LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

UCRL-TR-219003

Understanding the Mechanism of Human P450 CYP1A2 Using Coupled Quantum-Classical Simulations in a Dynamical Environment

E. W. Draeger, B. Bennion, F. Gygi, F. Lightstone

February 15, 2006

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

Understanding the Mechanism of Human P450 CYP1A2 Using Coupled Quantum-Classical Simulations in a Dynamical Environment

Principal Investigator:
Erik Draeger, Computation
Co-investigators:
Brian Bennion, Biosciences Directorate
François Gygi, Computation
Felice Lightstone, Biosciences Directorate

Abstract

The reaction mechanism of the human P450 CYP1A2 enzyme plays a fundamental role in understanding the effects of environmental carcinogens and mutagens on humans. Despite extensive experimental research on this enzyme system, key questions regarding its catalytic cycle and oxygen activation mechanism remain unanswered. In order to elucidate the reaction mechanism in human P450, new computational methods are needed to accurately represent this system. To enable us to perform computational simulations of unprecedented accuracy on these systems, we developed a dynamic quantum-classical (QM/MM) hybrid method, in which *ab initio* molecular dynamics are coupled with classical molecular mechanics. This will provide the accuracy needed to address such a complex, large biological system in a fully dynamic environment. We also present detailed calculations of the P450 active site, including the relative charge transfer between iron porphine and tetraphenyl porphyrin.

Biological Background

The cytochrome P450 superfamily of heme-thiolate enzymes are an ancient gene family that is found in almost all living organisms. The P450 enzymes are of practical interest because they are responsible for the biosynthesis of sex hormones, muscle-relaxing, anti-inflammatory and antihypertensive compounds, as well as metabolic conversion of both drugs and environmental xenobiotics. Isoforms such as CYP2D6, estimated to metabolize ~30% of known pharmaceutical products (Lewis, 1996), are of primary interest to drug discovery and development groups. Likewise, P450 isoforms such as CYP1A2 that oxidize environmental chemicals are of interest to those engaged in the study of environmental mutagens and carcinogens.

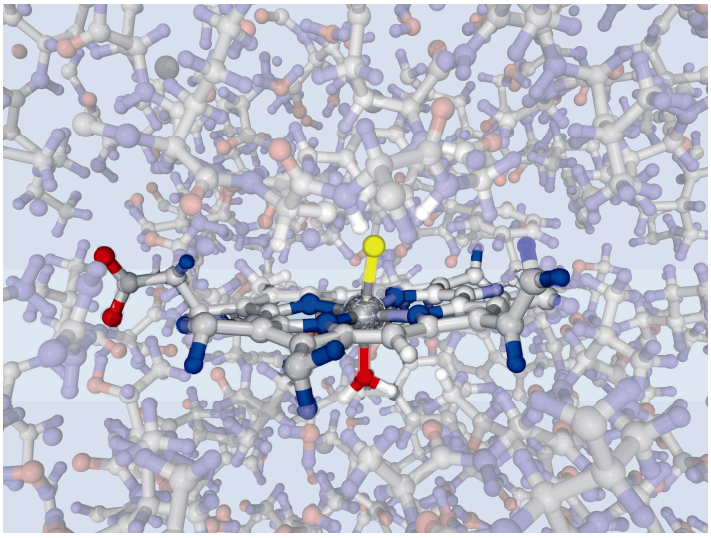


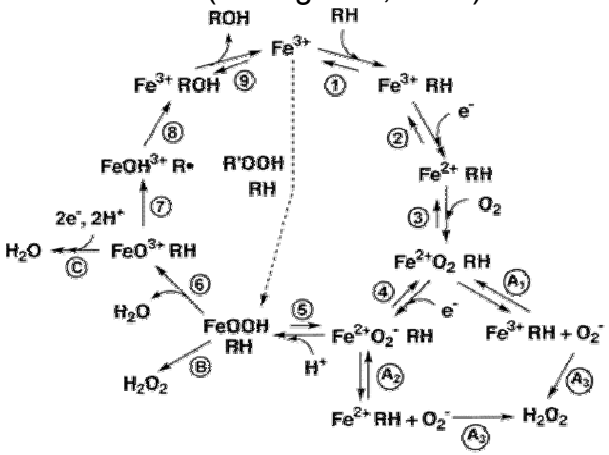
Figure 1: Active site of P450 CYP1A2 enzyme (quantum region) surrounded by protein (classical region).

Although this enzyme system has been extensively studied, there are still many questions about its catalytic cycle and oxygen activation mechanism which remain to be answered (Guengerich, 2001). Furthermore, because most P450 isoforms possess broad substrate specificities, the mechanism is specific to the reaction conditions, substrate and type of P450 involved. Careful computational study of catalytic mechanisms and enzyme kinetics can elucidate species which are difficult to capture by experiment but requires the accuracy of *ab initio* methods. Unfortunately, the application of *ab initio* approaches to these systems is computationally prohibitive; the system (i.e. 50,000 atoms) far exceeds the limits (i.e. 500 atoms) of state-of-the-art fully quantum mechanical computational techniques. Developing a quantum-classical hybrid method, in which *ab initio* molecular dynamics is coupled with classical molecular dynamics, would allow for the study of the human P450 enzyme oxygen activation mechanism in a fully dynamic environment including both the protein and surrounding solution, with *ab initio* accuracy at the active site (see Fig. 1).

