theory Comput.* **2005**, *1*, 2–13.

[33] Gao, J.; Alhambra, C. *J. Chem. Phys.* **1997**, *107*, 1212–1217.

[34] Walker, R. C.; Crowley, M. F.; Case, D. A. *J. Comput. Chem.* **2008**, *29*, 1019–1031.

[35] Yarne, D. A.; Tuckerman, M. E.; Martyna, G. J. *J. Chem. Phys.* **2001**, *115*, 3531–3539.

[36] Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. *J. Chem. Theory Comput.* **2005**, *1*, 1176–1184.

[37] Laino, T.; Mohamed, F.; Laio, A.; Parrinello, M. *J. Chem. Theory Comput.* **2006**, *2*, 1370–1378.

[38] Hünenberger, P. H.; McCammon, J. A. *Biophys. Chem.* **1999**, *78*, 69–88.

[39] Hünenberger, P. H.; McCammon, J. A. *J. Chem. Phys.* **1999**, *110*, 1856–1872.

[40] Kuwajima, S.; Warshel, A. *J. Chem. Phys.* **1988**, *89*, 3751–3759.

[41] Weber, W.; Hunenberger, P. H.; McCammon, J. A. *J. Phys. Chem. B* **2000**, *104*, 3668–3675.

[42] Friedman, H. L. *Mol. Phys.* **1975**, *29*, 1533–1543.

[43] Wang, L.; Hermans, J. *J. Phys. Chem.* **1995**, *99*, 12001–12007.

[44] Berkowitz, M.; McCammon, J. A. *Chem. Phys. Lett.* **1982**, *90*, 215–217.

[45] Brooks, C. L., III; Karplus, M. *J. Chem. Phys.* **1983**, *79*, 6312–6325.

[46] Brunger, A.; Brooks, C. L., III; Karplus, M. *Chem. Phys. Lett.* **1984**, *105*, 495–500.

[47] Brunger, A.; Brooks, C. L., III; Karplus, M. *Proc. Natl. Acad. Sci. USA* **1985**, *82*, 8458–8462.

[48] Lee, F. S.; Warshel, A. *J. Chem. Phys.* **1992**, *97*, 3100–3107.

[49] Alper, H.; Levy, R. M. *J. Chem. Phys.* **1993**, *99*, 9847–9852.

[50] Essex, J. W.; Jorgensen, W. L. *J. Comput. Chem.* **1995**, *16*, 951–972.

[51] Warshel, A.; King, G. *Chem. Phys. Lett.* **1985**, *121*, 124–129.

[52] Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451–5459.

[53] Beglov, D.; Roux, B. *J. Chem. Phys.* **1994**, *100*, 9050–9063.

[54] Im, W.; Bernèche, S.; Roux, B. *J.Chem. Phys.* **2001**, *114*, 2924–2937.

[55] Banavali, N. K.; Im, W.; Roux, B. *J. Chem. Phys.* **2002**, *117*, 7381–7388.

[56] Schaefer, P.; Riccardi, D.; Cui, Q. *J. Chem. Phys* **2005**, *123*, 014905/1–14.

[57] Elstner, M.; Porezag, D.; Jungnickel, G.; Elstner, J.; Haugk, M.; Frauenheim, T.; Suhai, T.; Seifert, G. *Phys. Rev. B* **1998**, *58*, 7260–7268.

[58] Riccardi, D.; Schaefer, P.; Yang, Y.; Yu, H.; Ghosh, N.; Prat-Resina, X.; König, P.; Li, G.; Xu, D.; Guo, H.; Elstner, M.; Cui, Q. *J. Phys. Chem. B* **2006**, *110*, 6458–6469.

[59] Riccardi, D.; Cui, Q. *J. Phys. Chem. A* **2007**, *111*, 5703–5711.

[60] Riccardi, D.; Schaefer, P.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 17715–17733.

[61] König, P. H.; Ghosh, N.; Hoffmann, M.; Elstner, M.; Tajkhorshid, E.; Frauenheim, T.; Cui, Q. *J. Phys. Chem. A* **2006**, *110*, 548–563.

[62] Ma, L.; Cui, Q. *J. Am. Chem. Soc.* **2007**, *129*, 10261–10268.

[63] Zhu, X.; Jethiray, A.; Cui, Q. *J. Chem. Theory Comput.* **2007**, *3*, 1538–1549.

[64] Riccardi, D.; König, P.; Prat-Resina, X.; Yu, H.; Elstner, M.; Frauenheim, T.; Cui, Q. *J. Am. Chem. Soc.* **2006**, *128*, 16302–16311.

[65] Riccardi, D.; Li, G. H.; Cui, Q. *J. Phys. Chem. B* **2004**, *108*, 6467–6478.

[66] Grimme, S. *J. Comp. Chem.* **2004**, *25*, 1463–1473.

[67] Antes, I.; Thiel, W. *J. Phys. Chem. A* **1999**, *103*, 9290–9295.

[68] Zhang, Y.; Lee, T.-S.; Yang, W. *J. Chem. Phys.* **1999**, *110*, 46–54.

[69] Bessac, F.; Alary, F.; Carissan, Y.; Heully, J.-L.; Daudey, J.-P.; Poteau, R. *THEOCHEM* **2003**, *632*, 43–59.

[70] DiLabio, G. A.; Hurley, M. M.; Christiansen, P. A. *J. Chem. Phys.* **2002**, *116*, 9578–9584.

[71] Yasuda, K.; Yamaki, D. *J. Chem. Phys.* **2004**, *121*, 3964–3972.

[72] Lilienfeld, O. A.; Tavernelli, I.; Rothlisberger, U.; Sebastiani, D. *J. Chem. Phys.* **2005**, *122*, 014113/1–6.

[73] Théry, V.; Rinaldi, D.; Rivail, J.-L.; Maigret, B.; Ferenczy, G. G. *J. Comput. Chem.* **1994**, *15*, 269–282.

[74] Murphy, R. B.; Philipp, D. M.; Friesner, R. A. *J. Comput. Chem.* **2000**, *21*, 1442–1457.

[75] Gao, J.; Amara, P.; Alhambra, C.; Field, M. J. *J. Phys. Chem. A* **1998**, *102*, 4714–4721.

[76] Sherwood, P.; de Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjovoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. *J. Mol. Struct. (THEOCHEM)* **2003**, *632*, 1–28.

[77] Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. *Proteins* **1986**, *1*, 47–59.

[78] Burden, R. L.; Faires, J. D. *Numerical Analysis*; Prindle, Weber and Schmidt: Boston, 1985.

[79] Jackson, J. D. *Classical Electrodynamics*; John Wiley and Sons: New York, 1975.

[80] Nicholls, A.; Honig, B. *J. Comput. Chem.* **1991**, *12*, 435–445.

[81] Press, W. H.; Flannery, B. P.; Teukolsky, S. A.; Vetterlig, W. T. *Partial Differential Equations.* In *Numerical Recipes in C*; Cambridge University Press: Cambridge, England, 1988; pp 673–680.

[82] Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420–1426.

[83] Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580–10594.

[84] Jensen, F. *Transition State Theory and Statistical Mechanics.* In *Introduction to Computational Chemistry*; John Wiley & Sons: Chichester; England, 1999; pp 301–304.

[85] Gonzalez-Lafont, A.; Truong, T. N.; Truhlar, D. G. *J. Phys. Chem.* **1991**, *95*, 4618–4627.

[86] Rossi, I.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *233*, 231–236.

[87] Bash, P. A.; Ho, L. L.; MacKerell Jr., A. D.; Levine, D.; Hallstrom, P. *Proc. Nat. Ac. Sci. U.S.A.* **1996**, *93*, 3698–3703.

[88] Lau, E. Y.; Kahn, K.; Bash, P. A.; Bruice, T. C. *Proc. Nat. Ac. Sci. U.S.A.* **2000**, *97*, 9937–9942.

[89] Ridder, L.; Rietjens, I. M. C. M.; Vervoort, J.; Mulholland, A. J. *J. Am. Chem. Soc.* **2002**, *124*, 9926–9936.

[90] Otte, N.; Scholten, M.; Thiel, W. *J. Phys. Chem. A* **2007**, *111*, 5751–5755.

[91] Amin, E. A.; Truhlar, D. G. *J. Chem. Theory Comput.* **2008**, *4*, 75–85.

[92] Mulliken, R. S. *J. Chem. Phys.* **1962**, *36*, 3428–3439.

[93] Patchkovskii, S.; Thiel, W. *J. Comp. Chem.* **1996**, *17*, 1318–1327.

[94] Yamaguchi, Y.; Osamura, Y.; Goddard, J. D.; Schaefer, H. F., III *A New Dimension to Quantum Chemistry: Analytic Derivative Methods in Ab Initio Molecular Electronic Structure Theory*; Oxford University Press: Oxford, 1994; pp 128–132.

[95] Gilson, M. K.; Sharp, K. A.; Honig, B. H. *J. Comput. Chem.* **1987**, *9*, 327–335.

[96] Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902–3909.

[97] Smith, W.; Forester, T. *J. Mol. Graph.* **1996**, *14*, 136–141.

[98] MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

[99] Wu, R.; Wang, S.; Zhou, N.; Cao, Z.; Zhang, Y. *J. Am. Chem. Soc.* **2010**, *132*, 9471–9479.

[100] Ke, Z.; Wang, S.; Xie, D.; Zhang, Y. *J. Phys. Chem. B* **2009**, *113*, 16705–16710.

[101] Klähn, M.; Braun-Sand, S.; Rosta, E.; Warshel, A. *J. Phys. Chem. B* **2005**, *109*, 15645–15650.

[102] Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1984**, *106*, 3049–3059.

[103] Chandrasekhar, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 2974–2975.

[104] Jorgensen, W. L. *Acc. Chem. Res.* **1989**, *22*, 184–189.

[105] Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1985**, *107*, 154–163.

[106] Kuhn, B.; Kollmann, P. A. *J. Am. Chem. Soc.* **2000**, *122*, 2586–2596.

[107] Kollmann, P. A.; Kuhn, B.; Donini, O.; Perakyla, M.; Stanton, R.; Bakowies, D. *Acc. Chem. Res.* **2001**, *34*, 72–79.

[108] Bentzien, J.; Mueller, R. P.; Floriá, J.; Warshel, A. *J. Phys. Chem. B* **1998**, *102*, 2293–2301.

[109] Olsson, M. H. M.; Hong, G.; Warshel, A. *J. Am. Chem. Soc.* **2003**, *125*, 5025–5039.

[110] Warshel, A. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 425–443.

[111] Cisneros, G. A.; Wang, M.; Silinski, P.; Fitzgerald, M. C.; Yang, W. *Biochemistry* **2004**, *43*, 6885–6892.

[112] Wang, M.; Lu, Z.; Yang, W. *J. Chem. Phys.* **2004**, *121*, 101–107.

[113] Donini, O.; Darden, T.; Kollman, P. A. *J. Am. Chem. Soc.* **2000**, *122*, 12270–12280.

[114] Benighaus, T.; Thiel, W. *J. Chem. Theory Comp.* **2008**, *4*, 1600–1609.

[115] Tannor, D. J.; Marten, B.; Murphy, R.; Friesner, R. A.; Sitkoff, D.; Nicholls, A.; Honig, B.; Ringnalda, M.; Goddard, W. A. *J. Am. Chem. Soc.* **1994**, *116*, 11875–11882.

[116] Hayik, S. A.; Liao, N.; Merz, K. M. *J. Chem. Theory Comput.* **2008**, *4*, 1200–1207.

[117] Im, W.; Beglov, D.; Roux, B. *Comp. Phys. Comm.* **1998**, *111*, 59–75.

[118] ChemShell. http://www.chemshell.org (accessed August 14, 2009).

[119] Thiel, W. *MNDO2004*; Max-Planck-Institut für Kohlenforschung: Mülheim an der Ruhr, Germany, 2004.

[120] Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.

[121] Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177–2186.

[122] Im, W.; Seefeld, S.; Roux, B. *Biophys. J.* **2000**, *79*, 788–801.

[123] Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511–519.

[124] Nosé, S. *Mol. Phys.* **1984**, *52*, 255–268.

[125] Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.

[126] Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635–2643.

[127] Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.

[128] Bandyopadhyay, P.; Gordon, M. S. *J. Chem. Phys.* **2000**, *113*, 1104–1109.

[129] Bandyopadhyay, P.; Gordon, M. S.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **2002**, *116*, 5023–5032.

[130] Gontrani, L.; Mennucci, B.; Tomasi, J. *J. Mol. Struct. (THEOCHEM)* **2000**, *500*, 113–127.

[131] Cui, Q. *J. Chem. Phys.* **2002**, *117*, 4720–4728.

[132] Ridder, L.; Mulholland, A. J.; Vervoort, J.; Rietjens, I. M. C. M. *J. Am. Chem. Soc.* **1998**, *120*, 7641–7642.

[133] Ridder, L.; Harvey, J. N.; Rietjens, I. M. C. M.; Vervoort, J.; Mulholland, A. J. *J. Phys. Chem. B* **2003**, *107*, 2118–2126.

[134] Billeter, S. R.; Hanser, C. F. W.; Mordasini, T. Z.; Scholten, M.; Thiel, W.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys.* **2001**, *3*, 688–695.

[135] Senn, H. M.; Thiel, W. *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.

[136] Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. *Chem. Rev.* **2005**, *105*, 2279–2328.

[137] Shaik, S.; Cohen, S.; Wang, Y.; Chen, H.; Kumar, D.; Thiel, W. *Chem. Rev.* **2010**, *110*, 949–1017.

[138] Altun, A.; Thiel, W. *J. Phys. Chem. B* **2005**, *109*, 1268–1280.

[139] Woo, H.-J.; Dinner, A. R.; Roux, B. *J. Chem. Phys.* **2004**, *121*, 6392–6400.

[140] Deng, Y.; Roux, B. *J. Chem. Phys.* **2008**, *128*, 115103/1–8.

[141] Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

[142] Becke, A. D. *Phys. Rev. A* **1988**, *38*, 3098–3100.

[143] Lee, C.; Yang, W.; Parr, R. G. *Phys. Rev. B* **1988**, *37*, 785–789.

[144] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

[145] Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571–2577.

[146] Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829–5835.

[147] Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Maves, S. A.; Benson, D. E.; Sweet, R. M.; Ringe, D.; Petsko, G. A.; Sligar, S. G. *Science* **2000**, *287*, 1615–1622.

[148] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 1372–1377.

[149] Wachters, A. J. H. *J. Chem. Phys.* **1970**, *52*, 1033–1036.

[150] Hay, P. J. *J. Chem. Phys.* **1977**, *66*, 4377–4384.

[151] Bauschlicher, C. W., Jr.; Langhoff, S. R.; Partridge, H.; Barnes, L. A. *J. Chem. Phys.* **1989**, *91*, 2399–2411.

[152] Senn, H. M.; Kästner, J.; Breidung, J.; Thiel, W. *Can. J. Chem.* **2009**, *87*, 1322–1337.

[153] Benighaus, T.; Thiel, W. *J. Chem. Theory Comput.* **2009**, *5*, 3114–3128.

[154] Kast, P.; Aisf-Ullah, M.; Hilvert, D. *Tetrahedron Lett.* **1996**, *37*, 2691–2694.

Appendix A

QM/MM Studies of the Coupling and Uncoupling Mechanisms in Cytochrome P450cam and its Mutants

In addition to the development work that is presented in this thesis, I contributed to an application project that adressed one step of the catalytic cycle of cytochrome P450cam. We used QM/MM methods to study the effect of mutation of two residues which are believed to be important for the catalytic effect (threonine 252 and aspartate 251) and proposed reaction mechanisms for this step in the wild type enzyme and in the mutants. This work was published in three articles that are appended to this thesis to document this work.

J|A|C|S
ARTICLES

Published on Web 03/12/2009

How is the Reactivity of Cytochrome P450cam Affected by Thr252X Mutation? A QM/MM Study for X = Serine, Valine, Alanine, Glycine

Muhannad Altarsha, Tobias Benighaus, Devesh Kumar, and Walter Thiel*

Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, D-45470 Mülheim an der Ruhr, Germany

Received November 7, 2008; E-mail: thiel@mpi-muelheim.mpg.de

Abstract: Proton transfer reactions play a vital role in the catalytic cycle of cytochrome P450cam and are responsible for the formation of the iron-oxo species called Compound I (Cpd I) that is supposed to be the active oxidant. Depending on the course of the proton transfer, protonation of the last observable intermediate (ferric hydroperoxo complex, Cpd 0) can lead to either the formation of Cpd I (coupling reaction) or the ferric resting state (uncoupling reaction). The ratio of these two processes is drastically affected by mutation of the Thr252 residue. In this work, we study the effect of Thr252X (X = serine, valine, alanine, glycine) mutations on the formation of Cpd I by means of hybrid quantum mechanical/molecular mechanical (QM/MM) calculations and classical simulations. In the wild-type enzyme, the coupling reaction is favored since its rate-limiting barrier is 13 kcal/mol lower than that for uncoupling. This difference is reduced to 7 kcal/mol in the serine mutant. In the case of valine, alanine, and glycine mutants, an additional water molecule enters the active site and lowers the activation energy of the uncoupling reaction significantly. With the additional water molecule, coupling and uncoupling have similar barriers in the valine mutant, and the uncoupling reaction becomes favored in the alanine and glycine mutants. These findings agree very well with experimental results and thus confirm the assumption that uncontrolled proton delivery by solvent water networks is responsible for the uncoupling reaction. The present study provides a detailed mechanistic understanding of the role of the Thr252 residue.

I. Introduction

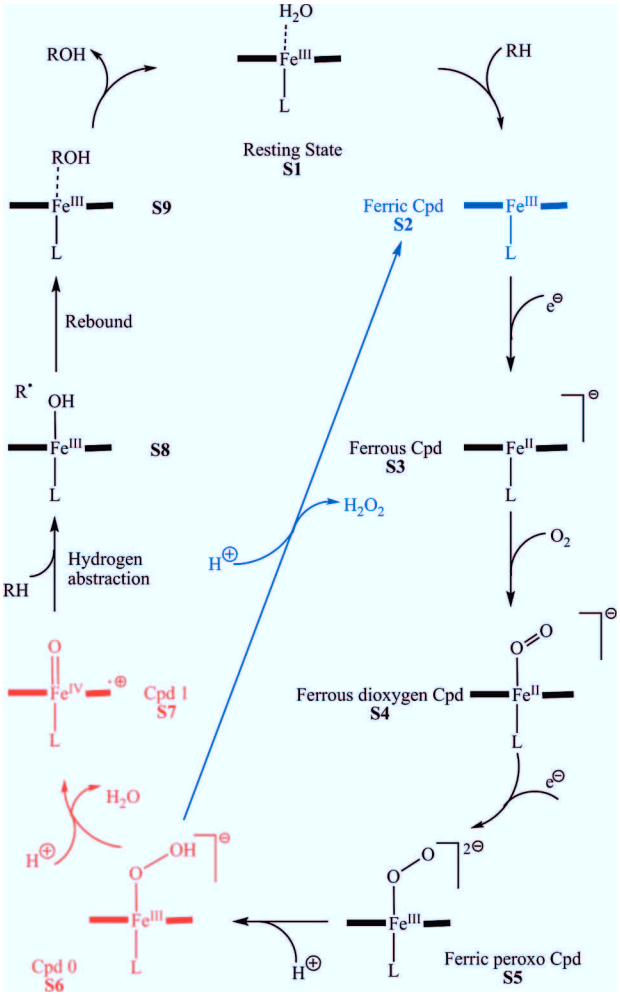
The heme protein monooxygenases known as cytochrome P450 play a vital role in the metabolism of xenobiotic substances in plants, fungi, bacteria, insects, and mammals.[1−3] They catalyze a variety of reactions including hydroxylation, epoxidation, and heteroatom oxidation. The consensus mechanism of P450 hydroxylation is shown in Scheme 1, where the proximal cysteinate ligand is abbreviated as L, and the porphyrin macrocycle is symbolized by the two bold lines flanking the iron.[4] The catalytic cycle begins with the resting state (**S1**) in which a water molecule is bound to the ferric ion in the distal side. When a substrate enters the protein pocket, the water molecule is displaced from the reactive center, leaving a ferric substrate-bound state (ferric resting state, **S2**). This complex is a slightly better electron acceptor than the resting state and can therefore take up an electron from reductase protein, leading to a high-spin ferrous complex (**S3**). Subsequent binding of molecular oxygen produces the oxyferrous state (**S4**). Addition of a second electron yields a peroxo-ferric derivative (**S5**) that, upon protonation of the distal oxygen, forms a hydroperoxo-ferric intermediate which is also known as Compound 0 (Cpd 0, **S6**). A second protonation then leads to O−O bond cleavage with loss of a water molecule and generates the "active oxygen" in the putatively active oxoferryl species Compound I (Cpd I, **S7**). Alternatively, protonation of the proximal oxygen of Cpd 0 (**S6**) leads to release of hydrogen peroxide and regenerates the ferric resting state (**S2** in Scheme 1).

In recent years, much attention has focused on possible mechanisms for the required proton delivery to the active site. On the basis of extensive site-directed mutation studies,[5−10] it has been suggested that the Asp251 and Thr252 residues participate in a controlled proton delivery pathway that involves solvent water and provides an active-site H-bond donor, which may be a trapped water molecule rather than Thr252.[5,8] The Thr252 residue is highly conserved in P450cam crystal structures and is believed to play an important role in the dioxygen activation machinery.[11]

(1) Guengerich, F. P. *J. Biol. Chem.* **1991**, *266*, 10019–10022.
(2) Porter, T. D.; Coon, M. J. *J. Biol. Chem.* **1991**, *266*, 13469–13472.
(3) (a) Ortiz de Montellano, P. R. *Cytochrome P450: Structure, Mechanism and Biochemistry*, 2nd ed.; Plenum Press: New York, 1995. (b) Ortiz de Montellano, P. R. *Cytochrome P450: Structure, Mechanism and Biochemistry*, 3rd ed.; Plenum Press: New York, 2005.
(4) Davydov, R.; Macdonald, I. D. G.; Makris, T. M.; Sugar, S. G.; Hoffman, B. M. *J. Am. Chem. Soc.* **1999**, *121*, 10654–10655.
(5) Vidakovic, M.; Sligar, S. G.; Li, H.; Poulos, T. L. *Biochemistry* **1998**, *37*, 9211–9219.
(6) Martinis, S. A.; Atkins, W. M.; Stayton, P. S.; Sligar, S. G. *J. Am. Chem. Soc.* **1989**, *111*, 9252–9253.
(7) Imai, M.; Shimada, H.; Watanabe, Y.; Matsuhima-Hibiya, Y.; Makino, R.; Koga, H.; Horiuchi, T.; Ishimura, Y. *Proc. Natl. Acad. Sci. U.S.A.* **1989**, *86*, 7823–7827.
(8) Gerber, N. C.; Sligar, S. G. *J. Biol. Chem.* **1994**, *269*, 4260–4266.
(9) Gerber, N. C.; Sligar, S. G. *J. Am. Chem. Soc.* **1992**, *114*, 8742–8743.
(10) Benson, D. E.; Suslick, K. S.; Sligar, S. G. *Biochemistry* **1997**, *36*, 5104–5107.

10.1021/ja808744k CCC: $40.75

Scheme 1. Cytochrome P450cam Reaction Cycle



Site-directed mutagenesis studies[6−9] indicate that the presence of Thr252 is essential for the formation of Cpd I in P450cam. Upon replacement by an aliphatic residue in mutants such as Thr252Ala (T252A)[7,12] and Thr252Gly (T252G),[7] there is an uncoupling of O_2 consumption from camphor hydroxylation: most of O_2 consumed is converted to H_2O_2 without O−O bond cleavage,[7] and only 5% (3%) of the T252A (T252G) mutants undergo hydroxylation. By contrast, in the wild-type enzyme, the coupling ratio of O_2 consumption and hydroxylation is 100%. According to a low-temperature EPR and ENDOR study, both wild-type P450cam and the T252A mutant form Cpd 0 at 77 K, but only the former gives hydroxylation upon warming while the latter shows uncoupling, thus confirming the key role of Thr252 during the delivery of the second proton in the catalytic cycle (from **S6** to **S7**).[4] Given that Cpd I formation is suppressed in the T252A mutant, the observed reactivity toward other substrates may suggest the existence of more than one oxidant in P450 enzymes.[13] For example, the T252A mutant is capable of epoxidizing the double bond of olefins such as 5-methylenylcamphor, presumably with Cpd 0 being the active species in this case.[14] However, Cpd 0 is generally found to be a sluggish oxidant that is not competitive with Cpd I and may only function as such in the absence of Cpd I.[14−17]

Substitution of Thr252 with valine (T252V) also causes a considerable uncoupling of oxygen consumption and hydroxylation, with about 30−40% of hydrogen peroxide formation.[7,18] Substitution with serine (T252S), a hydroxyamino acid, has a smaller effect: only approximately 15% of O_2 consumed is recovered as H_2O_2.[7,18,19] Replacing the Thr252 side-chain hydroxyl (OH) with a methoxy (OCH_3) group slows the hydroxylation reaction significantly, but the yield of 5-exohydroxycamphor is still 100% (no uncoupling).[19]

Taken together, these experimental findings indicate that the Thr252 residue plays a crucial mechanistic role and influences the ratio of hydroxylation versus uncoupling products. Although a free hydroxy group at position 252 is not a prerequisite for O−O bond cleavage in P450cam, the replacement of Thr252 by amino acids with non-hydrogen-bonding side chains virtually suppresses camphor hydroxylation in favor of the uncoupled reduction of O_2 to H_2O_2.[7,18,20]

In the crystal structure of the T252A mutant, there is a water molecule near the O_2 binding site of the mutant which is not present in the wild-type enzyme.[11,21] This suggests that the solvent may be responsible for the observed uncoupling of the enzyme turnover from camphor hydroxylation in the mutant:[12,21] solvent in contact with dioxygen may supply protons in an uncontrolled manner, promoting H_2O_2 production, rather than substrate hydroxylation.