The P450 superfamily of enzymes is composed of more than 300 enzymes and is essential for all living organisms (Lewis, 1996). In the past, many studies have been performed on prokaryotic P450s due to the fact that they are water soluble, as compared to the membrane bound mammalian P450s (Guallar et al., 2002;

Schoneboom et al., 2002). This has led to the crystallization of several prokaryotic P450s. In 2000, Williams et al were the first to publish a hybrid rabbit P450 CYP2C5 enzyme (Williams et al., 2000). This crystal structure has allowed us to use a mammalian template to build a homology model of human P450 CYP1A2.

Two of the most important reactions catalyzed by P450 enzymes is the oxidation of C—H to an alcohol, i.e. replacing the H with OH, and the dealkylaton of N, i.e. changing N—CH_3 to N—H. These reactions are only achieved by very few electrophilic reagents in the laboratory, while P450s can oxidize nonactivated C—H (or N—C) bonds with high stereoselectivity at only 37°C. Biologically, these reactions principally rely on transition-metal containing enzymes with complex catalytic cycles. The general P450 catalytic cycle is shown in Scheme 1 (Guengerich, 2001).



A summary of the proposed catalytic cycle is as follows:

1. The P450 iron starts in the ferric state and then the substrate binds the protein on the distal side of the iron porphyrin. This step may involve a change in the spin state of the iron. What is known is that the water which is coordinated to the ferric state disassociates from the iron when the substrate binds. Therefore steps 1 and 2 can be interchanged.
2. The ferrous P450 binds O_2 (step 3). This complex is highly unstable and can generate ferric iron and superoxide anion, $O_2^{\bullet-}$.
3. A second electron enters the system (step 4).
4. A proton is abstracted from either the surrounding protein or water environment to form a peroxide coordinated to the iron (step 5).
5. The O—O bond is cleaved while abstracting a second proton generating water and a complex shown as FeO^{3+} (step 6). This complex is highly controversial is its exact state and is comparable to the peroxidase compound I ($Fe^{4+}O_2$-porph$^{\bullet+}$). There have been numerous experimental and theoretical studies trying to determine the exact electronic state of this iron porphyrin complex. It is this complex that can be used to rationalize most of the P450 reactions, although there are alternate mechanisms.
6. Once the FeO^{3+} complex is formed, the radical immediately abstracts a hydrogen from the substrate, creating a carbon radical (step 7).
7. The carbon radical substrate immediately abstracts the hydroxide from the iron, completing the hydroxylation of the substrate and substrate leaves (steps 8 and 9).

One should note that Scheme 1 does not make any provisions for important conformational changes which are necessary to proceed to the next step in the catalytic cycle.

Computational Background

In 1985, a major break-through was made when Car and Parrinello (Car and Parrinello, 1985; Galli and Parrinello, 1991) developed a new technique based on DFT (Kohn and Sham, 1965) by treating electronic degrees of freedom at the same time as the nuclear equations of motion. Since the method employs QM theory to describe the entire system, it is often referred to as first principles molecular dynamics (FPMD). In the typical implementation of FPMD, only the chemically active valence electrons are explicitly described with an expansion in a plane-wave basis, while the chemically inert core electrons are represented by pseudopotentials (Yin and Cohen, 1982; Galli and Pasquarello, 1993). Since the pseudopotentials are transferable by design, this method does not require reparameterization when new systems are studied. In addition, the use of a plane wave basis set naturally lends itself to the application of periodic boundary conditions, so the method is well suited for modeling systems in the condensed phase.

The first applications of FPMD simulations were limited to small systems such as silicon (Car and Parrinello, 1985; Stich et al., 1989). As these methods have been continuously improved upon, and advanced computational resources have become available, such as the DOE teraflop scale supercomputers, it is now possible to investigate small biochemical systems containing several hundreds of atoms for picosecond timescales (Carloni and Alber, 1998; Pantano et al., 2000; Rovira and Parrinello, 2000). As the number of systems that have been investigated with this new approach increases, it is becoming clear that the increased computational expense is repaid in the form of extremely accurate structural and dynamical properties. In particular, such methods potentially allow for very accurate dynamical simulations of chemical phenomena including enzyme-catalyzed reactions.