The available crystal structures[22] of wild-type P450cam indicate three possible proton delivery pathways through the Asp251,[5,23,24] Glu366,[22,25] and Arg299[26] channels. The latter has been characterized through a combined site-directed mutagenesis and molecular dynamics study[27] but is blocked when substrate is present. In the Glu366 channel, the bridging water molecules (nos. 687, 566, 523, and 902 in PDB structure 1DZ8[22]) and the hydroxyl group of Thr252 form a network connecting the carboxyl group of Glu366 and the distal oxygen atom, which however terminates at Glu366 without any connection to the surface.[5] In this paper, we shall therefore focus on the Asp251 channel only.

(11) Raag, R.; Martinis, S. A.; Sligar, S. G.; Poulos, T. L. *Biochemistry* **1991**, *30*, 11420–11429.
(12) Nagano, S.; Cupp-Vickery, J. R.; Poulos, T. L. *J. Biol. Chem.* **2005**, *280*, 22102–22107.
(13) Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I. *Chem. Rev.* **2005**, *105*, 2253–2278.
(14) Jin, S.; Makris, T. M.; Bryson, T. A.; Sligar, S. G.; Dawson, J. H. *J. Am. Chem. Soc.* **2003**, *125*, 3406–3407.
(15) De Visser, S. P.; Kumar, D.; Shaik, S. *J. Inorg. Biochem.* **2004**, *98*, 1183–1193.
(16) Li, C.; Zhang, L.; Zhang, C.; Hirao, H.; Wu, W.; Shaik, S. *Angew. Chem., Int. Ed.* **2007**, *46*, 8168–8170.
(17) Hirao, H.; Kumar, D.; Shaik, S. *J. Inorg. Biochem.* **2006**, *100*, 2054–2068.
(18) Shimada, H.; Makino, R.; Imai, M.; Horiuchi, T.; Ishimura, Y. Mechanism of Oxygen Activation by Cytochrome P-450cam. *International Symposium on Oxygenases and Oxygen Activation*; Yamada Science Foundation, 1990; pp 133−136.
(19) Kimata, Y.; Shimada, H.; Hirose, T.; Ishimura, Y. *Biochem. Biophys. Res. Commun.* **1995**, *208*, 96–102.
(20) Aikens, J.; Sligar, S. G. *J. Am. Chem. Soc.* **1994**, *116*, 1143–1144.
(21) Hishiki, T.; Shimada, H.; Nagano, S.; Egawa, T.; Kanamori, Y.; Makino, R.; Park, S. Y.; Adachi, S. I.; Shiro, Y.; Ishimura, Y. *J. Biochem.* **2000**, *128*, 965–974.
(22) Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Maves, S. A.; Benson, D. E.; Sweet, R. M.; Ringe, D.; Petsko, G. A.; Sligar, S. G. *Science* **2000**, *287*, 1615–1622.
(23) Taraphder, S.; Hummer, G. *J. Am. Chem. Soc.* **2003**, *125*, 3931–3940.
(24) Kamachi, T.; Yoshizawa, K. *J. Am. Chem. Soc.* **2003**, *125*, 4652–4661.
(25) Guallar, V.; Friesner, R. A. *J. Am. Chem. Soc.* **2004**, *126*, 8501–8508.
(26) Oprea, T. I.; Hummer, G.; García, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 2133–2138.
(27) Loida, P. J.; Sligar, S. G.; Paulsen, M. D.; Arnold, G. E.; Ornstein, R. L. *J. Biol. Chem.* **1995**, *270*, 5326–5330.

Previous theoretical studies have addressed the formation of Cpd I from Cpd 0 in the wild-type enzyme using gas-phase model systems.[23,25,28−32] In the most elaborate such work, density functional theory (DFT) calculations on a large active site model with 96 atoms predict the existence of a protonated Cpd 0 species with significant barriers for the conversion both toward Cpd 0 and Cpd I.[30] According to a later quantum mechanics/molecular mechanics (QM/MM) study,[33] protonation of Cpd 0 by Glu366 via a hydrogen-bonding network may indeed generate such an intermediate, which however is very high in energy (more than 20 kcal/mol above Cpd 0) and mechanistically irrelevant since the barrier for its decay is only 3−4 kcal/mol both toward Cpd 0 and Cpd I. The QM/MM calculations suggest another mechanism instead: an initial homolytic O−O bond cleavage generates an OH species and one-electron reduced Cpd I, and subsequent proton transfer to this OH species with concomitant electron transfer from the heme yields Cpd I and water. The rate-limiting step at the QM/MM level is O−O cleavage with a barrier of about 13−14 kcal/mol in both the Asp251 and Glu366 channels.

Here, we extend this previous QM/MM work to address the effect of Thr252 mutations on the formation of Cpd I and on the competition between coupling and uncoupling reactions, that is, conversion of Cpd 0 to Cpd I and water as opposed to the ferric resting state and hydrogen peroxide. We cover the wild-type P450cam enzyme and its mutants T252S, T252V, T252A, and T252G. We also consider the effect of an extra water molecule in the active site by means of classical MD simulations and QM/MM calculations for different mutants.

II. Computational Details

Starting from a crystallographic structure (pdb file: 1DZ8), we generated a Cpd 0 model for the T252S, T252V, T252A, and T252G mutants by manually replacing the threonine residue in the wild-type enzyme with serine, valine, alanine, and glycine, respectively. Solvation and protonation procedures followed standard protocols used previously in our group.[34−36] Asp251 was protonated since it serves as proton source in the present study. The resulting model with a net charge of $-8e$ was neutralized by protonating selected ionic residues at the surface of the enzyme without affecting salt bridges or hydrogen bonds.

The system was solvated using 5891 TIP3P water molecules[37] (yielding a total of ca. 25 000 atoms). It was relaxed by energy minimizations and classical molecular dynamics (MD) simulations using the CHARMM22 force field[38] as implemented in the CHARMM[39] program. Throughout all MD simulations, the coordinates of the entire heme unit and the coordinating Cys357 as well as the outer 8 Å of the solvent layer were kept fixed.

The applied QM/MM methodology is analogous to that used in our group in related studies.[33−36] Therefore, only some aspects relevant to the present work are presented here. An electronic embedding scheme[40] was adopted in the QM/MM calculations; that is, the MM charges were included in the one-electron Hamiltonian of the QM part, and the QM/MM electrostatic interactions were evaluated from the QM electrostatic potential and MM partial charges. No cutoffs were introduced for the nonbonding MM and QM/MM interactions. Hydrogen link atoms and the charge shift model[41] were employed to treat the QM/MM boundary. The TURBOMOLE[42] program was used for the QM calculations, while the CHARMM22 force field was run through the DL_POLY[43] code for the treatment of the MM part. The QM/MM calculations were performed with the ChemShell package[44] that integrates the TURBOMOLE and DL_POLY programs, and geometry optimizations were carried out using the HDLC optimizer[45] implemented in ChemShell.

The QM part was treated by the UB3LYP[46] density functional method with a LACVP[47] small-core ECP basis set (B1), that is, double ζ on Fe and 6-31G[48] on the rest of the atoms. The MM part was described by the CHARMM22 force field. For improved accuracy, all calculations were repeated using the larger TZVP[49,50] basis set (B2) on all atoms.

The optimized active region was defined to include all residues and water molecules within 6 Å of any atom of the core region. This results in ca. 1400 atoms to be optimized, which belong to the iron−dioxygen−porphyrin complex, camphor, and the amino acid residues and water molecules around the active site. All minima and transition states (TS) reported in this paper were fully optimized.

QM Region: For our present QM/MM calculations, we employed QM regions analogous to those adopted for the wild-type enzyme in our previous work[33] (see Figure 1). For the wild-type enzyme (model A): iron−porphine (without heme side chains), sulfur atom of Cys357, distal O_2H moiety, the $C^{\beta}H-O^{\mu 1}H$ unit of Thr252, Wat901, and the $C^{\beta}H_2-C^{\mu}(=O^{\delta 1})(-O^{\delta 2}H)$ unit of Asp251 (Figure 1). For the serine mutant (model B), the threonine side chain was replaced by the serine side chain, and hence only the $C^{\mu 2}H_3-O^{\mu 1}H$ unit was included in the QM region. The QM regions for the valine, alanine, and glycine mutants were obtained analogously (see models C, E, and G in Figure 1), with the $C^{\mu 2}H_3-C^{\beta}H-C^{\mu 2}H_3$ unit (C) and the $C^{\mu 2}H_4$ unit (E) being part of the QM region. Finally, an extra water molecule (WatS) in the QM region extends models C, E, and G to models D, F, and H, respectively.

III. Results and Discussion

1. Classical MD Results. Classical MD simulations were performed to check whether the Asp251 channel can support an additional water molecule (WatS) to form a more extended hydrogen-bonding network. The native enzyme has been studied

(28) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. *Chem. Rev.* **2005**, *105*, 2279–2328.
(29) Ogliaro, F.; De Visser, S. P.; Cohen, S.; Sharma, P. K.; Shaik, S. *J. Am. Chem. Soc.* **2002**, *124*, 2806–2817.
(30) Kumar, D.; Hirao, H.; De Visser, S. P.; Zheng, J.; Wang, D.; Thiel, W.; Shaik, S. *J. Phys. Chem. B* **2005**, *109*, 19946–19951.
(31) Kamachi, T.; Shiota, Y.; Ohta, T.; Yoshizawa, K. *Bull. Chem. Soc. Jpn.* **2003**, *76*, 721–732.
(32) Harris, D. L.; Loew, G. H. *J. Am. Chem. Soc.* **1998**, *120*, 8941–8948.
(33) Zheng, J.; Wang, D.; Thiel, W.; Shaik, S. *J. Am. Chem. Soc.* **2006**, *128*, 13204–13215.
(34) Schöneboom, J. C.; Thiel, W. *J. Phys. Chem. B* **2004**, *108*, 7468–7478.
(35) Schöneboom, J. C.; Lin, H.; Reuter, N.; Thiel, W.; Cohen, S.; Ogliaro, F.; Shaik, S. *J. Am. Chem. Soc.* **2002**, *124*, 8142–8151.
(36) Altun, A.; Thiel, W. *J. Phys. Chem. B* **2005**, *109*, 1268–1280.
(37) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(38) MacKerell, A. D., Jr.; et al. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
(39) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
(40) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580–10594.
(41) Sherwood, P.; De Vries, A. H.; Collins, S. J.; Greatbanks, S. P.; Burton, N. A.; Vincent, M. A.; Hillier, I. H. *Faraday Discuss.* **1997**, *106*, 79–92.
(42) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kölmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.
(43) Smith, W.; Forester, T. R. *J. Mol. Graph.* **1996**, *14*, 136–141.
(44) Sherwood, P.; et al. *J. Mol. Struct.* **2003**, *632*, 1–28.
(45) Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177–2186.
(46) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.
(47) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299–310.
(48) Hehre, W. J.; Ditchfield, K.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.
(49) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829–5835.
(50) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571–2577.

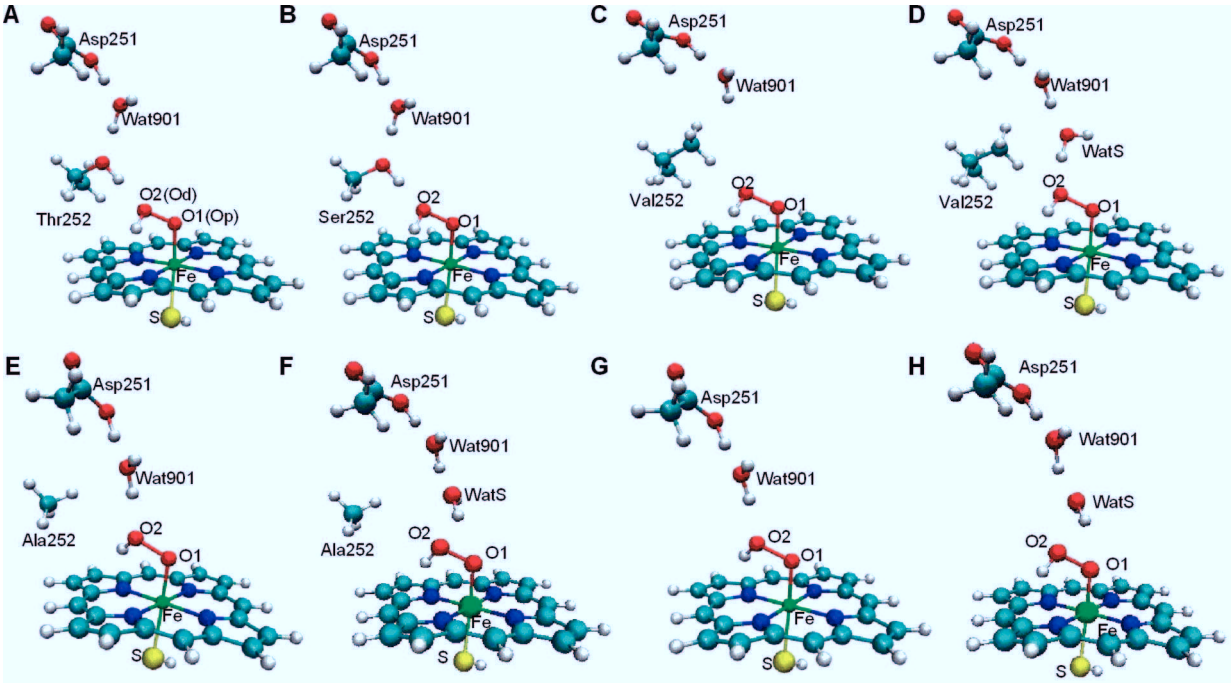


Figure 1. QM regions for the wild-type enzyme and its mutants. Models A, B, C, E, and G represent the native enzyme of P450cam, the T252S mutant, the T252V mutant, the T252A mutant, and the T252G mutant, respectively. Models D, F, and H extend models C, E, and G, respectively, with an extra water molecule (WatS).

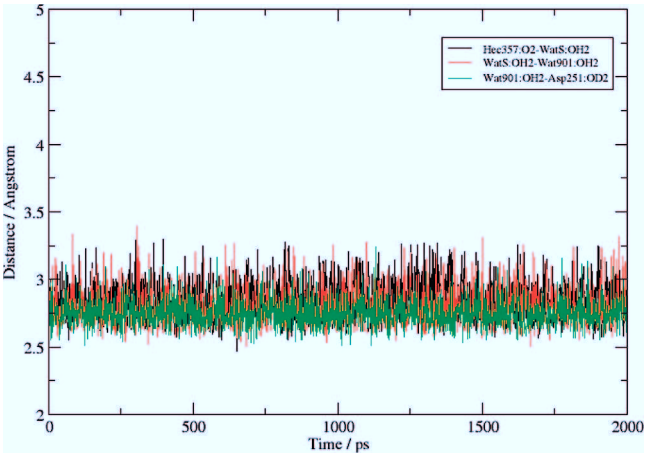


Figure 2. Monitoring the mobility of the extra water molecule (WatS) in the T252A mutant (for atom labels, see Figure 1F; Hec is the porphine-FeO2 unit).

previously in this regard, and it was found that the extra water molecule leaves the active site.[51] In the present MD simulations on the serine, valine, alanine, and glycine mutants of Cpd 0, we checked the mobility of WatS and, at the same time, tested the stability of the hydrogen-bonding network between Asp251 and the FeO_2H moiety.

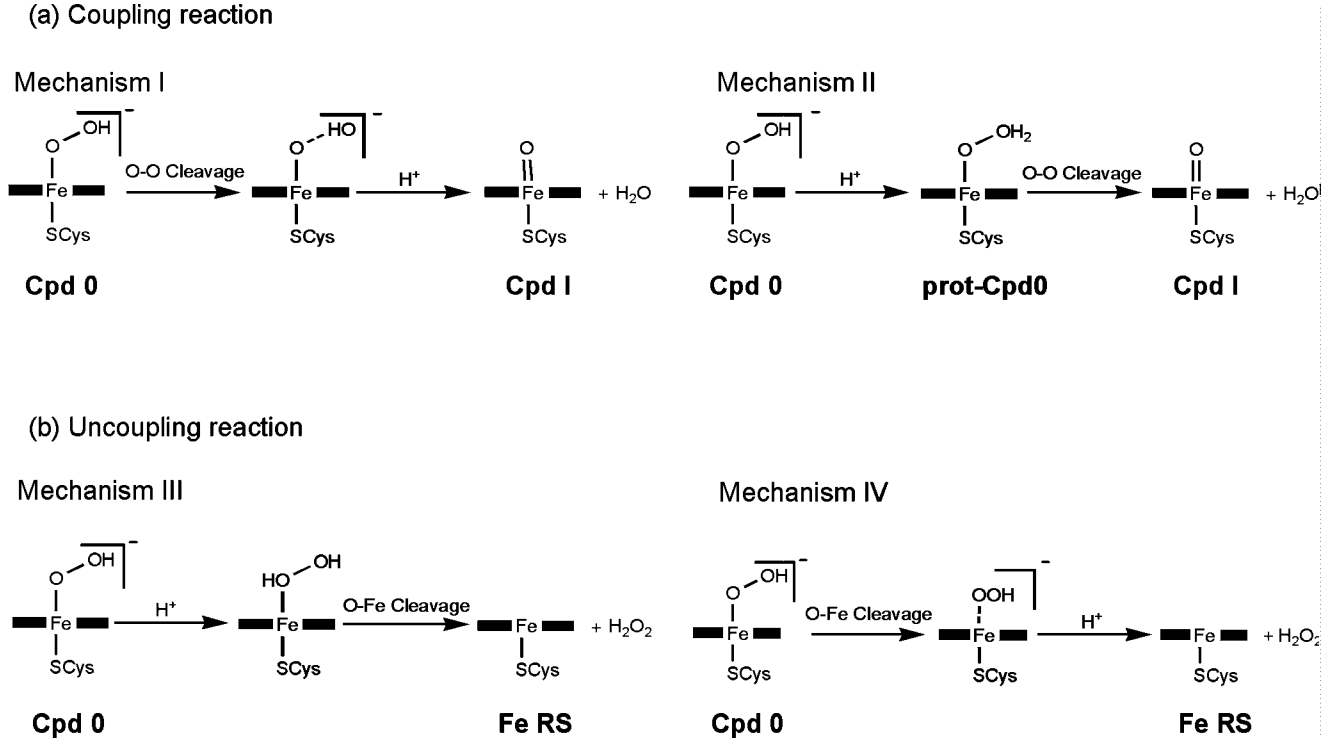
Substitution of threonine by alanine generates some empty space in the distal pocket so that WatS can find a stable position between Wat901 and the FeO_2H moiety. The classical MD results for the alanine mutant in Figure 2 show that Wat901 and WatS do not escape from the protein pocket during the 2 ns simulation. WatS stays close to the FeO_2H moiety for most of the time, whereas Wat901 forms two hydrogen bonds with WatS and the Asp251 residue (confirmed by average distances of 2.80 ± 0.13 and 2.75 ± 0.10 Å for WatS:OH2-Wat901:OH2 and Wat901:OH2-Asp251:OD2, respectively). The distance between the O_2H moiety and the oxygen atom of WatS also remains around 3 Å (2.85 ± 0.13 Å). Hence all these hydrogen bonds are conserved in the alanine mutant. This is consistent with the crystal structure of the T252A mutant which shows an extra water molecule (not present in the wild-type enzyme) near the O2 (distal oxygen) binding site of the mutant.[11,21] Likewise, the 2 ns MD simulations for the valine and glycine mutants confirm the stability of the additional water molecule (see Figures S2 and S3 in the Supporting Information). By contrast, the extra water molecule is not stable in the serine mutant since it escapes from the distal pocket very early in the MD simulation (see Figure S1 in the Supporting Information), in analogy to the behavior of the wild-type enzyme.

2. Survey of Possible Reaction Mechanisms. Scheme 2 shows the four mechanisms that were considered in the QM/MM calculations. In each case, the crystallographic water molecule W901 (in combination with WatS in models D, F, and H) supports proton transfer from Asp251 to the X252 residue (X = threonine, serine, valine, alanine, and glycine). The proton transferred to the FeO_2H moiety comes from the hydroxy group of threonine in the WT enzyme or serine in the T252S mutant. In the absence of a hydroxy group, such as in the valine, alanine, and glycine mutants, it is transferred from the closest water molecule, which is Wat901 in models C, E, and G, and WatS in models D, F, and H. The proton-donating sites are replenished by proton transfer along the hydrogen-bonding network in the Asp251 channel. The following mechanisms were studied for coupling (I,II) and uncoupling (III,IV).

Mechanism I: Initially, the O−O bond is cleaved homolytically to generate an OH species and a one-electron reduced Cpd

(51) Zheng, J.; Altun, A.; Thiel, W. *J. Comput. Chem.* **2007**, *28*, 2147–2158.

Scheme 2. Mechanisms for (a) Cpd I Formation (Coupling Reaction) and (b) Ferric Resting State Formation (Uncoupling Reaction)



I. A subsequent proton transfer to the OH species with a concomitant electron transfer from the heme yields Cpd I and water.[33]

Mechanism II: A proton is transferred to the distal oxygen atoms of the hydroperoxo group to form protonated Cpd 0 (prot-Cpd 0: $FeOOH_2$), followed by heterolytic O−O bond cleavage that generates Cpd I and water.[33] In the Asp251 channel, we were unable to locate a stable prot-Cpd 0 minimum for any of the mutants, as in the case of the native enzyme.[33] We thus conclude that mechanism I is preferred and only discuss this mechanism for the coupling reaction.

Mechanism III: A proton is transferred to the proximal oxygen atom of the hydroperoxo group to form FeH_2O_2, accompanied by heterolytic O−Fe bond cleavage that generates the ferric resting state and hydrogen peroxide.

Mechanism IV: Initially, the O−Fe bond is cleaved homolytically to generate an OOH radical, followed by a proton transfer to the OOH group to yield the ferric resting state and hydrogen peroxide. Mechanism IV is calculated to be much less favorable than mechanism III for the uncoupling reaction since the O−Fe cleavage reaction always requires much activation. Inclusion of an extra water molecule in the T252V, T252A, and T252G mutants has only a minor effect on this barrier which remains high. Therefore, we do not discuss the corresponding results here but only document them in the Supporting Information.

3. QM/MM Results. The QM/MM optimized structures for the QM region of Cpd 0 are shown in Figure 1 for the native enzyme and its mutants. In the native enzyme, Cpd 0 has a doublet ground state which is more stable than the quartet state both in the gas phase[29,31] and the enzyme environment.[33] The same also holds for the T252X mutants, and we shall therefore only discuss the doublet state in this paper. We have carried out full QM/MM optimizations for all relevant doublet species using both the B1 and B2 basis set. The geometries and energies obtained with the two basis sets are generally similar. Optimized geometries are shown in Figures 3 and 4 and in Figures S4−S16 of Supporting Information. Table 1 lists the relative energies of all optimized stationary points (both B1 and B2 values, relative to Cpd 0). In the text, we shall quote energy values computed with the larger B2 basis set (unless noted otherwise).

In addition, we have also performed full QM/MM optimizations for the lowest quartet and sextet states of Cpd 0 in the wild-type enzyme and each mutant using the B1 basis set. The Supporting Information documents the corresponding results along with those for the doublet ground state in detail, providing QM, MM, and QM/MM energies, spin densities, group charges, and selected geometric parameters (Tables S2−S13). The Supporting Information also contains spin densities and group charges for the doublet states of the optimized stationary points of mechanisms I, III, and IV (B1 basis).

3.1. Wild-Type Enzyme. Coupling Reaction: We have previously investigated the conversion of Cpd 0 to Cpd I in the wild-type enzyme at the QM/MM level using the B1 basis.[33] We have now reoptimized the stationary points using the B2 basis, which leads to minor changes only. The transition states for the initial O−O cleavage and the subsequent proton transfer lie 14.3 and 11.7 kcal/mol above Cpd 0, respectively, and the overall reaction energy is calculated to be −3.7 kcal/mol.

Uncoupling Reaction: In the favored mechanism III, the proton is transferred through the Asp251−Wat901−Thr252−HOOFe hydrogen-bonding chain to O1 (proximal oxygen) which triggers O−Fe bond cleavage and generates the ferric resting state and hydrogen peroxide. Figure 3 shows the corresponding QM/MM optimized structures. According to the QM/MM calculations, the reaction proceeds in a single step, with an activation barrier of 27.0 kcal/mol, and is endothermic by 6.4 kcal/mol. The formed hydrogen peroxide has almost zero spin density (see Supporting Information) in the product (ferric resting state, FeRS). The Fe−O1 distance increases from 1.87 Å in Cpd 0 to

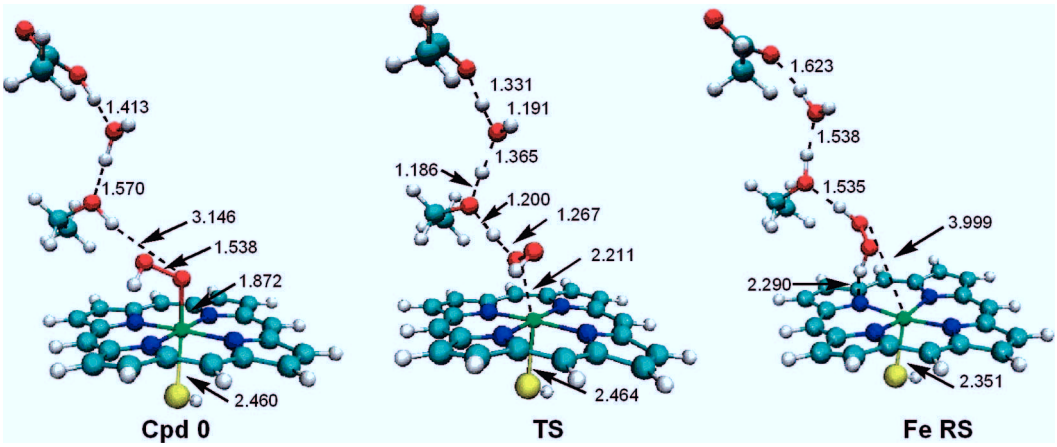


Figure 3. Optimized geometries (UB3LYP/B1/CHARMM) for the uncoupling reaction in the wild-type enzyme. Only the QM region is shown.

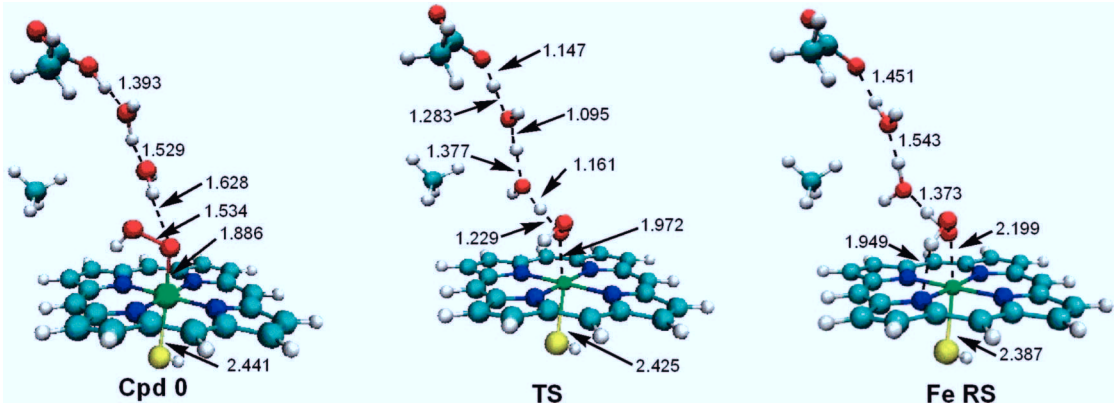


Figure 4. Optimized geometries (UB3LYP/B1/CHARMM) for the uncoupling reaction in the T252A mutant in the presence of an extra water molecule (WatS). Only the QM region is shown.

Table 1. Computed Relative Energies (kcal/mol) for Optimized Stationary Points Using Basis Sets B1/B2 (with respect to Cpd 0)[a]

mechanism	species	model A WT	model B T252S	model C T252V	model D T252V+W	model E T252A	model F T252A+W	model G T252G	model H T252G+W
I	TS1	14.4/14.3	15.8/15.6	17.1/17.1	19.0/18.9	17.4/17.1	17.5/17.3	18.8/17.9	19.3/18.9
	IC1	11.7/10.9	9.4/8.4	11.4/11.0	16.0/15.4	9.5/7.7	15.2/16.0	5.7/5.9	11.5/11.5
	TS2	14.3/11.7	11.4/11.6	14.6/15.0	18.1/18.1	12.3/10.6	16.7/17.0	7.1/7.8	11.7/11.9
	Cpd I	−0.2/−3.7	−0.4/−3.8	−0.2/−2.6	7.1/−0.6	−2.0/−2.3	2.5/−3.0	−2.6/−3.4	2.0/−5.0
III	TS	26.6/27.0	22.3/23.1	24.5/26.5	19.4/19.5	27.6/29.1	11.7/11.9	28.8/28.2	11.4/12.0
	Fe RS	7.2/6.4	11.8/8.0	7.5/6.4	13.4/7.2	5.6/−0.1	4.3/2.4	4.4/−1.0	4.8/1.7

[a] For notation see text; for models A−H, see Figure 1; +W denotes the presence of an extra water molecule. All energies refer to geometries that are fully optimized using the corresponding basis set (B1/B2).

4.00 Å in FeRS, and the O1−O2 distance is 1.51 Å, as expected for hydrogen peroxide.

In the native P450cam enzyme, the coupling reaction is thus computed to be much more facile than uncoupling, and there should be no release of hydrogen peroxide. This agrees with experiment where 5-exohydroxycamphor is found as the only product.[18]

3.2. T252S Mutant. In this mutant, there is only one water molecule (Wat901) in the distal pocket of the enzyme since a manually added water molecule (WatS) quickly leaves the pocket during MD simulation (see above and Figure S1 in Supporting Information). However, like threonine in the wild-type enzyme, serine has a hydroxyl group and can thus form a proton transfer channel from Asp251 to the FeOOH moiety through Wat901 and its hydroxyl group.

Coupling Reaction: Figure S4 in the Supporting Information presents the QM/MM optimized structures for the coupling reaction of the serine mutant. The barrier for O−O bond cleavage is 15.6 kcal/mol, and the energy of the first intermediate (IC1) is 8.4 kcal/mol. The transition state for the subsequent proton transfer from Asp251 via the water molecule (Wat901) to serine lies at 11.6 kcal/mol. The overall reaction is exothermic by −3.8 kcal/mol. After proton release, the side chain of Asp251 rotates back to form a salt bridge with Arg186 as shown in Figure S4 in Supporting Information. The spin density and charge of the OH group in the first intermediate (IC1) are −0.6 and −0.2, indicating that OH will not behave as a "perfect" radical in IC1 due to the strong hydrogen-bonding interactions with Ser252 and the FeO unit (as in the case of the wild-type enzyme[33]). Overall, the coupling reaction is very similar in the T252S mutant and the wild-type enzyme, both electronically and mechanistically. The rate-limiting barrier is slightly higher in the T252S mutant (15.6 vs 14.3 kcal/mol in the wild-type enzyme).

Uncoupling Reaction: The optimized structures are shown in Figure S5 in Supporting Information. Like in the wild-type enzyme, the reaction proceeds in one step, with an energy barrier of 23.1 kcal/mol, and is endothermic by 8.0 kcal/mol. In the transition state and the product, Ser252 and Asp251 have zero spin density, consistent with a proton-assisted heterolytic O−Fe bond cleavage. The O−O distance stays at around 1.5 Å during the reaction, whereas the Fe−O1 distance increases from 1.87 Å in Cpd 0 to 2.24 Å in TS1 and to 3.71 Å in the product, thus reflecting the formation of hydrogen peroxide and the cleavage of the O−Fe bond.

A comparison of the computed rate-limiting barriers for the T252S mutant shows that the coupling reaction is favored over the uncoupling reaction by 7.5 kcal/mol, less so than in the wild-type enzyme where the difference in the computed barriers is 12.8 kcal/mol. Experimentally, the serine mutant retains a high ratio in the coupling of oxygen consumption to *d*-camphor hydroxylation with some H_2O_2 formation being observed (15% of oxygen consumption relates to hydrogen peroxide and 85% to *d*-camphor hydroxylation).[7,19] The QM/MM calculations thus give the correct qualitative trend that uncoupling becomes more facile in the T252S mutant compared with the wild-type enzyme, but they are not accurate enough to provide quantitative predictions.

3.3. T252V Mutant. As noted before, the T252V mutant has enough space to accommodate an extra water molecule (WatS) in the distal pocket of the enzyme (see above and Figure S2 in Supporting Information). We have therefore studied its reactions without and with WatS (see models C and D in Figure 1). Figures S6−S9 in the Supporting Information show the QM/MM optimized structures (QM region only).

Coupling Reaction without WatS: Similar to the wild-type enzyme and the T252S mutant, the coupling reaction starts with O−O cleavage followed by proton transfer from Asp251 to the formed OH species either directly through Wat901 or through the methyl group of valine. The latter path through the methyl group is unfavorable (transition state at 26 kcal/mol) and will thus not be discussed. On the preferred path through Wat901, the transition states for the two steps lie at 17.1 and 15.0 kcal/mol, respectively, and the intermediate (IC1) is 11.0 kcal/mol above Cpd 0. Judging from the computed spin density (−0.97) IC1 contains an almost "pure" OH radical. After the proton transfer, Asp251 moves back to form a hydrogen bond with Arg186; in this product conformation, Asp251 has zero spin density and a Mulliken group charge of almost −0.5 e.

Coupling Reaction with WatS: The introduction of WatS extends the hydrogen-bonding network in the Asp251 channel. It does not affect the computed energy profile much. The two transition states are raised in energy slightly, to 18.9 and 18.1 kcal/mol, respectively. The intermediate again contains an OH radical (spin density of −0.98), and the formation of Cpd I is again roughly thermoneutral (−0.6 kcal/mol). The presence of an extra water molecule (WatS) in the protein pocket of the T252V mutant thus does not have a significant influence on the coupling reaction.

Uncoupling Reaction without WatS: As in the case of the T252V mutant, only a concerted process is found, with a barrier of 26.5 kcal/mol and an endothermicity of 6.4 kcal/mol. The O−Fe distance increases from 1.86 Å in Cpd 0 to 2.10 Å in the transition state and 4.04 Å in the product. The H_2O_2 moiety has zero spin density and a total charge of −0.1 in the product, consistent with the formation of hydrogen peroxide.

Uncoupling Reaction with WatS: The incorporation of an extra water molecule into the Asp251 channel lowers the computed barrier appreciably, by 7.0 to 19.5 kcal/mol, while the endothermicity changes only slightly to 7.2 kcal/mol. In the product, the spin density and Mulliken charge on H_2O_2 are 0.0 and −0.4 e, respectively.

Comparing the different mechanistic scenarios for the T252V mutant, we find similar barriers for coupling and uncoupling in the presence of WatS (18.9 vs 19.5 kcal/mol) and a slightly lower rate-limiting barrier for coupling in the absence of WatS (17.1 kcal/mol). Considering the limited accuracy of our calculations, we can only conclude qualitatively that these reactions should be competitive and that one should expect both *d*-camphor hydroxylation and H_2O_2 formation in the T252V mutant.

3.4. T252A Mutant. As in the case of the T252V mutant, we have studied the coupling and uncoupling reactions both in the absence and presence of an additional water molecule (WatS). The QM/MM optimized structures are shown in Figure 4 and Figures S10−S12 in the Supporting Information (QM regions only).

Coupling Reaction without WatS: The transition states for the two steps are located at 17.1 and 10.6 kcal/mol above Cpd 0. The rate-limiting barrier for the initial O−O cleavage is 2.8 kcal/mol higher than in the wild-type enzyme. The O−O distance increases from 1.53 Å in Cpd 0 to 2.60 Å in the intermediate (IC1), reflecting cleavage of the O−O bond. The O−Fe distance is 1.66 Å in the product, confirming the formation of Cpd I.

Coupling Reaction with WatS: Inclusion of an extra water molecule (i.e., going from model E to model F in Figure 1) has a negligible effect on the rate-limiting barrier for the initial O−O cleavage (17.3 kcal/mol) but raises the energies of the intermediate and the transition state for proton transfer appreciably. The overall exothermicity is also not affected much (−3.0 kcal/mol). After proton release, the side chain of Asp251 rotates back to form a salt bridge with Arg186 in the product.

Uncoupling Reaction without WatS: The proton transfer and the O−Fe bond cleavage occur in a concerted manner, with an activation barrier of 29.1 kcal/mol, which is about 12 kcal/mol higher than that for the coupling reaction in this mutant.

Uncoupling Reaction with WatS: Upon extension of the hydrogen-bonding network by WatS, the reaction remains concerted, but its barrier is reduced dramatically to 11.9 kcal/mol. The reaction energy does not change much (endothermic by 2.4 kcal/mol).

Comparing the computed barriers for the T252A mutant, we conclude that the uncoupling reaction is favored over the coupling reaction by about 5 kcal/mol, provided that an additional water molecule can enter the Asp251 channel close to the active site. If this is the case, the QM/MM results imply that formation of hydrogen peroxide should be dominant. This is consistent with the available experimental evidence.[12,21] Since Cpd I is not formed under these conditions, the experimentally observed epoxidation in the T252A mutant must be due to another oxidant, with Cpd 0 being an obvious candidate.[14] DFT calculations for gas-phase model systems indicate that Cpd 0 can indeed act as an oxidant for epoxidation, even though it is far less reactive than Cpd I.[17] It would seem worthwhile to address the possible competition between epoxidation and uncoupling in the T252A mutant in future QM/MM work.

3.5. T252G Mutant. Since the T252G mutant also provides sufficient space for an extra water molecule (WatS) close to

the active site, we studied the coupling and uncoupling reactions again in the absence and in the presence of WatS. Figures S13−S16 in the Supporting Information show the optimized QM/MM structures (QM regions only). Glycine is part of the MM region since it has no side chain that could be included in the QM region.

Coupling Reaction without WatS: The rate-limiting barrier for the initial O−O cleavage is 17.9 kcal/mol, that is, of similar magnitude as in the T252V and T252A mutants, but higher than in the wild-type enzyme. The intermediate (IC1) and the transition state for proton transfer from Asp251 to the OH radical lie 5.9 and 7.8 kcal/mol above Cpd 0, respectively. The overall reaction is exothermic by −3.4 kcal/mol.

Coupling Reaction with WatS: As in the other mutants, the presence of WatS molecule does not have much effect on the coupling reaction. Its inclusion raises the barrier for the initial O−O cleavage slightly to 18.9 kcal/mol and also increases the relative energies of the intermediate (11.5 kcal/mol) and the second transition (11.9 kcal/mol). The reaction remains exothermic (−5.0 kcal/mol).

Uncoupling Reaction without WatS: As in the case of the T252A mutant, the proton transfer from Asp251 via Wat901 to O1 and the cleavage of the Fe−O1 bond occur in a concerted reaction. The activation energy for the formation of H_2O_2 is 28.2 kcal/mol and thus about as high as that of the T252A mutant.

Uncoupling Reaction with WatS: The inclusion of WatS allows the formation of a more extended hydrogen-bonding network that provides a much better path for proton transfer between the proton source (Asp251) and the proton acceptor (O1). The barrier is thus reduced drastically to 12.0 kcal/mol, without affecting the reaction energy much (endothermic by 1.7 kcal/mol).

At first sight, it seems surprising that the uncoupling reaction with WatS has a significantly lower barrier in the T252A and T252G mutants than in the T252V mutant. Closer inspection of the QM/MM optimized structures of Cpd 0 offers an explanation: In the alanine and glycine mutants, WatS is hydrogen-bound to both oxygen atoms of the FeO_2 unit with distances of 1.5 to 1.7 Å. In the T252V mutant, however, the steric demands of the valine residue change the hydrogen-bonding network such that WatS is hydrogen-bound to the distal oxygen only so that the proton transfer to the proximal oxygen is hindered and requires more activation. As a consequence, a clear preference for uncoupling results only for the T252A and T252G mutants.

IV. Summary and Conclusions

Four mutants (Thr252Ala, Thr252Val, Thr252Gly, Thr252Ser) of P450cam have been targeted to investigate mutation effects on the formation of Cpd I. We have studied two competing reactions at the QM/MM level, which both originate from Cpd 0 and require proton transfer toward Cpd 0: the conversion of Cpd 0 to Cpd I and water (coupling reaction), and the regeneration of the ferric resting state with concomitant formation of hydrogen peroxide (uncoupling reaction). For each reaction, we have considered two mechanisms: either first bond cleavage in the FeOOH moiety followed by proton transfer in the Asp251 channel, or both steps in reverse order. Only the more facile of these two pathways has been discussed in the text for each reaction (mechanisms I and III), while the other results are only presented as Supporting Information.

The stability of an additional water molecule in the distal pocket has been tested for all four mutants by classical MD simulations. This extra water molecule quickly escapes from the distal pocket of the T252S mutant (as in the case of the native enzyme) but remains stable during 2 ns simulations of the T252A, T252G, and T252V mutants. Hence, for these three mutants, the reactions have been studied both without and with an additional water molecule.

In the coupling reaction, the initial and rate-limiting step is a homolytic O−O bond cleavage followed by a concomitant proton and electron transfer that yields Cpd I and water. In the wild-type enzyme, the O−O bond cleavage has a barrier of 14 kcal/mol that increases slightly to 16 kcal/mol in the T252S mutant and 17−18 kcal/mol in the T252A, T252V, and T252G mutants. The presence of the extra water molecule (WatS) in the T252A, T252V, and T252G mutants has only marginal effects on this barrier. The transition state for the initial O−O cleavage is rate-limiting in all five enzymes considered. Since the mutations as well as the additional water molecule mainly affect the subsequent proton transfer, it is reasonable that neither of these modifications causes any significant effect on the rate-limiting step of the coupling reaction.

The uncoupling reaction involves a proton transfer from Asp251 via a hydrogen-bonding network to the distal oxygen atom (O1) accompanied by O−Fe bond cleavage that leads to the ferric resting state and hydrogen peroxide. An alternative two-step mechanism with initial O−Fe bond cleavage is less favorable. Without the extra water molecule, the energy barrier of the concerted mechanism is 25−28 kcal/mol in the wild-type enzyme and the valine, alanine, and glycine mutants. The presence of an extra water molecule (WatS) drastically reduces this barrier to 11 kcal/mol for the T252A and T252G mutants and to 19 kcal/mol for the T252V mutant.

According to the experimental data, the coupling reaction is more favorable in the wild-type enzyme and the serine mutant, with 100 and 85% coupling (formation of Cpd I), respectively. In the serine mutant, uncoupling (formation of hydrogen peroxide) accounts for 15% of the O_2 consumption. The uncoupling reaction dominates in the T252A and T252G mutants, where only 3−5% of the O_2 is consumed to form Cpd I. In the T252V mutant, both reactions occur with similar probability.

If the effect of an additional water molecule is not taken into account, the present QM/MM calculations would disagree with experimental results; that is, all mutants would behave like the wild-type enzyme and prefer coupling over uncoupling. An additional water molecule in the Asp251 channel has only a minor influence on the formation of Cpd I. However, consistent with experimental observations that direct contact between solvent water and the heme active site is responsible for uncoupling, an extra water molecule has a significant effect on the uncoupling reaction. In case of the T252A and T252G mutants, the barrier of the uncoupling reaction is reduced dramatically such that the coupling reaction becomes disfavored, in line with experiment. For the T252V mutant, both reactions are observed experimentally, which can again only be rationalized if an additional water molecule is taken into account. The competition between the coupling and uncoupling reactions and the effects of the extra water molecule are illustrated by the energy profiles for the wild-type enzyme (Figure 5) and the T252A mutant (Figure 6). It is conceivable that there are other P450 systems (e.g., with other substrates and/or other mutations) where a single additional water molecule may also play a

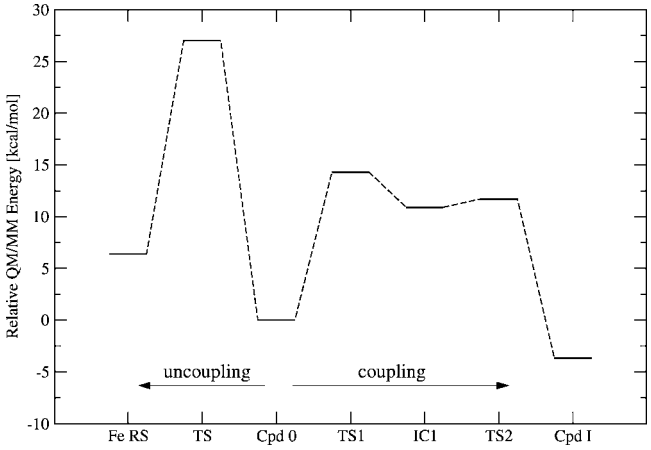


Figure 5. Energy profile (UB3LYP/B2/CHARMM) of the coupling and uncoupling reaction in the wild-type enzyme.

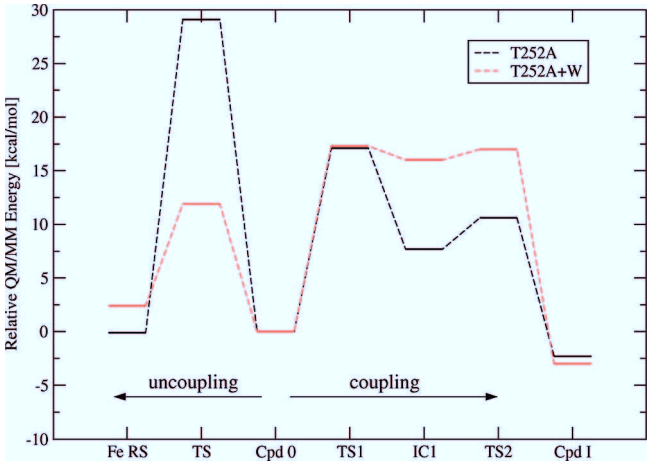


Figure 6. Energy profiles (UB3LYP/B2/CHARMM) of the coupling and uncoupling reaction with and without an extra water in the T252A mutant.

decisive mechanistic role, and it is therefore generally advisable in theoretical work to check this possibility by explicit simulation.

The present QM/MM results agree with the available experimental evidence and thus offer an atomistic explanation for the role of the Thr252 residue in the reactions of Cpd 0. Thr252 is a highly specific proton donor to the distal oxygen atom of Cpd 0, and the preference for the coupling reaction in the wild-type enzyme is based on this specificity. If Thr252 is replaced by small aliphatic residues, an additional water molecule can enter the active site and establish proton transfer channels to both oxygen atoms of the FeOOH moiety. Optimum hydrogen-bonding networks can be formed in the T252A and T252G mutants, which lower the barriers to proton transfer substantially and lead to a preference for uncoupling. This effect is less pronounced in the T252V mutant with the sterically more demanding valine residue, and hence both reactions can occur in this case. Replacing Thr252 with a structurally similar serine residue causes only relatively minor changes since the required proton transfer can make use of the OH group that is present in both residues, without the need to involve an extra water molecule. The difference between the computed rate-limiting barriers for coupling and uncoupling is smaller in the T252S mutant than in the wild-type enzyme, which is qualitatively consistent with the slight decrease in the observed specificity.

Acknowledgment. This work was supported by the Max Planck Society.

Supporting Information Available: MD results for the T252S, T252V, and T252G mutants. Energies, spin densities, and group charges for mechanisms I, III, and IV. Geometric parameters for mechanisms I and III. Optimized geometries of the stationary points of mechanisms I, III, and IV. Complete references 38 and 44. This material is available free of charge via the Internet at http://pubs.acs.org.

JA808744K

QM/MM Study of the Second Proton Transfer in the Catalytic Cycle of the D251N Mutant of Cytochrome P450cam

Muhannad Altarsha, Dongqi Wang,† Tobias Benighaus, Devesh Kumar, and Walter Thiel*

Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, D-45470 Mülheim an der Ruhr, Germany

Received: November 7, 2008; Revised Manuscript Received: May 20, 2009

Protonation of Compound 0 in the catalytic cycle of cytochrome P450cam may lead to the formation of either the reactive Compound I (coupling) or the ferric resting state (uncoupling). In this work, we investigate the effect of the D251N mutation on the coupling and uncoupling reaction by combined quantum mechanics/molecular mechanics (QM/MM) calculations. The mutated Asn251 residue has two possible orientations, i.e. directed toward the active site (no flip) or away from the active site (flip), with the latter one being preferred in classical molecular dynamics (MD) simulations. The possible proton transfer mechanisms in the coupling and uncoupling reaction were studied for three models of the D251N mutant, i.e. no flip (model I), flip (model II), and flip with an extra water (model III). According to the QM/MM calculations, the uncoupling reaction is always less favorable than the coupling reaction. The coupling reaction in the D251N mutant follows the same mechanism as in the wild-type enzyme, with initial O−O cleavage followed by proton transfer. The barrier for the initial step is similar in all D251N models, but the proton transfer is most facile in model III. The hydroxide anion formed in model III is not reprotonated easily by neighboring residues, while proton delivery from bulk solvent seems possible via a water network that remains intact during 2 ns classical MD simulation. The computational results are consistent with the experimental findings that the coupling reaction dominates the consumption of dioxygen in the D251N mutant, but with lower activity than in the wild-type enzyme.

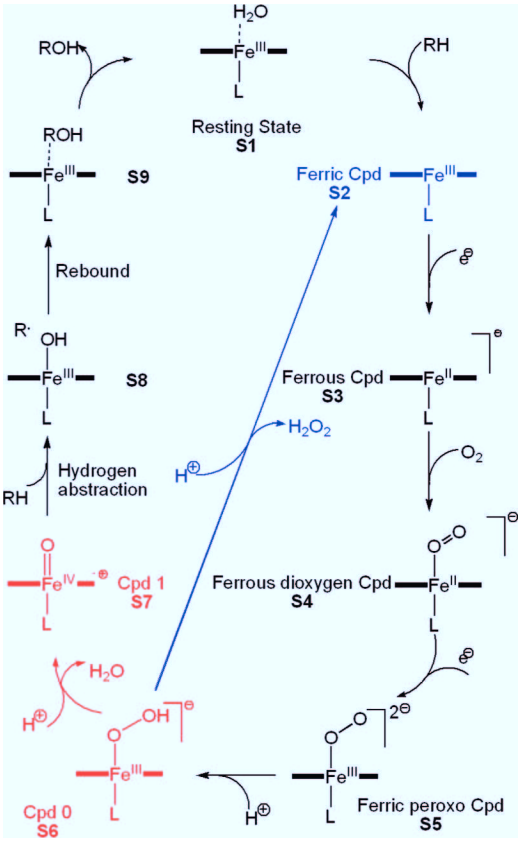
I. Introduction

Cytochromes P450 (P450s),[1] a ubiquitous family of heme containing monooxygenases, utilize dioxygen to insert an oxygen atom into inert hydrocarbon substrates. They play an important role in the biosynthesis of steroids, drug metabolism, and detoxification of xenobiotics.[2] Many studies of P450 have focused on the bacterial P450cam[3] with a camphor substrate, the first soluble P450 protein whose sequence and X-ray structure were determined.[4]

The catalytic cycle of P450cam is shown in Scheme 1.[3] The essential steps up to the formation of the active species involve (1) binding of the substrate, (2) reduction of the ferric cytochrome P450 to the ferrous state, (3) binding of molecular oxygen leading to the ferrous dioxygen complex, (4) second electron transfer and formation of the peroxo-iron(III) complex, (5) protonation of the distal oxygen, which leads to the formation of the ferric hydroperoxo complex (Compound 0, Cpd 0), and (6) second protonation of the distal oxygen with O−O bond cleavage, which generates the putative oxoferryl species (Compound I, Cpd I). Alternatively, protonation of the proximal oxygen in Cpd 0 leads to the uncoupling reaction that yields the ferric resting state and hydrogen peroxide instead of the hydroxylated product. Hence, a well-targeted proton transfer is indispensable for cleavage of the iron-bound dioxygen and formation of Cpd I.

Two proton delivery pathways have been proposed for P450cam that involve the highly conserved residues Asp251[5−10] and Glu366.[3,11,12] The crystallographic structure published by Schlichting et al.[3] indicated that there could be a Glu366 channel composed of Glu366, the bridging water molecules 687, 566,

SCHEME 1: Cytochrome P450cam Catalytic Cycle



* To whom correspondence should be addressed. E-mail: thiel@mpi-muelheim.mpg.de.

† Current address: Laboratory of Physical Chemistry, Swiss Federal Institute of Technology Zürich, ETH, CH-8093 Zürich, Switzerland.

10.1021/jp809838k CCC: $40.75
Published on Web 06/19/2009

523, and 902 (numbering as in PDB structure 1DZ8[3]), and the hydroxyl group of Thr252. The hydrogen bond network between the carboxyl group of Glu366 and the distal oxygen atom remains stable throughout molecular dynamics (MD) simulation.[13] However, this chain terminates at Glu366 without any connection to the protein surface,[10] and mutations of Glu366 show little influence on catalytic activity. These findings suggest that Glu366 does not play a major role in catalysis.[9,10,14]

Closer inspection of the Asp251 channel renders this a better candidate. First, the crystallographic structure indicates that the Asp251 residue may serve as a proton shuttle between the solvent accessible Lys178/Asp182/Arg186 triad and Thr252.[6,10,15,16] This is supported by MD simulations that confirm the flexibility of the Asp251 residue.[17−19] Second, extensive site-directed mutation studies clearly indicate that Asp251 and Thr252 play a vital role in the catalytic cycle.[6,10,20,21] Therefore, it has been suggested that the Asp251 and Thr252 residues are important in constituting a controlled proton delivery pathway that involves solvent water and provides an active-site H-bond donor. This may be a trapped water molecule or Thr252.[6,10]

An alternative proposal[22] assumes that the proton transfer from the solvent into the active site of P450 may proceed through the hydration cluster close to the heme propionates in the resting state. However, after the entry of the substrate, this pathway is blocked and will therefore not be considered.[3]

The D251N mutation causes a structural change in the vicinity of the active site. The new Asn251 amide side chain no longer favors the hydrogen bonds with Thr185 and Lys178 that keep Asp251 in an orientation toward the active site. Instead, Asn251 forms a new hydrogen bond with Asp182 and rotates away from the active site. This makes the active site more accessible for the solvent, such that an alternative solvent-based proton delivery channel may be established, as indicated by solvent kinetic isotope effect measurements.[4,10] Overall, the D251N mutant exhibits a greatly diminished rate of O_2 consumption for the coupling reaction. The product formation rate decreases by a factor of more than 30 from 820 nmol/min/nmol heme in the wild-type enzyme to 26 nmol/min/nmol heme in the D251N mutant; this has been interpreted in terms of a slower proton transfer to the iron-linked dioxygen.[10]

Mutation of the Thr252 residue by amino acids without hydrogen-bonding side chains virtually suppresses camphor hydroxylation in favor of the uncoupled reduction of O_2 to H_2O_2.[5,8,10,23] The crystal structure of the pentacoordinated ferric complex of the Thr252Ala mutant indicates that the solvent may be responsible for the observed uncoupling of the enzyme turnover from camphor hydroxylation in this mutant, since it contains a water molecule near the O_2 binding site, which is not present in the wild-type enzyme.[11,23]

Although the formation of Cpd I from Cpd 0 has been the subject of numerous theoretical studies, their results were partly inconsistent and mechanistic details such as the stability of a protonated Cpd 0 intermediate (prot-Cpd 0) depended strongly on the chosen model system.[11,17,24−28] Therefore, Zheng et al.[19] studied the formation of Cpd I in the full P450cam enzyme using a hybrid quantum mechanical/molecular mechanical (QM/MM) approach. They found that protonated Cpd 0 is an intermediate only in the Glu366 channel. It is very high in energy (more than 20 kcal/mol above Cpd 0) and the barrier for its decay is only 3−4 kcal/mol toward either Cpd 0 or Cpd I. In the Asp251 channel, protonated Cpd 0 was found to be unstable. Therefore, a novel mixed homolytic−heterolytic mechanism was proposed as the most favored pathway in both channels,[19] with the rate-limiting step being the initial O−O bond cleavage with a barrier of about 13−14 kcal/mol. The same methodology was used to study the effect of mutations at the Thr252 position to explain the important role of this residue in the proton delivery pathway and to rationalize the preference for coupling or uncoupling in the Thr252 mutants.[29]

Wang et al.[30] compared the role of the Asp251 and Asn251 residues during formation of Cpd 0 in the wild-type enzyme and the D251N mutant, respectively. In the D251N mutant, the Asn251 side chain was observed to be flexible during MD simulations and to flip from an orientation toward the active site to an orientation away from the active site, which generates some empty space between Arg186 and Wat901 that may be filled by an extra water molecule. QM/MM calculations on the resulting no flip, flip, and flip with extra water models led to the conclusion that the proton transfer in the Asn251 channel requires either a back-flip of the Asn251 side chain or the participation of an extra water molecule in the active site, and even then the barriers are higher than those in wild-type P450cam.

Until now, the effect of the D251N mutation on the coupling and uncoupling reactions of Cpd 0 has not been studied by QM/MM methods that account for the full enzyme. This article reports QM/MM calculations on both reaction pathways for the D251N mutant and the wild-type enzyme to elucidate the mechanistic role of the Asp251 residue and solvent molecules. In section II we briefly describe the computational methods employed. In section III we present and discuss the results of classical MD simulations and of QM/MM calculations for the different D251N models and compare them to results for the wild-type enzyme and to experimental data. Finally, section IV offers conclusions.

II. Computational Methodology and Proposed Mechanisms

QM/MM Setup. The available experimental X-ray structure of cytochrome P450cam (PDB code, 1DZ8;[3] resolution, 1.9 Å) was used as a starting point in our work. The same protonation and solvation protocol was employed as in previous studies.[31−33] Glu366 remains deprotonated, since we do not consider the Glu366 channel. Asn251 is neutral in the D251N mutant while Asp251 was protonated in our related previous QM/MM work on the wild-type enzyme.[19,30]

Both the wild-type and mutant system contain around 25000 atoms including 5895 TIP3P water molecules.[34] The initially prepared systems were relaxed by energy minimizations and MD simulations using the CHARMM22 force field[35] as implemented in the CHARMM program.[36] During the MD simulation, the heme unit, Cys357, and the outer 8 Å of the solvent layer were kept fixed.

The chosen QM/MM methodology is analogous to that used in previous studies.[19,30−33] Therefore, only those aspects relevant to the present work are mentioned here. Minimized snapshots from classical MD simulation trajectories were taken as the initial structures for QM/MM calculations. The QM region was described by unrestricted hybrid DFT (UB3LYP)[37] using the LACVP[38] small-core ECP basis set on Fe and 6-31G[39] on the rest of all atoms (B1) for geometry optimizations. Single-point calculations were carried out with the TZVP[40,41] basis set (B2) applied to all atoms. The CHARMM force field was run through the DL_POLY[42] code to treat the MM part of the system. All QM/MM calculations were performed with the ChemShell package[43] that integrates the TURBOMOLE[44] and DL_POLY programs and also provides the HDLC optimizer[45] for geometry optimizations. An electronic embedding scheme[46] was adopted

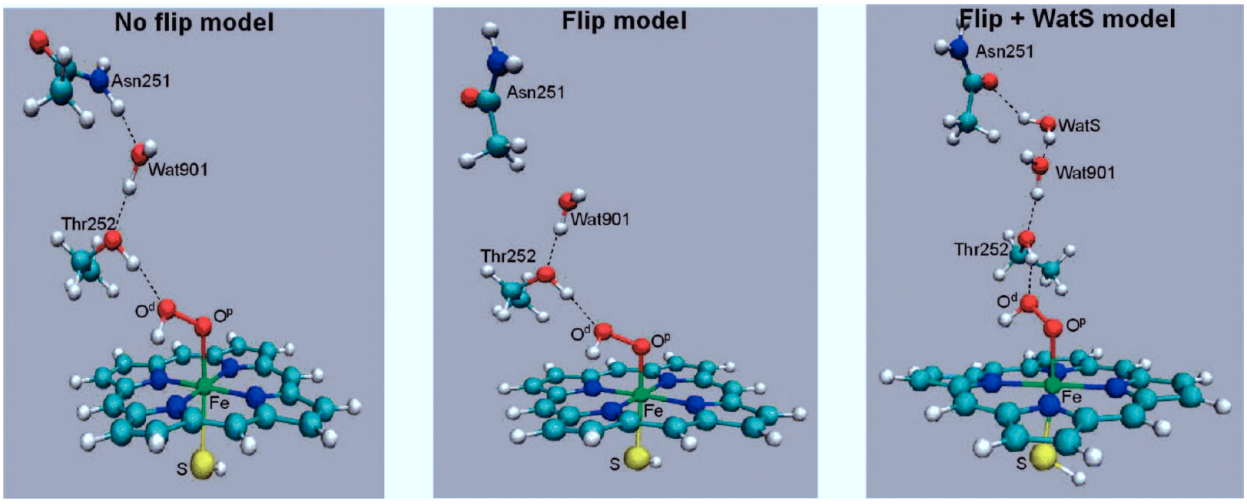
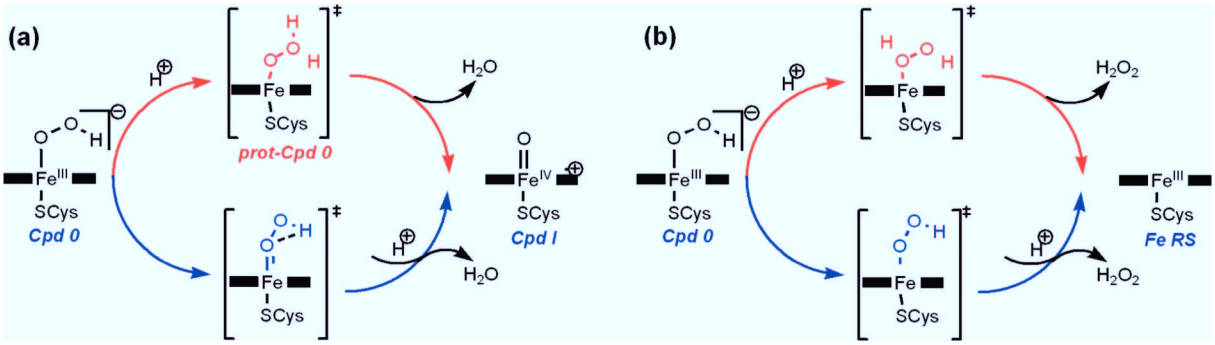


Figure 1. QM models for the Asn251 mutant: Model I represents the no flip configuration of the Asn251 side chain, model II represents the flip configuration, and model III extends model II with an extra water molecule (WatS).

SCHEME 2: (a) Two Proposed Reaction Mechanisms for Cpd I Formation (Coupling Reaction), and (b) Two Proposed Reaction Mechanisms for Ferric Resting State Formation (Uncoupling Reaction)



in the QM/MM calculations; that is, the MM charges were included in the one-electron Hamiltonian of the QM part, and QM/MM electrostatic interactions were evaluated as interaction of the QM electrostatic potential with MM partial charges. Hydrogen link atoms in combination with a charge shift model[47] were employed to treat the QM/MM boundary. All minima (reactants, products, and intermediates) and transition states (TS) reported in this paper were fully optimized.

The active region was defined to include all residues and water molecules within 6 Å of any non-hydrogen atom of the core region which contains the heme unit, Cys357, Glu366, Asn251, Thr252, Wat523, Wat566, Wat687, Wat901, and WatS. This results in ca. 1400 atoms to be optimized.

QM Region. We used an analogous QM region as in previous work on the wild-type enzyme.[19] In the D251N mutant, the Asp251 residue was manually replaced by Asn, and the QM region thus includes the following: porphyrin-FeOOH without side chains, SH ligand, Asn251 (CH_2CONH_2), Thr252 (CH_3CHOH), and Wat901. This QM region is shown in Figure 1 for models I and II, which represent the no flip and flip conformations encountered during 2 ns classical MD simulations (see below). In model III an extra water molecule (WatS) was manually placed in the empty space between Arg186 and Wat901. The stability of WatS in the active site was confirmed by means of classical MD simulation, and therefore, model II was extended to model III by including WatS. The three models are illustrated in Figure 1.

Possible Proton Transfer Pathways. Scheme 2 presents possible pathways that yield two different products through proton transfer, respectively. In the first two pathways (coupling reactions, Scheme 2a), the proton is transferred from the hydroxy group of threonine to the distal oxygen (O^d), while in the other two pathways (uncoupling, Scheme 2b), the proton is transferred to the proximal oxygen (O^p). In both cases, Wat901 bridges Asn251 and Thr252 (and WatS in model III) to construct the proton transfer channel. The details of these mechanisms are as follows.

In the coupling reaction (Scheme 2a), which leads to the formation of Cpd I and one water molecule, the proton is transferred to the proximal oxygen via two possible pathways. The first pathway represents the conventional mechanism (in red, Scheme 2a), in which the coupling reaction is initiated by proton transfer to the distal oxygen atom of the O_2H moiety and protonated Cpd 0 is formed, followed by a heterolytic O−O bond cleavage that yields Cpd I and water.[19] In the second mechanism, an initial O−O bond cleavage generates an OH species hydrogen-bonded to the FeO moiety. A subsequent proton transfer to this OH species with a concomitant electron transfer from the heme yields Cpd I and water[19] (blue pathway in Scheme 2a).

The uncoupling reaction (Scheme 2b) leads to the formation of the ferric resting state and hydrogen peroxide. Two possible mechanisms have been investigated also for this reaction. The first one (red pathway in Scheme 2b) starts with a proton transfer to the proximal oxygen to form Fe−O_2H_2, followed by heterolytic Fe−O bond cleavage that generates the ferric resting state and hydrogen peroxide. In the second mechanism, the

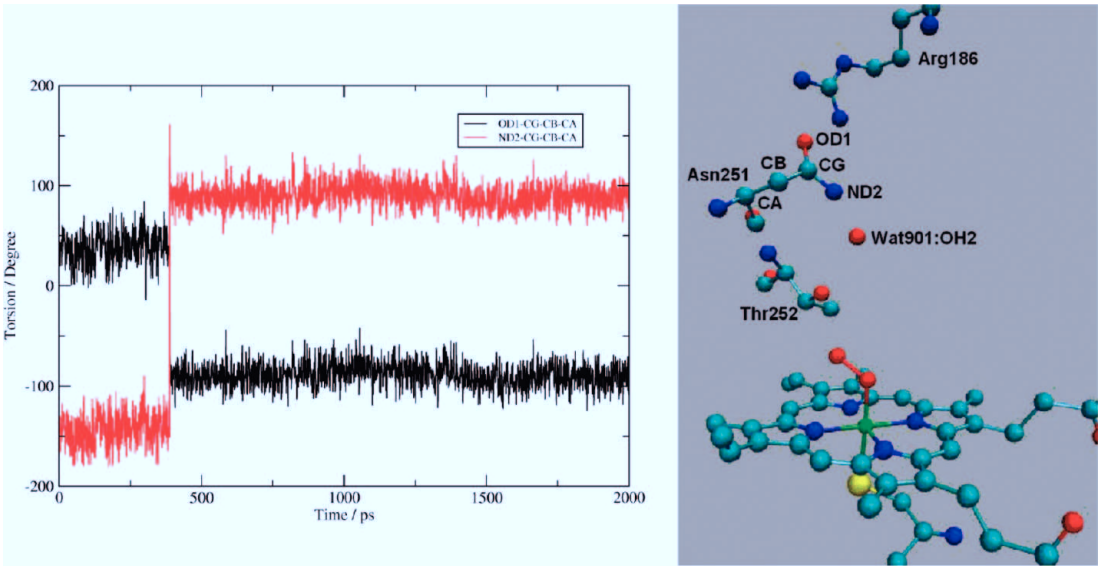


Figure 2. Motion of the Asn251 side chain during the MD simulation. Left: results for torsion angles OD1−CG−CB−CA and ND2−CG−CB−CA. Right: definition of atom labels.

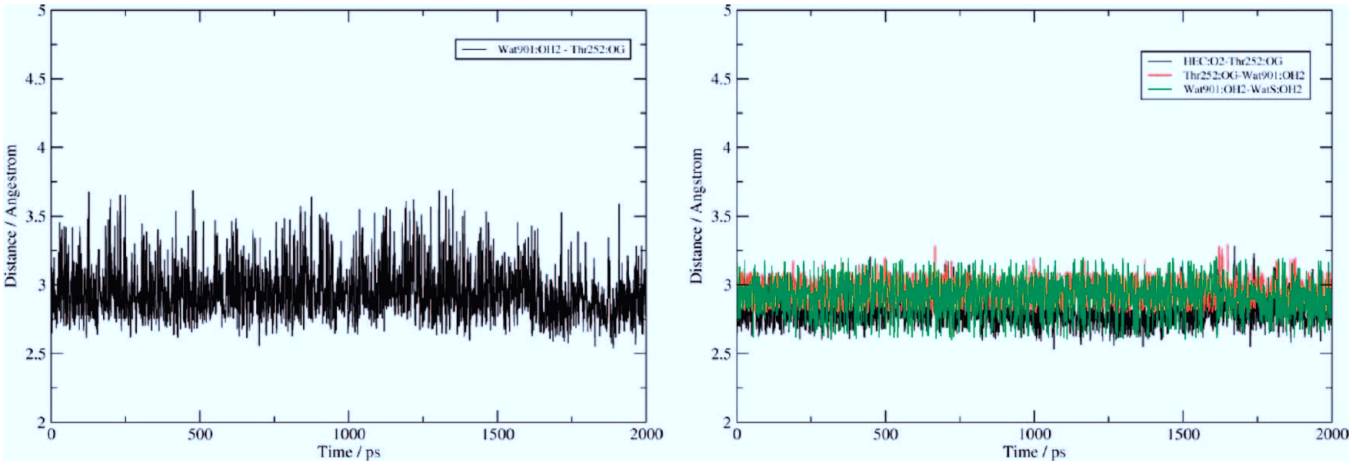


Figure 3. Monitoring the mobility of the crystallographic water molecule (Wat901) and the extra water molecule (WatS) during the MD simulation (for atom labels, see Figure 2). Left: Wat901 in model II. Right: Wat901 and WatS in model III.

Fe−O bond is cleaved first and a hydroperoxo radical is formed, which is then converted to hydrogen peroxide.

III. Results and Discussion

MD Simulation Results. The crystal structure of the D251N mutant (2A1N)[48] is a dimer which contains the Asn251 residue in two different orientations in the two units (normal and flipped orientation). In the former case, the Asn251 side chain shows a 25° rotation toward Wat901, thus establishing a connection to the active site.[48] This suggests that there is no strong interaction between Asn251 and Arg186, unlike the salt bridge which exists between Asp251 and Arg186 in the crystal structure of the wild-type enzyme (1DZ8). Consequently, a direct interaction between Asn251 and Wat901 is possible, and a proton transfer channel from Asn251 via the only crystallographic water molecule in this region (Wat901) to the active site may be formed.[10,18]

We performed a 2 ns classical MD simulation of the D251N mutant of Cpd 0 starting from model I (no flip) to study the stability of the Asn251 amide group. Its conformation is characterized by two torsion angles with the backbone (OD1−CG−CB−CA and ND2−CG−CB−CA). During the simulation (see Figure 2), the OD1−CG−CB−CA torsion angle decreases from 70° to −90°, whereas the ND2−CG−CB−CA angle increases from −108° to 88°. In the no flip conformation, the Asn251 residue forms hydrogen bonds with nearby residues, i.e. Arg186 and Wat901 (relevant average distances: Arg186:HH22−Asn251:OD1, 1.809 Å; Arg186:HH12−Asn251:OD1, 1.779 Å; Wat901:OH2−Asn251:HD22, 1.656 Å). After the flip of the Asn251 side chain (model II), the amide group of Asn251 retains its hydrogen bond with Arg186 and forms new hydrogen bonds with Thr181 and Asp182 (average distances: Thr181:OG1−Asn251:HD22, 2.049 Å; Asp182:OD1−Asn251:HD21, 1.651 Å).

Wat901 and WatS play a critical role in the proton delivery, and we have therefore checked their mobility by additional MD simulations. Figure 3 shows the results for Cpd 0 in model II (left) and model III (right). The connection between Wat901 and the side chain of Asn251 is interrupted in model II (flip) of the Asn251 mutant. However, since Wat901 keeps its hydrogen bond with the Thr252:OH group, it does not escape from the protein pocket during the 2 ns simulation but stays close to Thr252 (see Figure 3). Likewise, in model III (flip with extra water), Wat901 remains stable next to the Thr252 amino acid: the average value and the standard deviation for the distance between Wat901:OH2 and Thr252:OG1 is 2.955 ± 0.224 Å,

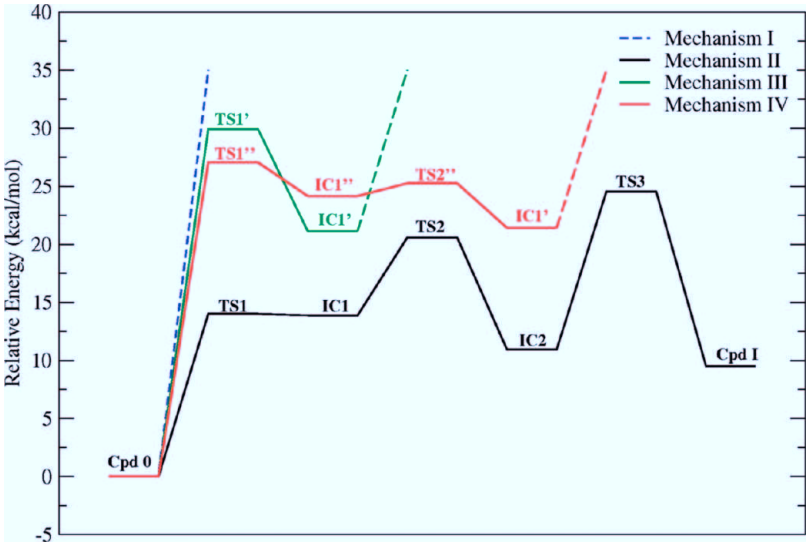


Figure 4. Energy profiles for the four possible mechanisms in the no flip conformation of the D251N mutant (I and II for coupling, III and IV for uncoupling). Relative energies are given in kcal/mol with respect to Cpd 0.

indicating that the hydrogen bond between Thr252 and Wat901 is conserved. The extra water molecule that is inserted in model III (WatS) does not escape from the distal pocket during the MD simulation: the average value and the standard deviation for the distances HEC:O2−Thr252:OG1, Thr252:OG1−Wat901:OH2, and Wat901:OH2−WatS:OH2 are 2.803 ± 0.102 Å, 2.940 ± 0.096 Å, and 2.925 ± 0.127 Å, respectively. These results confirm that both water molecules are stable in the distal pocket of the enzyme.

QM/MM Results. In the following, we present the results from QM/MM geometry optimizations and reaction path calculations for models I−III. Since previous work on the native enzyme and several mutants has established that the doublet state of Cpd 0 is more stable than the quartet state both in gas phase models[7,26] and in the actual enzyme environment,[19] we focus on the doublet spin state in this study. The QM/MM optimized structures of the QM region of the D251N mutant have already been shown in Figure 1 for Cpd 0 in models I−III. The figures in this section will present QM/MM optimized structures of all relevant stationary points (with data for selected geometrical parameters) as well as energy profiles (with relative energies obtained from the B1/B2 basis sets). Since the computed relative energies are not too sensitive to extension of the basis set, we shall discuss the B1 values in the text (B1 is the basis used for geometry optimization). Other computational results (such as spin densities, Mulliken charges, and additional reaction profiles) are documented in the Supporting Information (SI), which also provides an overview picture of the active-site structure (Figure S38).

A. No Flip Model. As already mentioned, four different reaction mechanisms were studied for the coupling and uncoupling reactions of Cpd 0. The coupling mechanisms I and II yield Cpd I and water, while the uncoupling mechanisms III and IV lead to the formation of the ferric resting state and hydrogen peroxide. Figure 4 shows the energy profiles of all mechanisms for the no flip model. In the case of mechanism I (blue line in Figure 4), the energy profile goes uphill only, and we were unable to locate a stable prot-Cpd 0 intermediate. In previous QM/MM studies by our group on the native enzyme and the Thr252X (X = Ser, Ala, Val, Gly) mutants in the Asp251 channel, prot-Cpd 0 was also found to be unstable.[19,29] Mechanism II proceeds in three steps (black line in Figure 4). The first one is O−O bond cleavage with an energy barrier of 14.0 kcal/mol. In the second step, the hydrogen on Thr252 is transferred to the OH moiety with an energy barrier of 8.0 kcal/mol followed by the deprotonation of Asn251 with a barrier of 13.6 kcal/mol relative to IC2.

In contrast to mechanism II, mechanisms III and IV give the ferric resting state and hydrogen peroxide. Mechanism III (green line in Figure 4) starts with a direct proton transfer from the Thr252 residue to the proximal oxygen (O1) of the FeOOH moiety, accompanied by a spontaneous O−Fe bond cleavage. This step has an energy barrier of 29.9 kcal/mol. After this proton transfer, the first intermediate (IC1′), comprising the ferric resting state and hydrogen peroxide with an O-Thr252 anion, is calculated to be 21.1 kcal/mol above the reactant. The recovery of the natural state of Thr252 by abstracting a proton from Asn251 seems to be unrealistic in this model, since deprotonated Asn251 is a significantly stronger base than deprotonated Thr252 in IC1′. Therefore, the proton channel is interrupted and this mechanism does not lead to the desired product.

In mechanism IV (red line in Figure 4), the initial step is O−Fe bond cleavage with a barrier of 27.1 kcal/mol. The resulting intermediate IC1″ contains an OOH moiety and has an energy of ca. 24 kcal/mol. In the next step, a proton is transferred from Thr252 to the proximal oxygen of the OOH moiety with a tiny barrier of only around 1 kcal/mol relative to IC1″. This step leads to an intermediate that is very similar to the one of mechanism III and has an energy of 21.4 kcal/mol. However, in mechanism IV the product of the final proton transfer from Asn251 to the deprotonated Thr252 is again not stable.

In summary, we find that mechanism I is not a realistic pathway, since the formation of its first intermediate (prot-Cpd 0) is difficult. Mechanisms III and IV involve significantly higher activation energies than mechanism II already in the initial stage of the reaction. Qualitatively similar results were also obtained for models II and III. Hence, mechanism II emerges as the most favorable pathway, and we will discuss only this mechanism in detail for all three models. Further information about the other mechanisms is given in the Supporting Information.

Figure 5 presents the QM/MM optimized structures and the energy profile of mechanism II in the no flip model of the D251N mutant. In this conformation, the hydrogen bonding

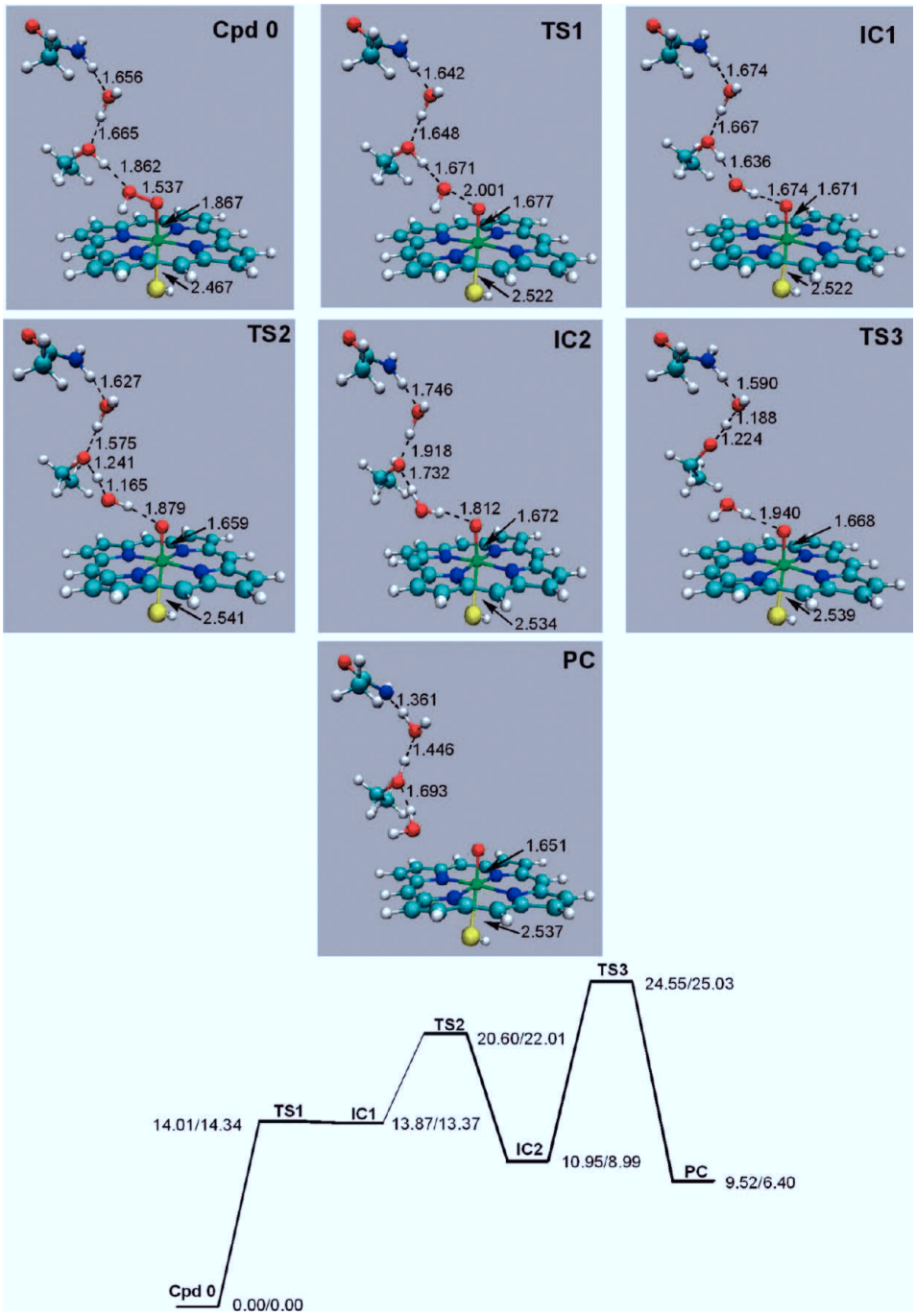


Figure 5. Asn251 mutant, no flip model. (a) Optimized geometries of Cpd 0, TS1, IC1, TS2, IC2, TS3, and PC (UB3LYP/B1/CHARMM). Only the QM region is shown. (b) Energy profile of mechanism II. Energies in kcal/mol relative to Cpd 0 (B1/B2).

network between the proton source (Asn251) and the distal oxygen (O^d) is conserved. The O^d-O^p bond cleavage has a barrier of 14.0 kcal/mol (i.e., very similar to the wild-type enzyme) and leads to the first intermediate (IC1) with an energy of 13.9 kcal/mol. During the first step, the O^d-O^p distance increases from 1.537 (Cpd 0) to 2.001 (TS1) and 2.524 Å (IC1), indicating cleavage of the O−O bond. Meanwhile, the $Fe-O^p$ distance decreases noticeably from 1.867 to 1.677 Å from reactant to first transition state, and the Fe−S distance elongates from 2.467 (Cpd 0) to 2.522 Å (TS1, IC1).

During the reaction, the hydrogen bond between Thr252 and the distal oxygen atom becomes stronger due to the displacement of the OH moiety toward the Thr252 residue. In IC1, a new strong FeO−HO hydrogen bond with a distance of 1.674 Å is formed. The spin density and charge of the OH moiety in IC1 are −0.7 and −0.1, respectively, which indicates that OH will not behave as a "perfect" radical in IC1. This is due to the strong hydrogen-bonding interactions of the OH moiety with both the heme (through FeO) and the protein environment (through Thr252). A similar spin density was reported previously for the wild-type enzyme.[19]

The subsequent proton transfer from Asn251 to the distal oxygen proceeds in two steps. First, a hydrogen atom is transferred from Thr252 to the OH moiety, and then Thr252 is reprotonated via Wat901 by Asn251 with a concomitant electron transfer from the heme. During the first step, the system has to pass over a high energy barrier (TS2) of 20.6 kcal/mol to reach a stable intermediate (IC2) with an energy of 11.0 kcal/mol. In IC2, the spin density of the OH group reduces to almost zero and the hydrogen atom of Thr252 is transferred to the distal

oxygen (0.997 Å) to form a water molecule. The spin density and Mulliken charge on Thr252 are −1.0 and −0.2, respectively. The radical character of Thr252 in IC2 is confirmed by QM/MM calculations with different QM regions, functionals, basis sets, and initial guesses for the density matrix (see Tables S27−S29 of the Supporting Information). The occurrence of such an IC2 radical intermediate in the D251N mutant contrasts with the situation in the wild-type enzyme where the proton transfer from Asp251 is more facile (due to the much higher acidity of Asp251 compared with Asn251) and proceeds in a concerted manner. The transition state of the last step (TS3) in the mutant has an energy of 24.6 kcal/mol above Cpd 0, representing the highest point of the whole pathway. The overall process is endothermic by 9.5 kcal/mol. Reaction path calculations indicate that a concerted proton transfer in the mutant requires an activation of about 39 kcal/mol relative to Cpd 0 (see the Supporting Information).

It is also worthwhile to discuss the nature of IC1 and TS1. In a previous study that employed the same model of the wild-type enzyme, we found a similar reactive high-energy species whose stability depends on the size of the QM region.[19] In view of the energetic similarities of TS1 and IC1 (IC1 is about 0.2 kcal/mol more stable than TS1) and the geometric similarities of IC1 and TS2, it seems probable that TS1 and IC1 are artifacts resulting from the limited size of the QM region and the absence of conformational sampling. Hence, it is more realistic to regard the no flip mechanism in the mutant as a two-step process. The first step then corresponds to O−O bond cleavage, hydrogen transfer, and water formation with an overall barrier of 20.6 kcal/mol. The second step is the deprotonation of the mutated Asn251 residue with a barrier of 13.6 kcal/mol.

The major difference between the wild-type enzyme and the D251N mutant is 2-fold: (1) The barrier of the rate-determining step increases from 14.4 kcal/mol in the wild-type enzyme to 20.6 kcal/mol in the D251N mutant. (2) The formation of Cpd I is a one-step process in the wild-type enzyme, while it becomes a two-step reaction in the D251N mutant with a stable deprotonated Thr252 radical intermediate. Both effects result from the much higher pK_a value of the mutated Asn251 residue compared to Asp251 that acts as the proton source in the wild-type enzyme.

Finally, we tested if an extension of the QM region has a significant effect on the relative energies by including the Arg186 residue into the QM region. We found only very minor changes in relative energies and geometries (see the Supporting Information).

B. Flip Model. In the flipped conformer, rotation of the Asn251 residue causes an interruption of the hydrogen bond network so that it cannot serve as a proton source. Figure 6 shows the QM/MM optimized structures with selected geometric parameters and the energy profile of mechanism II in the flip model. The first step, cleavage of the O−O bond, remains unaffected by the Asn251 flip: the transition state (TS1) and the first intermediate (IC1) have energies of 14.0 and 9.0 kcal/mol relative to Cpd 0, respectively. The spin densities of OH (−1.0) and Fe═O (2.1) in IC1 indicate that the Fe═O moiety carries two unpaired electrons, and the third unpaired electron is mainly located at the OH moiety. The hydrogen bond interactions with the proximal oxygen of the heme and Thr252 that stabilize the OH radical are structurally similar to those of model I but provide a more efficient energetic stabilization of ca. 5 kcal/mol (relative to TS1). Subsequently, the reaction follows a radical mechanism as in the no flip model. The second step is the hydrogen transfer from Thr252 to the OH radical with a transition state (TS2) energy of 17.6 kcal/mol and an intermediate complex (IC2) at 12.4 kcal/mol. The computed spin densities and charges confirm that the proton transfer leads to the formation of a water molecule and the O-Thr radical species (see Tables S6 and S28 of the Supporting Information). Finally, a proton from water (Wat901) and an electron from the heme are transported to the O-Thr252 radical. The corresponding transition state (TS3) and the product complex lie 21.1 and 19.2 kcal/mol above Cpd 0, respectively.

The resulting product complex contains a negatively charged OH species with weak radical character, as indicated by the computed spin density and Mulliken charge (−0.2 and −0.5, respectively). This species is stabilized by strong hydrogen bonds with Thr252 and Asn251 (distances of 1.427 and 1.888 Å, respectively). The reaction pathway is blocked at this stage, since the Asn251 residue does not act as a proton source in the flip model and no other proton source exists in the vicinity. We have also included the Arg186 side chain into the QM region to check if Arg186 may serve as an alternative proton source but found that the distance to the OH anion is too large to enable proton transfer (details see in the Supporting Information). We thus have to conclude that flip conformations of this kind are unlikely to contribute to the formation of Cpd I in the D251N mutant.

C. Flip Model with an Extra Water WatS. As mentioned before, an extra water fits into the space which is released by the flip of the Asn251 side chain (see Figure 1). The introduction of WatS improves the hydrogen bond network with the Arg186 residue that is a potential proton source. Figure 7 presents the QM/MM optimized geometries and the energy profile for the flip model in the presence of the WatS molecule. In this pathway, the O^d-O^p bond cleavage has an energy barrier of 14.0 kcal/mol, similar to cases of the no flip and flip models. The energy of the first intermediate complex (IC1) is 11.8 kcal/mol relative to that of Cpd 0. The spin densities and Mulliken charges of the OH group and the proximal oxygen atom also confirm the cleavage of the O−O bond. The OH group in IC1 forms two strong hydrogen bonds with the heme and the Thr252 residue (distances of 1.447 and 1.690 Å, respectively).

The energy barrier of the second step which involves proton transfer from Wat901 to the OH moiety via Thr252 and concomitant electron transfer from the heme is 0.3 kcal/mol. The whole reaction is endothermic, with the product lying 7.7 kcal/mol above Cpd 0. Both the OH spin density (−0.2) and Mulliken charge (−0.5) indicate that an anionic OH species with only weak radical character is formed, which is stabilized by two hydrogen bonds with Thr252 and WatS (distances of 1.236 and 1.638 Å, respectively). Evidently, the proton transfer is facilitated by the additional water molecule, and the overall endothermicity decreases from 19.2 kcal/mol in model II to 7.7 kcal/mol in model III. Since the barrier for proton transfer is reduced to less than 0.5 kcal/mol, it is appropriate to regard the overall process essentially as a one-step reaction similar to model I.

Starting from the product complex, we tried to move one proton from WatS to the OH anion of Wat901. However, this proton always moved back to the emerging hydroxide anion upon full QM/MM optimization, indicating that WatS is an even weaker Brönsted acid than Wat901 in the given environment (see SI for detailed information). Closer inspection of these structures indicates that WatS also forms a hydrogen bond with the Arg186 residue, thus making Arg186 a potential proton source. To explore the role that Arg186 plays in the presence of WatS, we included the side chain of the Arg186 residue into

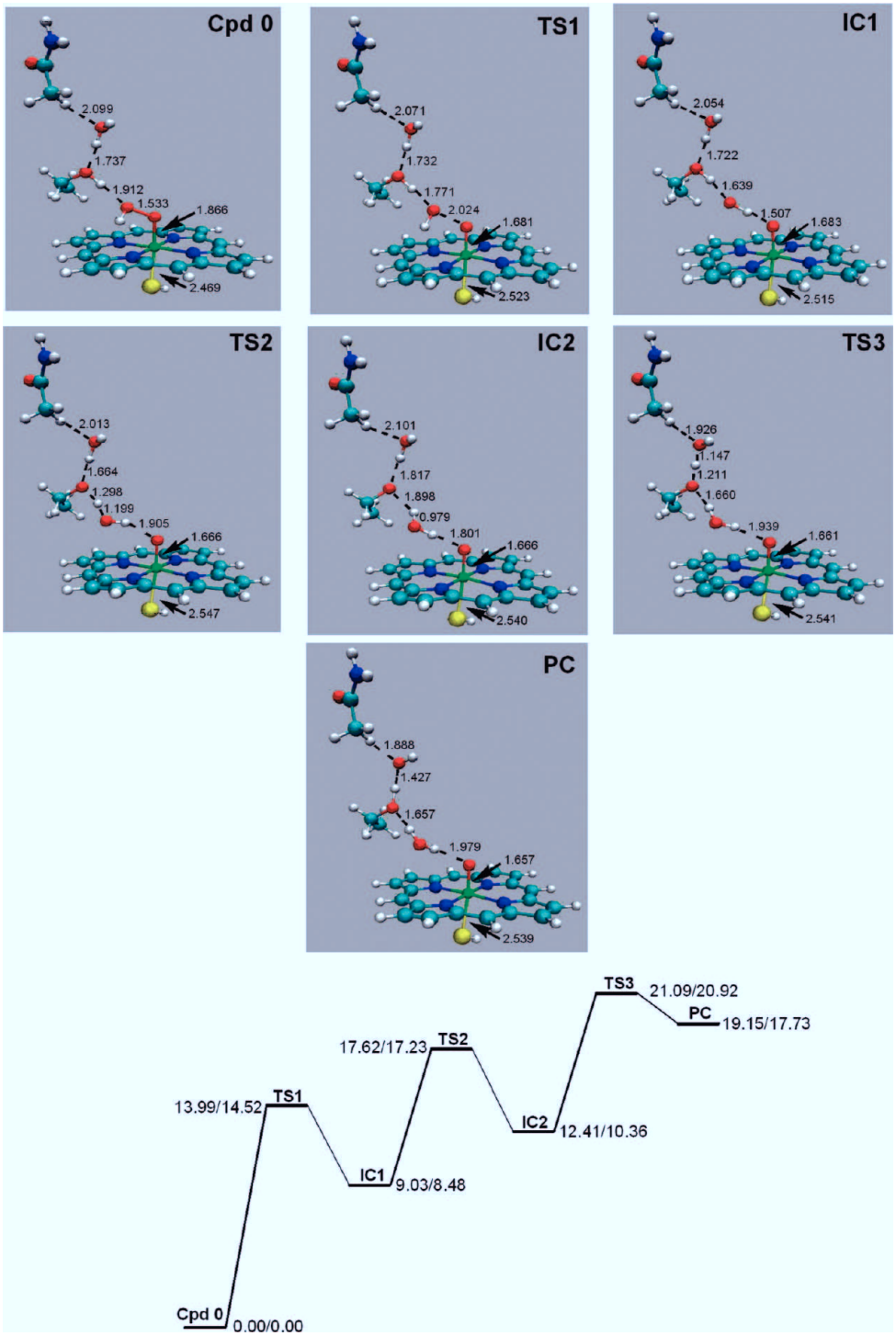


Figure 6. Asn251 mutant, flip model. (a) Optimized geometries of Cpd 0, TS1, IC1, TS2, IC2, TS3, and PC (UB3LYP/B1/CHARMM). Only the QM region is shown. (b) Energy profile of mechanism II. Energies in kcal/mol relative to Cpd 0 (B1/B2).

the QM region and tested if it can act as a proton source. Figure 8 presents the QM/MM optimized geometries of the reactant, first intermediate, and product complex of the flip model with an extra water molecule and Arg186 included in the QM region. The product complex (PC) contains an OH anion (Wat901) that is stabilized by strong hydrogen bonds to Thr252 (1.334 Å) and Wat901 (1.549 Å). We tried to locate the alternative product complex (PC′) with deprotonated Arg186 and two water molecules that would be formed by proton transfer from Arg186 to the OH anion (Wat901). However, constrained geometry optimizations indicate that PC′ is less stable than PC by about 8 kcal/mol and that the rearrangement from PC′ to PC is barrierless. This implies that in the given environment Wat901 is a better proton donor than Arg186 due to the favorable stabilization of the formed OH anion through hydrogen bond interactions. We have to conclude that deprotonated Wat901 will not be replenished by proton transfer via Arg186.

Further inspection of the structure of the product complex (PC) reveals that the hydroxide anion of Wat901 may be connected to the bulk solvent via WatS and three crystallographic water molecules (Wat149, Wat148, and Wat133) and that this network can further be improved by including an additional water molecule in the vacant space between WatS and Wat149. We have confirmed that this extended water

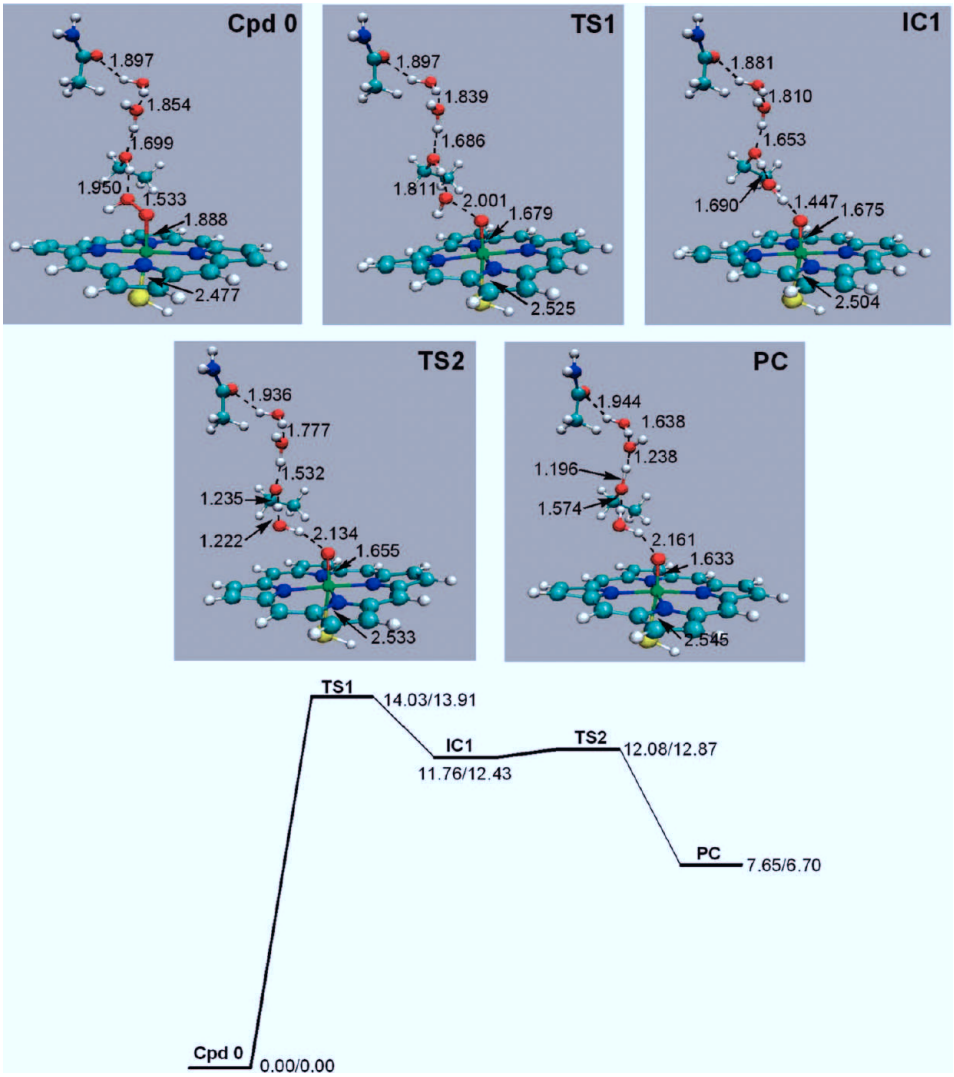


Figure 7. Asn251 mutant, flip model with an extra water. (a) Optimized geometries of Cpd 0, TS1, IC1, TS2, and PC (UB3LYP/B1/CHARMM). Only the QM region is shown. (b) Energy profile of mechanism II. Energies in kcal/mol relative to Cpd 0 (B1/B2).

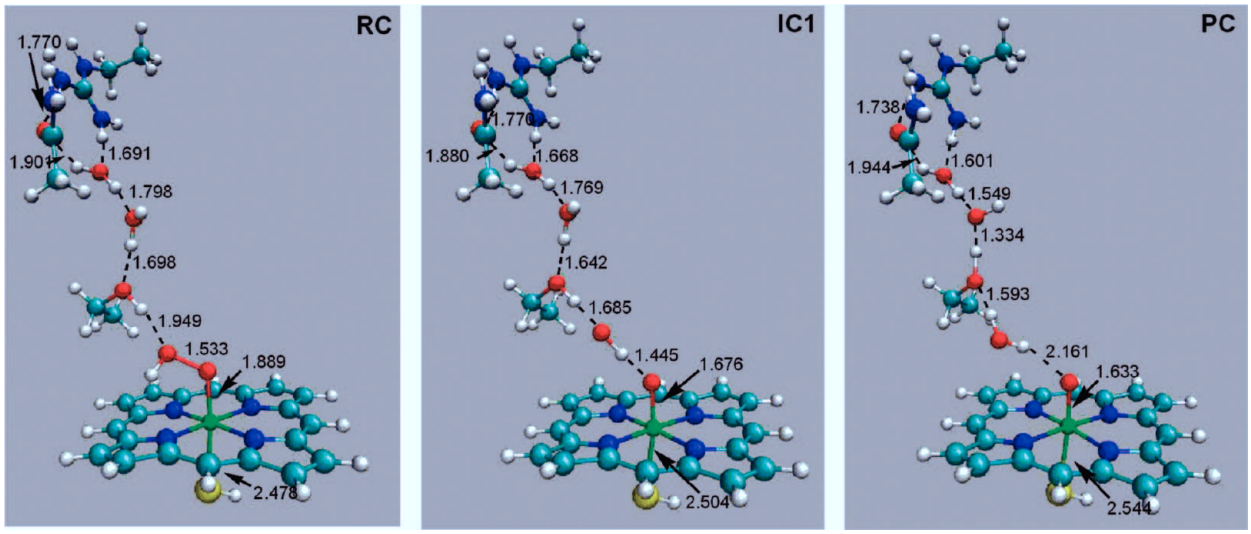


Figure 8. Asn251 mutant, flip model with an extra water and the Arg186 residue included in the QM region: Optimized geometries of Cpd 0, IC1, and PC (UB3LYP/B1/CHARMM). Only the QM region is shown.

network remains intact during 2 ns of classical MD simulation (see Figures S36−S37 and Table S26 in the Supporting Information). This suggests that the formed hydroxide anion may be reprotonated from the bulk solvent by a Grotthuss-type mechanism. It is well-known[49] that the free energy barriers for such proton transfers are quite low in liquid water, for the migration both of excess protons and of proton holes (involving hydroxide anions),[49−51] and similarly low barriers have recently

also been reported for the proton transfer along a wirelike water network in bacteriorhodopsin.[52] We thus consider it likely that such a low-barrier mechanism operates also in our case, but in view of the conformational complexity, we have not attempted to locate any of the corresponding transition states for reprotonation from bulk solvent.

IV. Discussion

In previous QM/MM work,[30] we have investigated the first proton transfer in the catalytic cycle of wild-type P450cam and its D251N mutant that leads to the formation of Cpd 0. In the mutant, the Asn251 side chain was found to be flexible in MD simulations and to flip to an orientation away from the active site, thus generating some empty space that can be occupied by an extra water molecule. Participation of this additional water molecule in the hydrogen-bonding network between Arg186/Asn251 and the heme was shown to provide a viable proton-transfer path in the D251N mutant, even though the resulting rate-limiting barrier remained higher than that in the wild-type enzyme.

In the present QM/MM study we address the second proton transfer in the catalytic cycle that converts Cpd 0 into Cpd I. As in the case of wild-type P450cam,[19] we find that the textbook mechanism with initial protonation of the distal oxygen atom of the FeOOH moiety and subsequent heterolytic O−O cleavage does not operate in the mutant. Instead a mixed homolytic−heterolytic mechanism is again more favorable, with initial O−O cleavage followed by proton transfer in the Asp251/Asn251 channel. The initial step is thus the same in both cases, and the corresponding barriers are indeed almost identical in the wild-type enzyme and the D251N mutant (ca. 14 kcal/mol). This is not surprising, since the cleavage of the O−O bond in the FeOOH moiety should not be influenced much by the Asp251/Asn251 replacement.

This substitution does however affect the subsequent proton transfer step. In wild-type P450cam, protonated Asp251 can act as a proton source, and there is a rather facile concerted pathway for proton delivery to the initially formed, hydrogen-bonded OH species. In the D251N mutant, this path requires significant activation according to the current QM/MM calculations for model I (no flip), and it is not available for models II (flip) and III (flip+WatS), where Asn251 does not serve as a proton donor. As in our previous QM/MM work on the first proton transfer,[30] the preferred arrangement in the mutant involves a flipped Asn251 conformation with an extra water molecule (WatS) that bridges the Asn251 residue and the crystallographic water molecule (Wat901) close to Thr252. The hydrogen-bonding network thus formed facilitates the second proton transfer, and the corresponding transition state lies indeed only about 12−13 kcal/mol above Cpd 0 (i.e., slightly below the transition state for O−O cleavage). However, the resulting product complex (ca. 7−8 kcal/mol above Cpd 0) still contains an OH anion (Wat901) which cannot be protonated by the Arg186 residue (via WatS). This is in contrast to the mechanism of the first proton transfer in the D251N mutant, where such reprotonation was found to be feasible. We note in this context that the two heme species being protonated in the catalytic cycle differ in their total charge (−2 for the reduced oxyheme complex and −1 for Cpd 0) so that one may expect from general electrostatic arguments that reprotonation from the bulk should be less facile in the case of Cpd 0.

In summary, we have identified one viable path for the second proton transfer in the D251N mutant which terminates at Wat901 and is thus not complete, since it does not provide a route for reprotonating the formed OH anion (Wat901). We anticipate, however, that reprotonation may be achieved through a water network that connects this OH anion with the bulk solvent. Inspection of the active-site geometry reveals that there is enough space between the residues Arg186, Asn251, and Asn255 to accommodate another water molecule which could form a stable network connecting WatS and Wat149 (at the boundary to the bulk). Classical MD simulations support this idea and reveal that the hydrogen-bonded network remains intact during 2 ns simulations. Proton delivery along such a network via a Grotthuss-type mechanism seems feasible but has not been studied at present.

Experimentally, the D251N mutant catalyzes the hydroxylation of camphor, but it is significantly less active than wild-type P450cam. The observed decrease in the product formation rate by a factor of 30 implies that the rate-limiting barrier should be about 2 kcal/mol higher in the D251N mutant compared with the wild-type enzyme. It is not clear which step in the catalytic cycle is responsible for this reduced activity. Concerning the conversion from Cpd 0 to Cpd I, we find similar barriers for the initial O−O cleavage in both systems, suggesting that the subsequent proton transfer makes the difference. The QM/MM calculations indicate that the active-site proton transfer events are rather facile both in the wild-type enzyme (from Asp251) and in the flip+WatS model of the D251N mutant (from Wat901). It is thus conceivable that a more difficult reprotonation from bulk solvent contributes to the reduced activity of the mutant.

Experimental solvent kinetic isotope effects (SKIEs) provide further mechanistic information. A recent study[53] reported an SKIE value (H/D) of 1.8 at 190 K (corresponding roughly to 1.6−1.7 at ambient temperature) for the second proton transfer (Cpd 0 → Cpd I) in wild-type P450cam, indicating some solvent participation in this process. Measurements of the turnover rates in various protium−deuterium mixtures gave SKIE values (H/D) of 1.8 for wild-type P450cam and of 10 for the D251N mutant.[10] These steady-state data do not refer to well-defined elementary reaction steps and can thus not be directly related to our computational results, but the dramatic increase for the D251N mutant suggests that solvent water molecules are more heavily involved in the reactions of this mutant compared with the wild-type enzyme. According to proton inventory analysis, the number of protons involved in the rate-limiting step appears to be far larger in the mutant (five to seven) than in wild-type P450cam.[10] These experimental findings are not at odds with the mechanistic scenario outlined above for the D251N mutant (i.e., proton delivery from bulk solvent to the formed OH anion through a water network).

In general, Cpd 0 can undergo two different protonation reactions, namely coupling (formation of Cpd I and water) and uncoupling (formation of the ferric resting state and hydrogen peroxide). According to the present QM/MM calculations, coupling is favored over uncoupling by a large margin in the D251N mutant. This is consistent with the experimental result that hydroxylation is the dominant reaction also in the D251N mutant, which implies the formation of Cpd I as the crucial reactive species in the P450cam consensus mechanism.

V. Conclusion

The present QM/MM study of the D251N mutant of cytochrome P450cam addresses the mechanism of the protonation reactions involving Cpd 0. The QM/MM calculations were performed at the UB3LYP/CHARMM level with two different

basis sets. For all reactions, only minor differences in the computed relative energies were observed upon basis set extension.

Classical MD simulations indicate that the side chain of the Asn251 residue can adopt two conformations: pointing toward the active site (no flip) and pointing toward the protein surface (flip). The flip of the Asn251 side chain breaks the H-bond network that connects the FeOOH moiety and Asn251, and releases enough space to accommodate an additional water molecule. The stability of the extra water molecule (WatS) was confirmed by MD simulation. Hence, three models were considered for the D251N mutant: no flip, flip, and flip with an extra water molecule.

We investigated two mechanisms that correspond to the formation of Cpd I and water (coupling reaction) and two mechanisms that lead to the formation of hydrogen peroxide and the ferric resting state (uncoupling reaction). The results clearly show, in agreement with experimental[54] findings, that the uncoupling reaction is unfavorable in the D251N mutant. Likewise, the mechanism that involves initial protonation of Cpd 0 in the coupling reaction is less likely, since the formation of protonated Cpd 0 is difficult. The coupling reaction in the D251N mutant is thus predicted to follow a stepwise mechanism that involves initial cleavage of the O−O bond and subsequent proton transfer to the distal oxygen. As this initial step does not require the participation of a proton source, there is no significant effect due to the D251N mutation.

The course of the subsequent protonation is found to be model-dependent. In model I (no flip), the Asn251 residue serves as the proton source, in spite of its high pK_a value. The first hydrogen transfer has an effective barrier of 20.6 kcal/mol and leads to the formation of Cpd I and deprotonated Thr252, which is then restored in a second step by accepting a proton from Asn251 via Wat901 and an electron from the heme (transition state 24.5 kcal/mol above Cpd 0). The reprotonation of Asn251 is expected to be facile, since Asn251 is in close contact with the bulk solvent.

In model II (flip) the protonation requires three steps. The path is blocked after the protonation of the Thr252 residue, however, since the Asn251 residue does not act as proton source and there is no other proton source in the vicinity. The corresponding product complex lies 19.2 kcal/mol above Cpd 0. Including the Arg186 residue in this model does not help, since it is too far away to enable proton transfer.

In model III (flipped with an extra water molecule in the active site), there is a well-connected H-bond network that facilitates the formation of Cpd I and the reprotonation of Thr252 from Wat901. The proton transfer is effectively concerted with a barrier of around 14 kcal/mol and an endothermicity of 7.7 kcal/mol. The presence of an extra water thus lowers the barrier appreciably and makes model III most realistic. However, we were unable to reprotonate the formed hydroxide anion (Wat901) from Arg186, so that the most likely scenario is reprotonation from the bulk solvent via a water network that remains intact during 2 ns of classical MD simulation. This process has not been studied at the QM/MM level, but it may well require additional activation and thus increase the effective barrier for Cpd I formation.

Acknowledgment. This work was supported by the Max Planck Society.

Supporting Information Available: Energies, spin densities, group charges, and optimized geometries for mechanisms II, III, and IV. QM/MM energy profiles from pathway calculations. Relative QM/MM energies for mechanism II (with partitioning into QM and MM contributions). MD simulation of the water network in the product complex (mechanism II, model III). Spin densities and group charges of intermediate IC2 in models I and II for different combinations of functional, basis set, and QM region. An overview figure showing all relevant residues. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Omura, T.; Sato, R. *J. Biol. Chem.* **1962**, *237*, 1375–1376.
(2) Ortiz de Montellano, P. R. *Cytochrome P450: Structure, Mechanism and Biochemistry*, 3rd ed.; Plenum Press: New York, 2005.
(3) Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A. M.; Maves, S. A.; Benson, D. E.; Sweet, R. M.; Ringe, D.; Petsko, G. A.; Sligar, S. G. *Science* **2000**, *287*, 1615–1622.
(4) Poulos, T. L.; Finzel, B. C.; Howard, A. J. *J. Mol. Biol.* **1987**, *195*, 687–700.
(5) Aikens, J.; Sligar, S. G. *J. Am. Chem. Soc.* **1994**, *116*, 1143–1144.
(6) Gerber, N. C.; Sligar, S. G. *J. Biol. Chem.* **1994**, *269*, 4260–4266.
(7) Kamachi, T.; Shiota, Y.; Ohta, T.; Yoshizawa, K. *Bull. Chem. Soc. Jpn.* **2003**, *76*, 721–732.
(8) Martinis, S. A.; Atkins, W. M.; Stayton, P. S.; Sligar, S. G. *J. Am. Chem. Soc.* **1989**, *111*, 9252–9253.
(9) Shimada, H.; Makino, R.; Unno, M.; Horiuchi, T.; Ishimura, Y. In *Cytochrome P450cam 8th International Conference*; Libbey, J., Ed.; Eurotext:Paris, 1994, p 299.
(10) Vidakovic, M.; Sligar, S. G.; Li, H.; Poulos, T. L. *Biochemistry* **1998**, *37*, 9211–9219.
(11) Guallar, V.; Friesner, R. A. *J. Am. Chem. Soc.* **2004**, *126*, 8501–8508.
(12) Harris, D. L.; Loew, G. H. *J. Am. Chem. Soc.* **1996**, *118*, 6377–6387.
(13) Harris, D. L.; Loew, G. H. *J. Am. Chem. Soc.* **1994**, *116*, 11671–11674.
(14) Sligar, S. G.; Markis, T. M. Private communications, February 2005.
(15) Gerber, N. C.; Sligar, S. G. *J. Am. Chem. Soc.* **1992**, *114*, 8742–8743.
(16) Lounnas, V.; Wade, R. C. *Biochemistry* **1997**, *36*, 5402–5417.
(17) Harris, D. L.; Loew, G. H. *J. Am. Chem. Soc.* **1998**, *120*, 8941–8948.
(18) Taraphder, S.; Hummer, G. *J. Am. Chem. Soc.* **2003**, *125*, 3931–3940.
(19) Zheng, J.; Wang, D.; Thiel, W.; Shaik, S. *J. Am. Chem. Soc.* **2006**, *128*, 13204–13215.
(20) Benson, D. E.; Suslick, K. S.; Sligar, S. G. *Biochemistry* **1997**, *36*, 5104–5107.
(21) Imai, M.; Shimada, H.; Watanabe, Y.; Matsuhima-Hibiya, Y.; Makino, R.; Koga, H.; Horiuchi, T.; Ishimura, Y. *Proc. Natl. Acad. Sci. U. S. A.* **1989**, *86*, 7823–7827.
(22) Oprea, T. I.; Hummer, G.; Garcïa, A. E. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 2133–2138.
(23) Raag, R.; Martinis, S. A.; Sligar, S. G.; Poulos, T. L. *Biochemistry* **1997**, *36*, 5104–5107.
(24) Kamachi, T.; Yoshizawa, K. *J. Am. Chem. Soc.* **2003**, *125*, 4652–4661.
(25) Kumar, D.; Hirao, H.; De Visser, S. P.; Zheng, J.; Wang, D.; Thiel, W.; Shaik, S. *J. Phys. Chem. B* **2005**, *109*, 19946–19951.
(26) Ogliaro, F.; De Visser, S. P.; Cohen, S.; Sharma, P. K.; Shaik, S. *J. Am. Chem. Soc.* **2002**, *124*, 2806–2817.
(27) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. *Chem. Rev.* **2005**, *105*, 2279–2328.
(28) Hishiki, T.; Shimada, H.; Nagano, S.; Egawa, T.; Kanamori, Y.; Makino, R.; Park, S. Y.; Adachi, S. I.; Shiro, Y.; Ishimura, Y. *J. Biochem.* **2000**, *128*, 965–974.
(29) Altarsha, M.; Benighaus, T.; Kumar, D.; Thiel, W. *J. Am. Chem. Soc.* **2009**, *131*, 4755–4763.
(30) Wang, D.; Zheng, J.; Shaik, S.; Thiel, W. *J. Phys. Chem. B* **2008**, *112*, 5126–5138.
(31) Altun, A.; Thiel, W. *J. Phys. Chem. B* **2005**, *109*, 1268–1280.
(32) Schöneboom, J. C.; Lin, H.; Reuter, N.; Thiel, W.; Cohen, S.; Ogliaro, F.; Shaik, S. *J. Am. Chem. Soc.* **2002**, *124*, 8142–8151.
(33) Schöneboom, J. C.; Thiel, W. *J. Phys. Chem. B* **2004**, *108*, 7468–7478.
(34) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(35) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.;

Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E., III; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(36) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.

(37) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648–5652.

(38) Hay, P. J.; Wadt, W. R. *J. Chem. Phys.* **1985**, *82*, 299–310.

(39) Hehre, W. J.; Ditchfield, K.; Pople, J. A. *J. Chem. Phys.* **1972**, *56*, 2257–2261.

(40) Schäfer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571–2577.

(41) Schäfer, A.; Huber, C.; Ahlrichs, R. *J. Chem. Phys.* **1994**, *100*, 5829–5835.

(42) Smith, W.; Forester, T. R. *J. Mol. Graphics* **1996**, *14*, 136–141.

(43) Sherwood, P.; De Vries, A. H.; Guest, M. F.; Schreckenbach, G.; Catlow, C. R. A.; French, S. A.; Sokol, A. A.; Bromley, S. T.; Thiel, W.; Turner, A. J.; Billeter, S.; Terstegen, F.; Thiel, S.; Kendrick, J.; Rogers, S. C.; Casci, J.; Watson, M.; King, F.; Karlsen, E.; Sjøvoll, M.; Fahmi, A.; Schäfer, A.; Lennartz, C. *THEOCHEM* **2003**, *632*, 1–28.

(44) Ahlrichs, R.; Bär, M.; Häser, M.; Horn, H.; Kälmel, C. *Chem. Phys. Lett.* **1989**, *162*, 165–169.

(45) Billeter, S. R.; Turner, A. J.; Thiel, W. *Phys. Chem. Chem. Phys.* **2000**, *2*, 2177–2186.

(46) Bakowies, D.; Thiel, W. *J. Phys. Chem.* **1996**, *100*, 10580–10594.

(47) Sherwood, P.; De Vries, A. H.; Collins, S. J.; Greatbanks, S. P.; Burton, N. A.; Vincent, M. A.; Hillier, I. H. *Faraday Discuss.* **1997**, *106*, 79–92.

(48) Nagano, S.; Poulos, T. L. *J. Biol. Chem.* **2005**, *280*, 31659–31663.

(49) Marx, D. *ChemPhysChem* **2006**, *7*, 1848–1870.

(50) Muller, R. P.; Warshel, A. *J. Phys. Chem.* **1995**, *99*, 17516–17524.

(51) Tuckerman, M. E.; Chandra, A.; Marx, D. *Acc. Chem. Res.* **2006**, *39*, 151–158.

(52) Mathias, G.; Marx, D. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 6980–6985.

(53) Kim, S. H.; Yang, T.-C.; Perera, R.; Jin, S.; Bryson, T. A.; Sono, M.; Davydov, R.; Dawson, J. H.; Hoffman, B. M. *Dalton Trans.* **2005**, 3464–3469.

(54) Markis, T. M.; Denisov, I. G.; Schlichting, I.; Sligar, S. G. In *Cytochrome P450: Sturcture, Mechanism and Biochemistry*, 2nd ed.; Plenum Press: New York, 1995.

JP809838K

J Biol Inorg Chem
DOI 10.1007/s00775-009-0608-3

ORIGINAL PAPER

Coupling and uncoupling mechanisms in the methoxythreonine mutant of cytochrome P450cam: a quantum mechanical/molecular mechanical study

Muhannad Altarsha · Tobias Benighaus · Devesh Kumar · Walter Thiel

Received: 6 August 2009 / Accepted: 4 November 2009

Abstract The Thr252 residue plays a vital role in the catalytic cycle of cytochrome P450cam during the formation of the active species (Compound I) from its precursor (Compound 0). We investigate the effect of replacing Thr252 by methoxythreonine (MeO-Thr) on this protonation reaction (coupling) and on the competing formation of the ferric resting state and H_2O_2 (uncoupling) by combined quantum mechanical/molecular mechanical (QM/MM) methods. For each reaction, two possible mechanisms are studied, and for each of these the residues Asp251 and Glu366 are considered as proton sources. The computed QM/MM barriers indicate that uncoupling is unfavorable in the case of the Thr252MeO-Thr mutant, whereas there are two energetically feasible proton transfer pathways for coupling. The corresponding rate-limiting barriers for the formation of Compound I are higher in the mutant than in the wild-type enzyme. These findings are consistent with the experimental observations that the Thr252MeO-Thr mutant forms the alcohol product exclusively (via Compound I), but at lower reaction rates compared with the wild-type enzyme.

Keywords Cytochrome P450 · Methoxythreonine · *O*-Methylthreonine · Proton transfer · Quantum mechanics/Molecular mechanics

Electronic supplementary material The online version of this article (doi:10.1007/s00775-009-0608-3) contains supplementary material, which is available to authorized users.

M. Altarsha · T. Benighaus · D. Kumar · W. Thiel (✉)
Max-Planck-Institut für Kohlenforschung,
Kaiser-Wilhelm-Platz 1,
45470 Mülheim, Germany
e-mail: thiel@mpi-muelheim.mpg.de

Introduction

Cytochrome P450 is one of the most versatile enzymes in nature [1, 2]. It uses dioxygen to catalyze a great variety of stereospecific and regioselective processes of oxygen insertion into organic compounds [3–8]. These processes are of vital importance in biosystems, where the enzyme participates in detoxification and in biosyntheses [1]. Since the activation of inert C–H bonds is one of the holy grails of chemistry [9], the facility to carry out this process makes the P450 enzyme superfamily a model for creative mimetic chemistry [10] designed to generate novel catalysts that can perform C–H activation.

The bacterial enzyme P450cam (CYP101) is the workhorse of P450 research which has generated much insight into the role of the protein in regulating the activity of the enzyme and the effects of site-directed mutations [4, 11–14]. Its active site contains a heme unit that consists of an iron protoporphyrin IX complex with Cys as the proximal axial ligand. The active catalytic species, with an Fe(IV)–oxo moiety, is commonly denoted as Compound I. It has been observed in a related chloroperoxidase, but is still elusive for P450 enzymes, where it has only been identified by transient spectroscopy [8].

Site-directed mutagenesis studies [15–18] in combination with X-ray structural analyses [19, 20] indicate that the conserved P450 residue Thr252 at the active site plays a crucial role in the catalysis, in particular during the formation of Compound I [21, 22]. Hence, it was no surprise that Thr252 became an early target for mutagenesis [23, 24]. Several mutants such as Thr252Ala and Thr252Gly show an uncoupling of O_2 consumption from D-camphor hydroxylation, most of the O_2 consumed being converted to H_2O_2 without cleaving the O–O bond (Scheme 1), whereas Thr252Ser retains significant coupling of O_2

Published online: 20 November 2009

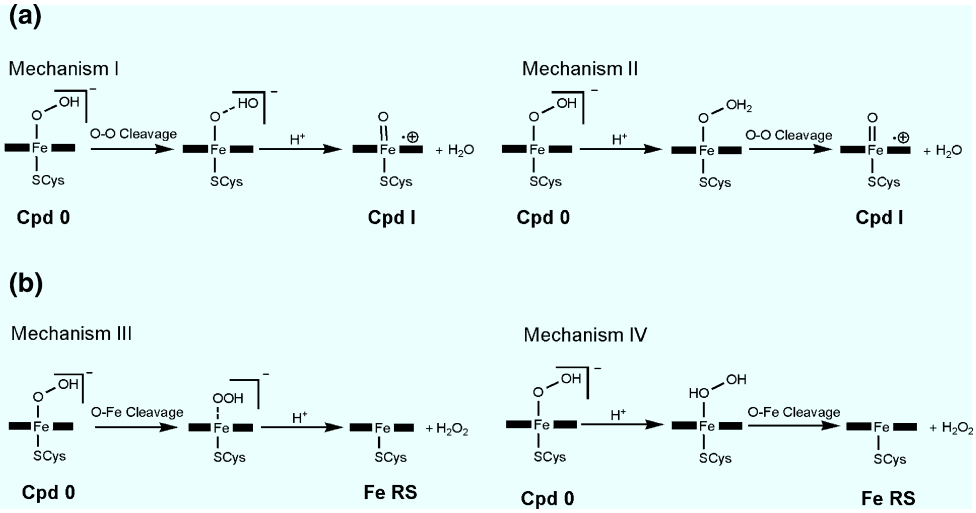
consumption with D-camphor hydroxylation [15]. When Thr252 is replaced by methoxythreonine (MeO-Thr), the resulting Thr252MeO-Thr mutant gives 100% formation of 5-*exo*-hydroxycamphor (no uncoupling), but the rate of reaction is slowed down to one third compared with that for the wild-type enzyme [25]. None of the other Thr252X mutants studied preserves the coupling activity to a similar extent [25]. It is generally assumed that site-directed Thr252X mutagenesis will disrupt the proton relay that converts Compound 0 to Compound I in P450 enzymes [26, 27].

Several theoretical studies have addressed the protonation reactions that generate Compound I [28–37]. The commonly formulated mechanism is protonation of Compound 0 at the distal oxygen atom followed by O–O bond cleavage. Recent density functional theory calculations on a large gas-phase active-site model (96 atoms) indeed gave a stable protonated Compound 0 species with significant barriers for the conversion toward both Compound 0 and Compound I [32]. However, subsequent quantum mechanical/molecular mechanical (QM/MM) calculations showed that such an intermediate is quite instable in the enzyme (more than 20 kcal/mol above Compound 0) and mechanistically irrelevant (barriers of only 3–4 kcal/mol for the decay to Compound 0 and Compound I) [37]. An alternative mechanism was proposed that involves an initial O–O bond cleavage followed by a proton transfer to the OH species formed (via a hydrogen-bonding network in the Asp251 channel), with a concomitant electron transfer from the heme (yielding Compound I and water) [37]. A similar mechanistic scenario was considered in heme oxygenase and chloroperoxidase [38, 39]. The latest QM/MM work [40] on this topic investigated both the coupling and the uncoupling reactions of Compound 0 in the wild-type P450cam enzyme and in four Thr252X mutants (X = Ser, Val, Ala, Gly). It was found that the formation of Compound I (coupling) always proceeds through the two-step mechanism with initial O–O bond cleavage [40]. By contrast, the uncoupling reaction is always concerted. Its barrier is always higher than that of the coupling reaction if the Asp251 channel contains only residue 252, the crystallographic water molecule Wat901, and protonated Asp251. Molecular dynamics (MD) simulations indicate, however, that an additional water molecule is stable in the Asp251 channel for X = Val, Ala, and Gly, which leads to much smaller barriers for uncoupling owing to a much more favorable hydrogen-bonding network. Including this extra water molecule in the QM region makes uncoupling competitive with coupling in the case of X = Val and renders it more facile for X = Ala and Gly [40], in qualitative agreement with experiment [15].

Here, we extend our previous QM/MM work by considering the effect of the Thr252MeO-Thr mutation. We address both the coupling and the uncoupling reactions and attempt to answer the question whether the Thr252MeO-Thr mutation will indeed disrupt the proton relay channel that is commonly viewed as being an essential prerequisite for the conversion of Compound 0 to Compound I.

Computational methods and proposed mechanisms

The initial structure was taken from the MD trajectory of the native enzyme studied earlier [37]. Thr was mutated into MeO-Thr by manually replacing the OH group in the Thr252 residue by OCH_3. The same solvation and protonation schemes were applied as in previous studies [41–43]. Glu366 and Asp251 were considered as possible proton sources [17, 31, 44], and the corresponding two



Scheme 1 **a** Two mechanisms for the conversion of Compound 0 (*Cpd 0*) to Compound I (*Cpd I*, coupling reaction). **b** Two mechanisms for ferric resting state (*Fe RS*) formation (uncoupling reaction)

protonation schemes were adopted as in the standard setup used previously [41–43] (i.e., protonated Glu366 and deprotonated Asp251 in the Glu366 channel, and deprotonated Glu366 and protonated Asp251 in the Asp251 channel). Both setups consisted of 24,988 atoms in total, including 5,891 TIP3P water molecules [45]. The solvated systems were relaxed by performing classical energy minimizations and MD simulations at the MM level using the CHARMM22 force field [46] as implemented in the CHARMM program [47]. The heme units with the Cys357 and OOH ligands as well as the outer 8 Å of the solvent layer were kept fixed during these initial runs.

The QM/MM method chosen was analogous to that used in our previous studies [41–43]. Here, we briefly mention some aspects relevant to the present work. Minimized snapshots from the MD trajectories were taken as initial structures for QM/MM optimizations. In the QM/MM calculations, the QM part was treated by unrestricted hybrid density functional theory (UB3LYP) [48] with the LACVP [49] small-core effective core potential basis set on iron and 6-31G [50] on the remaining atoms (B1) for geometry optimizations, while the MM part was described by the CHARMM22 force field. Single-point calculations were carried out with the TZVP [51, 52] basis set (B2).

An electronic embedding scheme [53] was adopted in the QM/MM calculations, i.e., interactions with MM charges were incorporated into the one-electron Hamiltonian of the QM calculation. No cutoffs were introduced for the nonbonding MM and QM/MM interactions. Hydrogen link atoms with the charge shift model were employed to treat the QM/MM boundary. The TURBOMOLE program [54] was used for the QM treatment in the QM/MM calculations as well as in pure QM calculations. The CHARMM22 force field was run through the DL_POLY [55] program to handle the MM part of the systems. The QM/MM calculations were performed with the ChemShell package [56], which integrates the TURBOMOLE and DL_POLY programs and performs geometry optimization with the HDLC optimizer [57].

Possible proton transfer pathways

Scheme 1 shows the four proposed mechanisms that were investigated for both protonation channels (Glu366 and Asp251).

In mechanism I, initially the O–O bond is cleaved to generate an OH radical and one-electron-reduced Compound I. Subsequently, a proton is transferred to the OH radical with a concomitant electron transfer from the heme that yields Compound I and water [37].

In mechanism II, a proton is transferred to the distal oxygen atoms of the hydroperoxo group to form protonated Compound 0 (containing $FeOOH_2$), followed by heterolytic O–O bond cleavage that generates Compound I and water.

Mechanisms I and II both give Compound I and correspond to the coupling reaction (Scheme 1a).

In mechanism III, initially the Fe–O bond is cleaved to generate an OOH radical, followed by a proton transfer to the OOH group that yields the ferric resting state and hydrogen peroxide.

In mechanism IV, a proton is transferred to the proximal oxygen atom of the hydroperoxo group to form an FeH_2O_2 moiety, followed by heterolytic cleavage of the O–Fe bond generating the ferric resting state and hydrogen peroxide.

Mechanisms III and IV both yield the ferric resting state and hydrogen peroxide (uncoupling reaction, Scheme 1b).

QM region

In the QM/MM calculations, we employed QM regions analogous to those adopted for the wild-type enzyme in our previous work [37] (Fig. 1), except that the Thr252 residue was replaced by MeO-Thr. In both channels (Asp251 and Glu366), the QM region included: iron porphine (without heme side chains), the sulfur atom of Cys357, the axial OOH moiety, and MeO-Thr (represented by $CH_3OCH_2CH_3$). In addition, the QM region also contained Wat901 and Asp251 (represented by CH_3COOH) in the case of the Asp251 channel, and Wat523, Wat566, Wat687, Wat902, and Glu366 (represented by CH_3COOH) in the case of the Glu366 channel (Fig. 1). Hence, the water molecules that may be involved in the proton transfer are part of the QM region for each channel.

Compound 0 can exist in a doublet, quartet or sextet state. It has a doublet ground state both in the wild-type P450cam enzyme and in the Thr252X mutants. According to the QM/MM calculations, the lowest quartet and sextet states lie 8.3 and 9.0 kcal/mol above the doublet ground state of the Thr252MeO-Thr mutant, respectively. Therefore, we only studied the reactions in the doublet state of the mutant, as was done previously in the case of the wild-type enzyme [37].

Results

Figures 2, 3, 4, 5, 6, 7, and 8 show the optimized QM/MM geometries of the QM regions for all relevant minima and transition states (Thr252MeO-Thr mutant, mechanisms I–IV, Glu366 and Asp251 channels). The computed relative QM/MM energies of the stationary points are summarized for basis sets B1/B2 in Table 1 (coupling reaction, mechanisms I and II) and Table 2 (uncoupling reaction, mechanisms III and IV). The single-point energies obtained

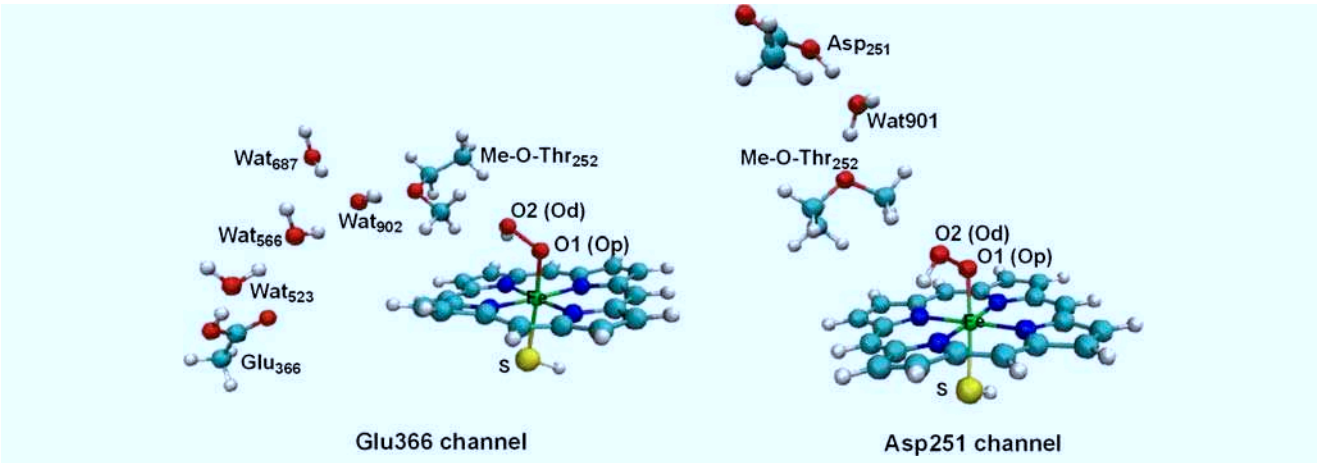


Fig. 1 Quantum mechanical region for the Thr252MeO-Thr mutant in the Glu366 and Asp251 channels

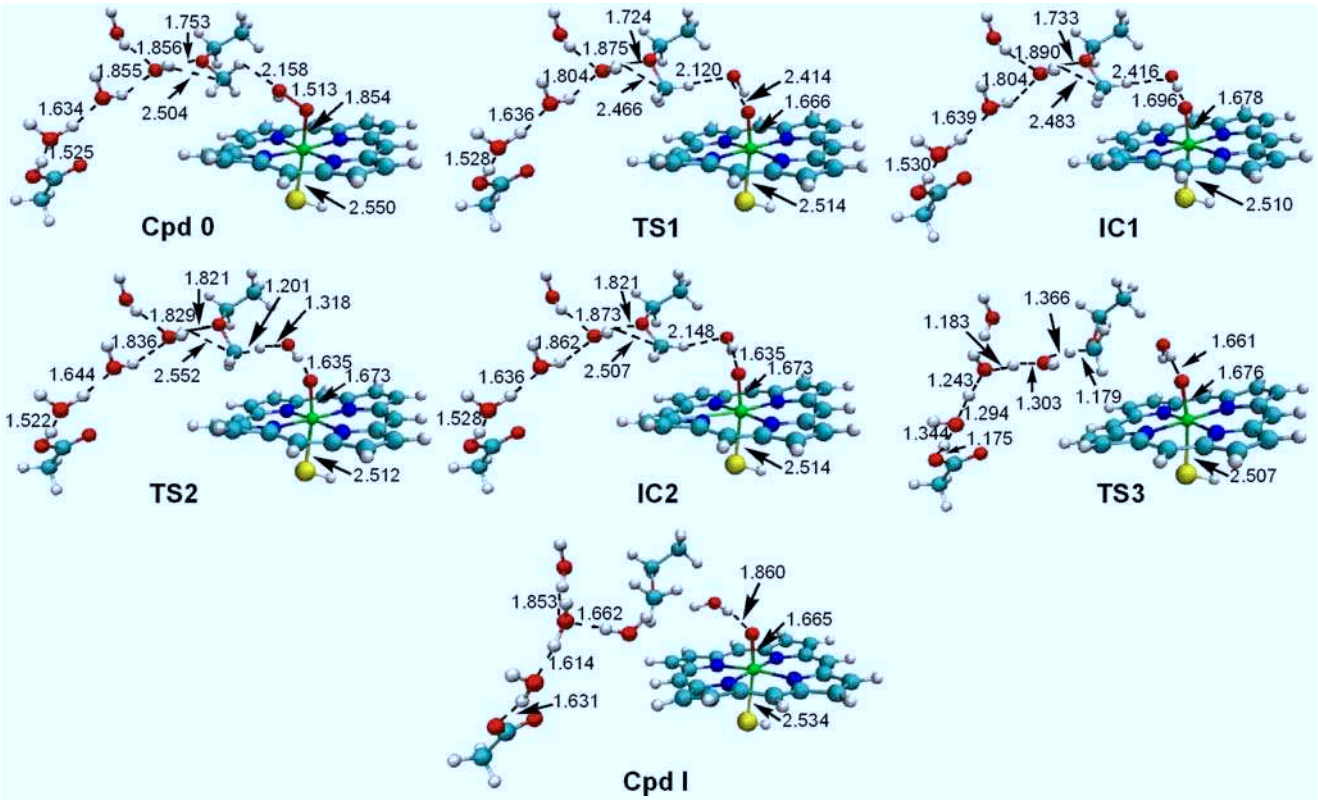


Fig. 2 Optimized geometries (UB3LYP/B1/CHARMM) for mechanism I (coupling reaction) in the Glu366 channel

with the larger TZVP basis (B2) at the corresponding optimized QM/MM geometries (B1) are generally quite similar to those obtained with the smaller basis (B1), although they are consistently slightly higher relative to Compound 0, typically by 1–3 kcal/mol. A similar behavior was also observed for the wild-type enzyme [37]. In the following discussion, we shall only quote B1 results for the sake of consistency (energies, geometries, etc.). Formation of the correct intermediates and products was verified by analysis of the spin densities and Mulliken charges. These data and selected geometrical parameters are documented in the electronic supplementary material.

Mechanism I: homolytic O–O bond cleavage followed by coupled proton–electron transfer

Glu366 channel

The first step passes over a barrier of 18.1 kcal/mol and leads to an intermediate (IC1), in which the OH moiety

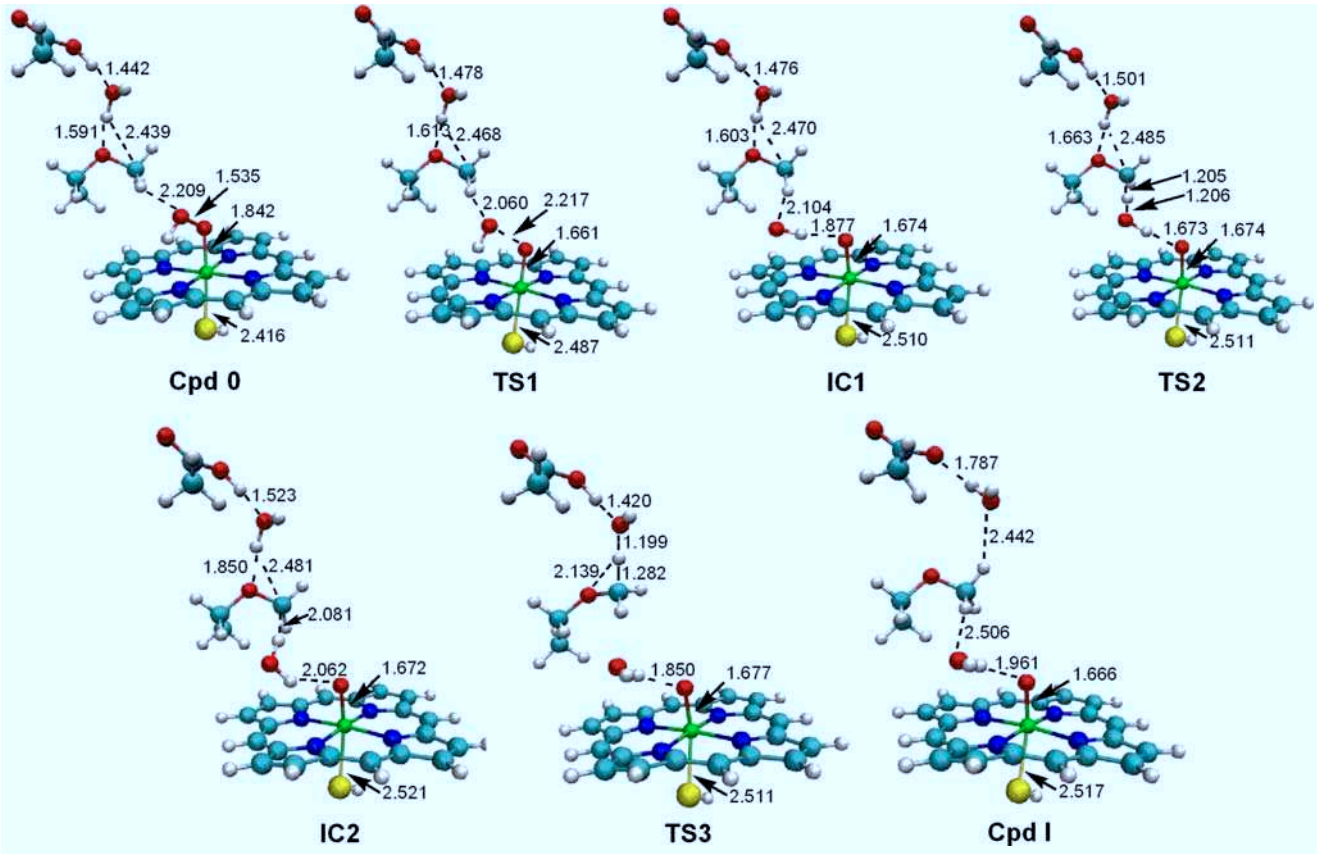


Fig. 3 Optimized geometries (UB3LYP/B1/CHARMM) for mechanism I (coupling reaction) in the Asp251 channel

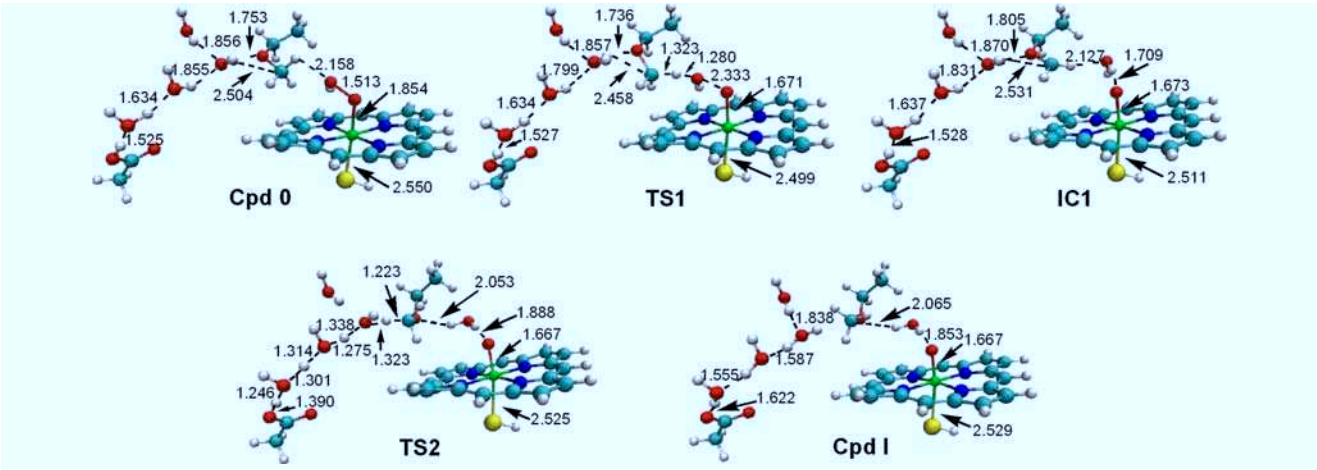


Fig. 4 Optimized geometries (UB3LYP/B1/CHARMM) for mechanism II (coupling reaction) in the Glu366 channel

forms two hydrogen bonds with MeO-Thr and with Fe=O (Fig. 2). During this step, the Fe–O bond shortens to 1.67 Å in TS1 and then remains at 1.68 Å in IC1. These structural features are similar to those reported for the wild-type enzyme [37]. The spin density and partial charge of the OH group in the first intermediate (IC1) are −0.93 and −0.04, indicating that IC1 contains an OH radical and one-electron-reduced Compound I. This suggests that the O–O bond cleavage is homolytic: the Fe=O moiety carries two unpaired electrons, and the third unpaired electron is mainly located on the OH moiety. IC1 is stabilized by hydrogen-bonding interactions of OH with FeO and MeO-Thr252, and therefore lies only 10.3 kcal/mol above the reactant.

The second step is a hydrogen transfer from the MeO-Thr group to the OH moiety which yields Compound I and

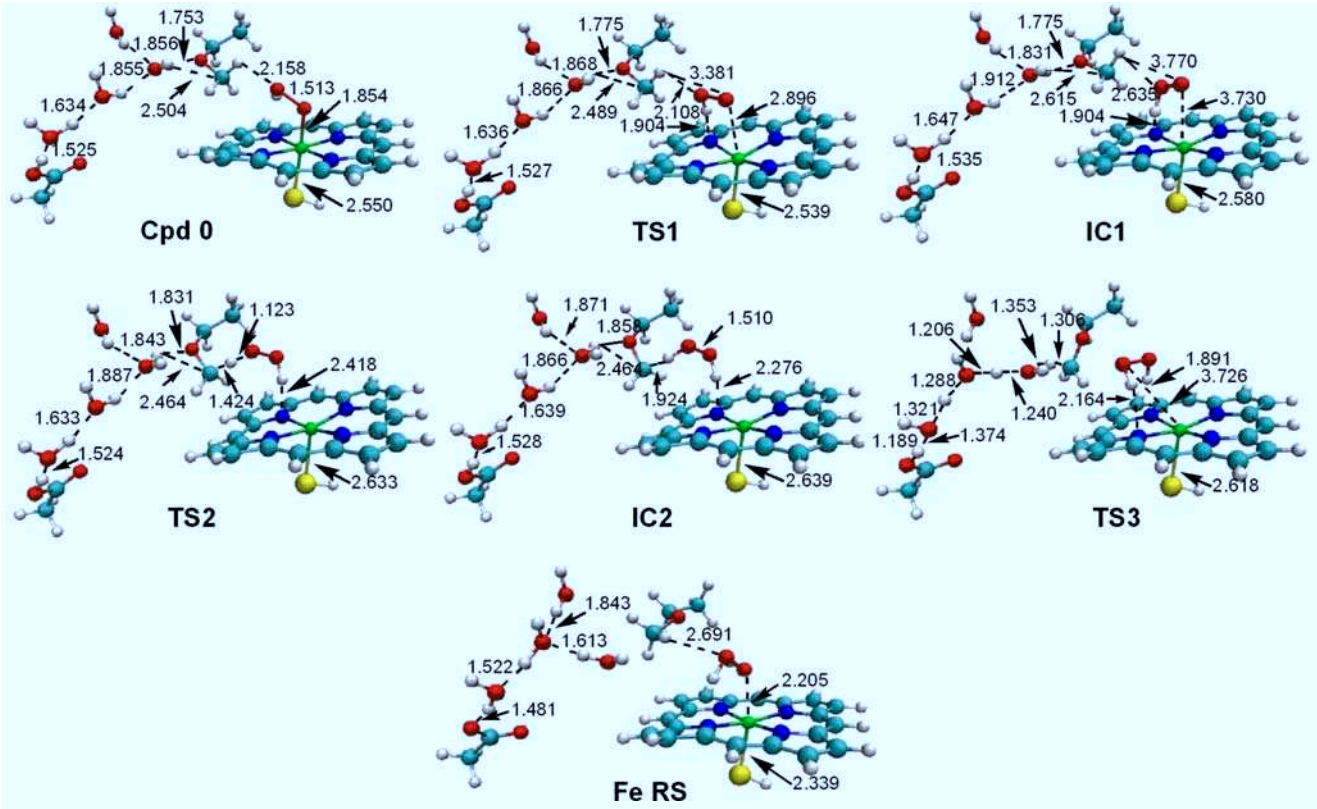


Fig. 5 Optimized geometries (UB3LYP/B1/CHARMM) for mechanism III (uncoupling reaction) in the Glu366 channel

water. The corresponding transition state (TS2) lies 18.5 kcal/mol above Compound 0, and the intermediate complex of CH_2O-Thr radical with Compound I (IC2) is quite stable, with an energy of 1.1 kcal/mol relative to Compound 0. The OH moiety is obviously reactive enough to abstract a proton from the methoxy group, and the resulting intermediate (IC2) is stabilized by Wat902 and the water molecule formed via two hydrogen bonds.

In the last step, a proton is transported from Glu366 to MeO-Thr in a concerted process via three bridging water molecules. Simultaneously, an electron is transferred from the heme to the methylene group to regenerate the MeO-Thr and form a π cation radical at the heme. The transition state (TS3) and the product (Compound I) lie 17.2 and 8.0 kcal/mol above Compound 0, respectively. The hydrogen-bonding network between Glu366 and MeO-Thr is reoriented after the proton transfer. Overall, the rate-limiting step is the hydrogen abstraction from the methoxy group with a barrier of 18.5 kcal/mol (TS2).

Asp251 channel

In this channel, the barrier of O–O bond cleavage is 18.6 kcal/mol (TS1 in Fig. 3), similar to the corresponding barrier in the Glu366 channel (18.1 kcal/mol). The intermediate (IC1, OH moiety and one-electron-reduced Compound I) is rather high in energy (14.4 kcal/mol). For the conversion of IC1 to Compound I, a proton needs to be transported from the Asp251 carboxyl group via Wat901 and MeO-Thr to OH, with a concomitant electron transfer from the heme. The spin density and partial charge of the OH group in IC1 are −0.79 and −0.12, indicating that OH will not behave as a "perfect" radical in IC1 owing to the strong hydrogen-bonding interactions with the methoxy group (2.10 Å) and the FeO unit (1.88 Å). In contrast to the wild-type enzyme [37], the subsequent proton delivery proceeds in two steps. As in the Glu366 channel, a hydrogen atom is first transferred from the methoxy group of MeO-Thr (TS2 at 22.6 kcal/mol, i.e., 8.2 kcal/mol above IC1). The intermediate formed (IC2 at 11.4 kcal/mol) then receives a proton through the Asp251 channel and an electron from the heme in a simultaneous process (TS3 at 23.0 kcal/mol). After releasing its proton, the side chain of Asp251 rotates back into a salt bridge with Arg186, as shown in Fig. 3.

Comparison

In each channel, the three transition states lie at similar energies relative to Compound 0. The highest point in the reaction profile is TS2 (TS3) in the Glu366 (Asp251)

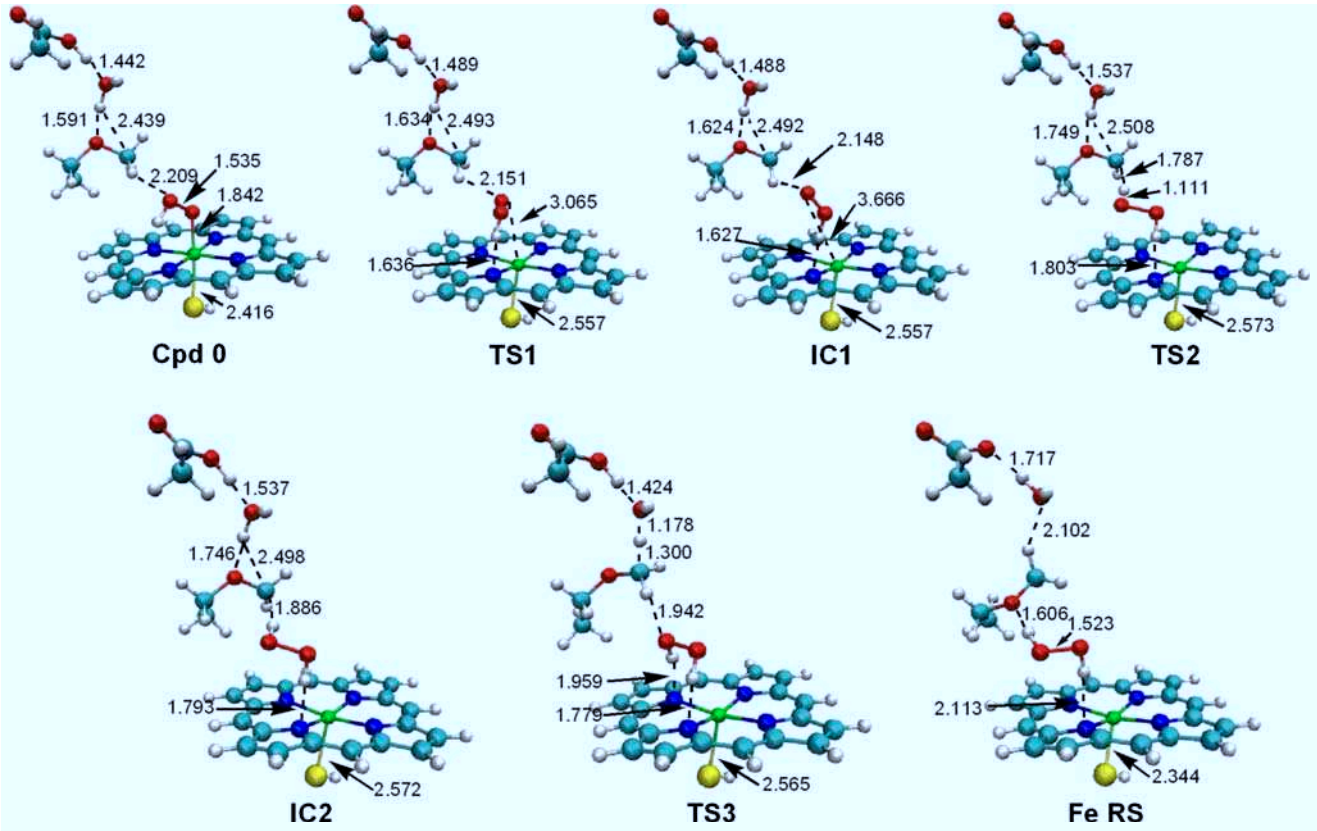


Fig. 6 Optimized geometries (UB3LYP/B1/CHARMM) for mechanism III (uncoupling reaction) in the Asp251 channel

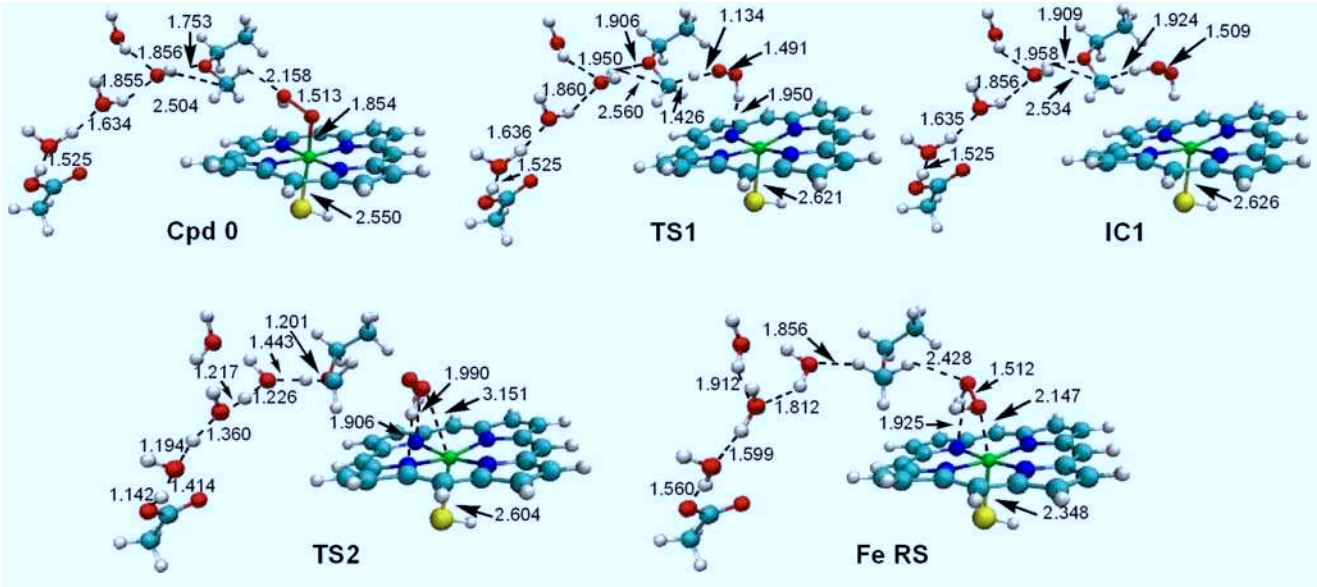


Fig. 7 Optimized geometries (UB3LYP/B1/CHARMM) for mechanism IV (uncoupling reaction) in the Glu366 channel

channel at 18.5 (23.0) kcal/mol (see Table 1), i.e., about 4–8 kcal/mol higher than in the wild-type enzyme [37]. The conversion of Compound 0 to Compound I via mechanism I should thus be much slower in the Thr252MeO-Thr mutant compared with the wild-type enzyme.

Additional snapshot

To ensure that the snapshot used in this study is representative for the system, reaction mechanism I in the Asp251 channel was also studied in an analogous manner

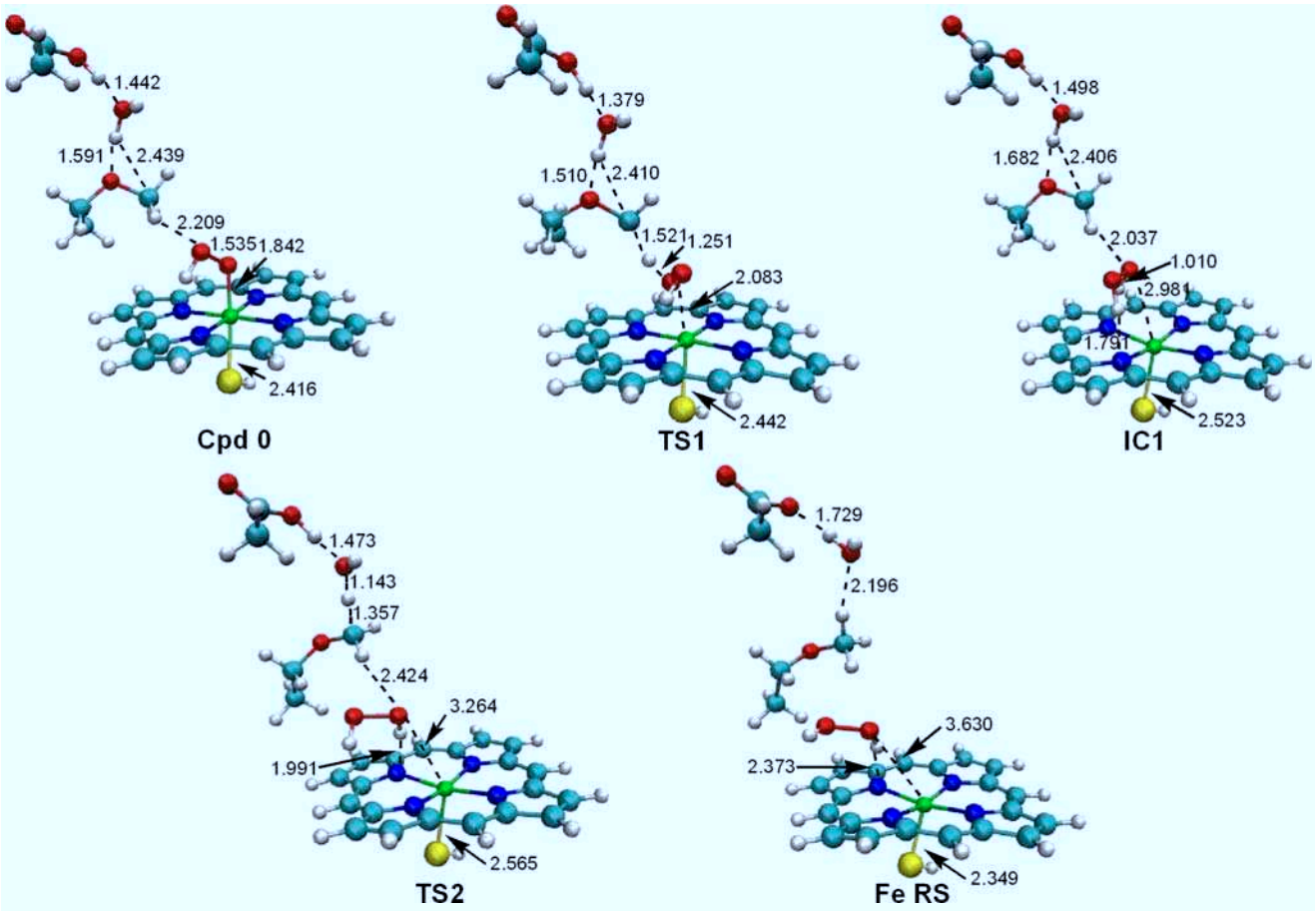


Fig. 8 Optimized geometries (UB3LYP/B1/CHARMM) for mechanism IV (uncoupling reaction) in the Asp251 channel

Table 1 Quantum mechanical/molecular mechanical (QM/MM) relative energies (kcal/mol) for optimized structures of mechanisms I and II of the coupling reaction for the Glu366 and Asp251 channels using basis sets B1/B2 (relative to Compound 0)

	Mechanism I						Mechanism II			
	TS1	IC1	TS2	IC2	TS3	Compound1	TS1	IC1	TS2	Compound I
Glu366 channel	18.1/20.7	10.3/13.0	18.5/21.2	1.1/4.5	17.2/20.3	8.0/12.2	20.8/23.8	0.8/2.3	20.0/24.5	8.0/13.5
Asp251 channel	18.6/19.0	14.4/15.4	22.6/23.6	11.4/14.0	23.0/25.9	7.7/11.3				

TS transition state, *IC* intermediate compound

Table 2 QM/MM relative energies (kcal/mol) for optimized structures of mechanisms III and IV of the uncoupling reaction for the Glu366 and Asp251 channels using basis sets B1/B2 (relative to Compound 0)

	Mechanism III						Mechanism IV			
	TS1	IC1	TS2	IC2	TS3	Fe RS	TS1	IC1	TS2	Fe RS
Glu366 channel	30.3/32.2	17.7/20.0	42.4/43.8	40.6/41.7	60.7/64.0	28.4/30.0	40.3/42.0	38.0/36.1	58.3/59.6	25.2/27.7
Asp251 channel	28.1/30.0	25.7/28.6	47.6/50.1	47.4/49.0	58.4/61.5	28.8/30.9	52.9/55.1	47.6/50.7	54.9/55.8	26.4/29.1

Fe RS ferric resting state

with a different snapshot which was drawn after 1,500 ps of MD simulation. The computed relative energies of all stationary points (Table S23) agree with those from the first snapshot (Table 1) to within 1 kcal/mol. The highest point in the reaction profile (TS3) is at 23.2 kcal/mol, very close to the value of 23.0 kcal/mol from the first snapshot (see earlier). The results from both snapshots are thus entirely consistent with each other.

Mechanism II: proton-assisted heterolytic O–O bond cleavage

Glu366 channel

The energy barrier for direct hydrogen atom transfer from MeO-Thr to FeOOH is 20.8 kcal/mol, and the resulting intermediate (IC1, Fig. 4) lies 0.8 kcal/mol above Compound 0. The unpaired electron is mainly located on the iron atom (iron spin density of 1.37). In contrast to the reaction in the wild-type enzyme, IC1 is not a protonated Compound 0 species, since the O–O bond is cleaved in the first step [37]. However, mechanism II differs from mechanism I, since the hydrogen transfer is part of the first step. In the second step, the concomitant transport of one proton (from Glu366) and one electron (from the heme) leads to formation of Compound I. The relative energies of TS2 and Compound I are 20.0 and 8.0 kcal/mol, respectively.

Asp251 channel

In the Asp251 channel, we chose several different reaction coordinates to convert Compound 0 to protonated Compound 0 by proton transfer from Asp251 to the distal oxygen atom of the hydroperoxo group. However, all energy scans led to continuously increasing energy profiles, and we were unable to locate protonated Compound 0. Similar problems have also been reported in previous QM/MM calculations for the wild-type enzyme [37].

Mechanism III: homolytic O_1–Fe bond cleavage followed by coupled proton–electron transfer

Glu366 channel

The optimized geometries are presented in Fig. 5. The barrier (TS1) for homolytic breaking of the O_1–Fe bond is 30.3 kcal/mol, and the intermediate (IC1) consisting of iron-bound heme and the OOH radical lies 17.7 kcal/mol above Compound 0. The subsequent hydrogen transfer from MeO-Thr to OOH is very difficult (TS2 at 42.4 kcal/mol, thus 24.7 kcal/mol above IC1), and the second intermediate (IC2) with iron-bound heme and the CH_2O-Thr radical is a shallow minimum (IC2 at 40.6 kcal/mol). The barrier for final proton transfer from Glu366 to CH_2O-Thr with concomitant electron transfer from the heme is prohibitively high (TS3 at 60.7 kcal/mol). The overall reaction is endothermic by 28.4 kcal/mol.

Asp251 channel

Figure 6 shows the optimized geometries. In general, the barriers are quite similar to those in the Glu366 channel. The barrier (TS1) for homolytic cleavage of the O_1–Fe bond is 28.1 kcal/mol. In the resulting intermediate (IC1 at 25.7 kcal/mol), the spin densities of OOH (−0.97) and iron (1.98) indicate that iron has two unpaired electrons and that OOH is present as a radical. The following hydrogen transfer from MeO-Thr to OOH is again difficult (TS2 at 47.6 kcal/mol, hence 21.9 kcal/mol above IC1) and leads to a very shallow intermediate (IC2 at 47.4 kcal/mol) with a CH_2O-Thr radical (spin density of −0.92). The final proton transfer from Asp251 to CH_2O-Thr requires much activation (TS3 at 58.4 kcal/mol, i.e., 11.0 kcal/mol above IC2). At the end of the reaction, Asp251 rotates to rebuild the salt bridge with Arg186, as also found in mechanism I. The overall reaction is endothermic by 28.8 kcal/mol.

Mechanism IV: proton-assisted heterolytic O–Fe bond cleavage

Glu366 channel

As can be seen from Fig. 7, the first step involves O–Fe bond cleavage combined with a hydrogen transfer from MeO-Thr to the proximal oxygen atom. The corresponding barrier is high (TS1 at 40.3 kcal/mol), and the shallow intermediate (IC1 at 38.0 kcal/mol) contains essentially neutral hydrogen peroxide with almost zero spin density and an O_1–O_2 distance of 1.51 Å; the Fe–O_1 distance increases from 1.85 Å (Compound 0) to 3.75 Å (IC1). The subsequent proton transfer from Glu366 to CH_2O-Thr again needs much activation (TS2 at 58.3 kcal/mol, thus 20.3 kcal/mol above IC1). The product (ferric resting state and hydrogen peroxide) lies 25.2 kcal/mol above Compound 0.

Asp251 channel

Figure 8 presents the optimized geometries. The O–Fe bond cleavage with formation of hydrogen peroxide again occurs in the first step, which has a very high barrier (TS1 at 52.9 kcal/mol). The intermediate (IC1 at 47.6 kcal/mol) contains hydrogen peroxide and a CH_2O-Thr radical (spin density of −0.90). The transition state for proton transfer from Asp251 to CH_2O-Thr (TS2 at 54.9 kcal/mol) lies 7.3 kcal/mol above IC1. The overall reaction is endothermic by 26.4 kcal/mol.

Discussion and conclusions

In this work, the coupling and uncoupling reactions in the Thr252MeO-Thr mutant of cytochrome P450cam were investigated for two proton delivery channels (Glu366 and Asp251) and four possible mechanisms by means of

QM/MM calculations. It is obvious from the QM/MM results that the uncoupling reaction (formation of the ferric resting state and hydrogen peroxide) is strongly disfavored. Regardless of mechanistic details, it suffers from high endothermicities of 25–30 kcal/mol and extremely high overall activation energies of 55–61 kcal/mol. We note in this context that our previous QM/MM study [37] of the uncoupling reaction in the wild-type enzyme yielded a barrier of 27 kcal/mol in the Asp251 channel, with the required proton being provided via the Asp251–Wat901–Thr252 network (mechanism similar to mechanism IV). Such proton delivery is expected to be more facile in the wild-type enzyme than the corresponding process in the Thr252MeO-Thr mutant since the hydroxyl group in Thr is a much better proton donor than the methyl group in MeO-Thr, and it is thus not surprising that the uncoupling reaction requires more activation in the Thr252MeO-Thr mutant.

In our recent QM/MM study [40], we discovered that the barrier for uncoupling is dramatically reduced in the Thr252X mutants (X = Val, Ala, Gly) when an extra water molecule enters the Asp251 channel and becomes part of a well-connected hydrogen-bonding network that provides a good proton delivery pathway. In these mutants, the extra water molecule remains present in 2-ns MD simulations, whereas it escapes from the channel for X = Ser and X = Thr. Also in the Thr252MeO-Thr mutant, the stability of an additional water molecule was tested by means of 2-ns MD simulations. It was observed that the additional water molecule escapes from the active site in both channels in the course of the MD simulations (Figs. S2, S3). This is in agreement with our previous findings for the wild-type enzyme, since MeO-Thr is sterically more demanding than Thr.

According to the present QM/MM results for the Thr252MeO-Thr mutant, the coupling reaction (mechanisms I and II) is endothermic by about 8 kcal/mol and requires an overall activation of 18–23 kcal/mol, depending on the channel and mechanism. It thus seems feasible and is clearly preferred over the uncoupling reaction. For both the coupling reaction and the uncoupling reaction, the highest point on the computed QM/MM energy profiles corresponds to hydrogen abstraction by OH and OOH species that are present in the intermediates formed. It is well known, e.g., from QM studies on small model systems [58], that such reactions are intrinsically more facile and more exothermic with OH than with OOH. For example, at the UB3LYP/6-31+G* level, the barrier (reaction energy) for hydrogen abstraction from ethyl methyl ether is 2.1 (−15.0) kcal/mol for OH and 12.6 (10.0) kcal/mol for OOH (Fig. S1). These intrinsic preferences are reflected in the QM/MM energies (Tables 1, 2).

Experimentally, the Thr252MeO-Thr mutant gives 100% coupling reaction and no uncoupling reaction [25], consistent with our QM/MM results. Furthermore, the observed rate constant for the formation of 5-*exo*-hydroxycamphor is one third of that of the wild-type enzyme [25]. This is in qualitative agreement with the QM/MM finding that the rate-limiting barriers for the coupling reaction are higher in the Thr252MeO-Thr mutant than in the wild-type enzyme. There are two caveats, however; first, it is not certain that the differences in the observed rate constants are actually due to different rates of Compound I formation; second, a factor 3 in the rate constant translates to a rather small difference in free-energy barriers of 0.7 kcal/mol (much smaller than the differences of 4–9 kcal/mol in the rate-limiting QM/MM barriers for the wild-type enzyme and the Thr252MeO-Thr mutant).

We finally address the question of the preferred coupling mechanism in the Thr252MeO-Thr mutant. At face value, the rate-limiting barriers are somewhat lower in the Glu366 channel than in the Asp251 channel (Table 1). One should keep in mind, however, that the Asp251 channel is in contact with bulk water, so it should be rather facile to reprotonate Asp251 after each coupling reaction that involves proton transfer in the Asp251 channel. This is not true for Glu366, which resides in a hydrophobic pocket and is thus difficult to reprotonate. Protonation via the Asp251 channel may thus actually be a more realistic scenario for the coupling reaction, as in the case of the wild-type enzyme. In this scenario, the barrier for the initial homolytic cleavage is predicted to rise from 14.3 kcal/mol in the wild-type enzyme to 18.6 kcal/mol in the mutant. This increase in activation energy can be rationalized by an analysis of the hydrogen-bonding network. In the wild-type enzyme, the OH radical is stabilized in the transition state by hydrogen bonds to Thr252 and the FeO unit with distances of 1.64 and 2.06 Å, respectively [37]. Stabilization is less effective in the Thr252MeO-Thr mutant, where these distances increase to 2.06 and 2.22 Å, respectively (mechanism I in the Asp251 channel). The subsequent protonation, with concomitant electron transfer from the heme, is essentially downhill in the wild-type enzyme and requires some activation in the mutant (Table 1). This is not surprising, and gas-phase QM modeling of this process indeed confirms the qualitative expectation that the hydroxyl group in Thr252 is a better proton donor than the methoxy group in MeO-Thr252 (see Sect. 9 in the electronic supplementary material). In the mutant enzyme, this process is split into two steps (hydrogen transfer from the methoxy group to OH followed by a simultaneous proton and electron transfer in the Asp251 channel, see mechanism I) which make it energetically feasible, through the stabilization of the resulting intermediates by strong hydrogen bonds to two water molecules. Regardless of their limited quantitative accuracy, the present QM/MM results thus raise the possibility that residue 252 may play

an active role in the proton delivery mechanism both for the wild-type enzyme and for the Thr252MeO-Thr mutant, whereas previous interpretations of the experimental data view this residue mainly as a structural factor for coordinating water molecules that deliver protons to the FeOOH unit [25].

We end with a cautionary note. The favored mechanism I in the Asp251 channel involves an incipient OH radical in the initial intermediate (IC1) that might be expected to undergo competing side reactions such as attack at the *meso* position of the porphyrin to affect heme degradation or demethylation of the methoxy group to regenerate Thr252. We note again in this context that the initially formed OH species is stabilized by hydrogen-bonding interactions with surrounding partners which lead to reduced OH spin density and hence presumably also to lower radical reactivity, in analogy to the situation in the wild-type enzyme [37]. A more reliable assessment will require QM/MM studies of the competing side reactions which are beyond the scope of this article.

References

1. Ortiz de Montellano PR (1998) Cytochrome P450: structure, mechanism and biochemistry. Plenum Press, New York
2. Ortiz de Montellano PR (1986) In: Ortiz de Montellano PR (ed) Cytochrome P450: structure, mechanism and biochemistry. Plenum Press, New York, pp 216–271
3. Dawson JH, Sono M (1987) Chem Rev 87:1255–1276
4. Sono M, Roach MP, Coulter ED, Dawson JH (1996) Chem Rev 96:2841–2888
5. Woggon WD (1997) Top Curr Chem 184:39–96
6. Ortiz de Montellano PR, De Voss JJ (2002) Nat Prod Rep 19:471–499
7. Watanabe Y (2001) J Biol Inorg Chem 8:846–856
8. Denisov IG, Makris TM, Sligar SG, Schlichting I (2005) Chem Rev 105:2253–2278
9. Schröder D, Schwarz H, Shaik S (2000) Struct Bond 97:91–123
10. Groves JT (2000) J Porphyrins Phthalocyanines 4:350–352
11. Meunier B (2000) J Porphyrins Phthalocyanines 4:353
12. Woggon WD, Wagenknecht HA, Claude C (2001) J Inorg Biochem 83:289–300
13. Costas M, Mehn MP, Jensen MP, Que L Jr (2004) Chem Rev 104:939–986
14. Rohde JU, In JH, Lim MH, Brennessel WW, Bukowski MR, Stubna A, Münck E, Nam W, Que L Jr (2003) Science 299:1037–1039
15. Imai M, Shimada H, Watanabe Y, Matsuhima-Hibiya Y, Makino R, Koga H, Horiuchi T, Ishimura Y (1989) Proc Natl Acad Sci USA 86:7823–7827
16. Martinis SA, Atkins WM, Stayton PS, Sligar SG (1989) J Am Chem Soc 111:9252–9253
17. Gerber NC, Sligar SG (1994) J Biol Chem 269:4260–4266
18. Gerber NC, Sligar SG (1992) J Am Chem Soc 114:8742–8743
19. Loew GH, Rohmer MM (1980) J Am Chem Soc 102:3655–3657
20. Poulos TL, Anthony-Cahill SJ, Griffith MC, Schultz PG (1989) Science 244:182–188
21. Harris DL, Loew GH (1994) J Am Chem Soc 116:11671–11674
22. Makris TM, Denisov IG, Sligar SG (2003) Biochem Soc Trans 31:516–519
23. Davydov R, Kappl R, Huttermann J, Peterson JA (1991) FEBS Lett 295:113–115
24. Davydov R, Macdonald IDG, Makris TM, Sligar SG, Hoffman BM (1999) J Am Chem Soc 121:10654–10655
25. Kimata Y, Shimada H, Hirose T, Ishimura Y (1995) Biochem Biophys Res Commun 208:96–102
26. Newcomb M, Toy PH (2000) Acc Chem Res 33:449–455
27. Vaz ADN, McGinnity DF, Coon MJ (1998) Proc Natl Acad Sci USA 95:3555–3560
28. Harris DL, Loew GH (1998) J Am Chem Soc 120:8941–8948
29. Guallar V, Friesner RA (2004) J Am Chem Soc 126:8501–8508
30. Kamachi T, Shiota Y, Ohta T, Yoshizawa K (2003) Bull Chem Soc Jpn 76:721–732
31. Taraphder S, Hummer G (2003) J Am Chem Soc 125:3931–3940
32. Kumar D, Hirao H, de Visser SP, Zheng J, Wang D, Thiel W, Shaik S (2005) J Phys Chem B 109:19946–19951
33. Shaik S, Kumar D, de Visser SP, Altun A, Thiel W (2005) Chem Rev 105:2279–2328
34. Ogliaro F, de Visser SP, Cohen S, Sharma PK, Shaik S (2002) J Am Chem Soc 124:2806–2817
35. Derat E, Shaik S (2006) J Phys Chem B 110:10526–10533
36. Sen K, Hackett JC (2009) J Phys Chem B 113:8170–8182
37. Zheng J, Wang D, Thiel W, Shaik S (2006) J Am Chem Soc 128:13204–13215
38. Chen H, Moreau Y, Derat E, Shaik S (2008) J Am Chem Soc 130:1953–1965
39. Chen H, Hirao H, Derat E, Schlichting I, Shaik S (2008) J Phys Chem B 112:9490–9500
40. Altarsha M, Benighaus T, Kumar D, Thiel W (2009) J Am Chem Soc 131:4755–4763
41. Schöneboom JC, Lin H, Reuter N, Thiel W, Cohen S, Ogliaro F, Shaik S (2002) J Am Chem Soc 124:8142–8151
42. Schöneboom JC, Thiel W (2004) J Phys Chem B 108:7468–7478
43. Altun A, Thiel W (2005) J Phys Chem B 109:1268–1280
44. Vidakovic M, Sligar SG, Li H, Poulos TL (1998) Biochemistry 37:9211–9219
45. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926–935
46. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M (1998) J Phys Chem B 102:3586–3616
47. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comput Chem 4:187–217
48. Becke AD (1993) J Chem Phys 98:5648–5652
49. Hay PJ, Wadt WR (1985) J Chem Phys 82:299–310
50. Hehre WJ, Ditchfield K, Pople JA (1972) J Chem Phys 56:2257–2261
51. Schäfer A, Horn H, Ahlrichs R (1992) J Chem Phys 97:2571–2577
52. Schäfer A, Huber C, Ahlrichs R (1994) J Chem Phys 100:5829–5835
53. Bakowies D, Thiel W (1996) J Phys Chem 100:10580–10594
54. Ahlrichs R, Bär M, Häser M, Horn H, Kölmel C (1989) Chem Phys Lett 162:165–169
55. Smith W, Forester TR (1996) J Mol Graph 14:136–141

56. Sherwood P, De Vries AH, Guest MF, Schreckenbach G, Catlow CRA, French SA, Sokol AA, Bromley ST, Thiel W, Turner AJ, Billeter S, Terstegen F, Thiel S, Kendrick J, Rogers SC, Casci J, Watson M, King F, Karlsen E, Sjøvoll M, Fahmi A, Schäfer A, Lennartz C (2003) J Mol Struct 632:1–28
57. Billeter SR, Turner AJ, Thiel W (2000) Phys Chem Chem Phys 2:2177–2186
58. Siegbahn PEM, Blomberg MRA, Crabtree RH (1997) Theor Chem Acc 97:289–300