QM/MM Method Description

Systems like P450 where a small chemically-active region is surrounded by a large biological environment are ideally suited to quantum-classical hybrid (QM/MM) methods. By dynamically coupling a first-principles molecular dynamics simulation to a classical molecular dynamics simulation, one can accurately reproduce the chemistry at the active site while still reproducing the full biological environment. Making such a coupling accurate and stable is quite challenging, particularly in the case where the first-principles calculation is performed within a plane wave basis. Several groups (Eichinger *et al.*, 1999, Laio *et al.,* 2002*)* have successfully demonstrated that QM/MM calculations involving a plane wave first-principles code can be done accurately and efficiently, by carefully defining the interactions between the systems so as to avoid introducing instabilities or systematic errors.

The total energy of a coupled QM/MM system can be written as:

$$E_{tot} = E_{QM}(\rho, R_{QM}) + E_{MM}(R_{MM}) + E_{QM/MM}(\rho, R_{QM}, R_{MM})$$

where ρ is the electronic density distribution of the quantum system, R_{QM} contains the coordinates of the quantum ions, and R_{MM} contains the coordinates of the classical atoms. E_{QM} and E_{MM} contain all contributions to the total energy which are functions only of the individual components of the quantum and classical systems, respectively. For nonbonded systems, i.e. systems in which there are no covalent bonds between the quantum and classical portions, these energies are simply the total energies of each system in isolation. When such covalent bonds exist, however, these individual energy terms contain artifacts from the bonds (or lack thereof) between the two systems, and thus do not have an easily identified physical meaning on their own.

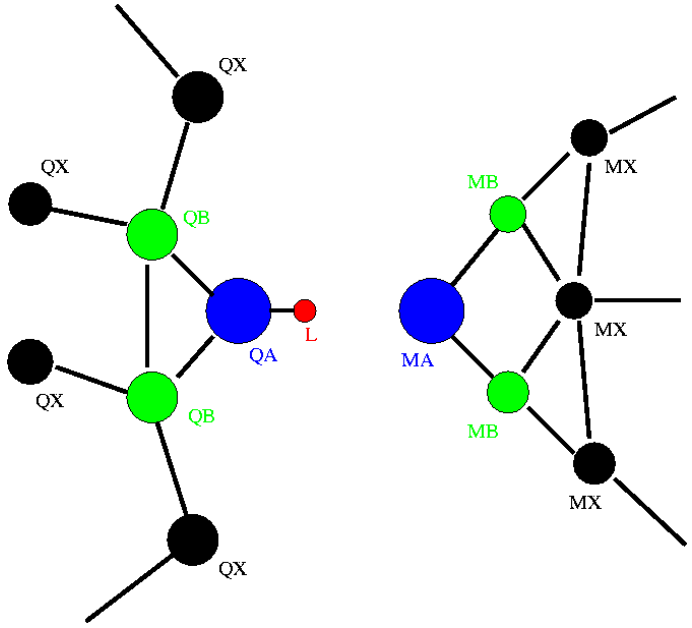


Figure 2: Cartoon QM/MM system, showing grouping of atoms into those involved in bonds (QA, MA), those bonded to those involved in bonds (QB, MB) and all other atoms (QX, MX).

$E_{QM/MM}$ contains all interactions which are functions of variables of both systems. To clearly define these interactions, it is necessary to separate the ionic coordinates of each system into subsets:

$$R_{QM} = \{R_{QA}, R_{QB}, R_{QX}\}$$
$$R_{MM} = \{R_{MA}, R_{MB}, R_{MX}\}.$$

The distinct subsets of R_{QM} are defined as follows: R_{QA} contain the coordinates of QM atoms which are bonded to MM atoms, R_{QB} contain the coordinates of QM atoms which are bonded to atoms in R_{QA}, and R_{QX} contain all remaining atoms in R_{QM} which are members of neither R_{QA} or R_{QB}. The subsets of R_{MM} are defined similarly (see Fig.

1). Thus, by definition QM/MM bonds exist only between R_{QA} and R_{MA}, and no bonds can exist between R_{QA} and R_{QX} atoms, or between R_{MA} and R_{MX} atoms.

Using these definitions, $E_{QM/MM}$ can be separated into several distinct terms:

$$E_{QM/MM} = E_{coul}^{ion\text{-}ion}(R_{QX};R_{MX}) + E_{coul}^{ion\text{-}rho}(\rho;R_{MX}) + E_{vdw}(R_{QX};R_{MX}) + E_{bond}(R_{QA},R_{QB};R_{MA},R_{MB})$$

The first three terms contain the non-bonded Coulomb and van der Waals interactions, while the last term contains all interactions needed to simulate covalent bonds between the two systems. The calculation of all four terms is described below:

a. Calculating E_{coul}

The Coulomb interaction energy is given by:

$$E_{coul}^{ion\text{-}ion}(R_{QX};R_{MX}) = \sum_{i\in R_{QX}} \sum_{j\in R_{MX}} v(|\mathbf{r}_i - \mathbf{r}_j|)$$

where $v(|\mathbf{r}_i - \mathbf{r}_j|)$ is the Coulomb pair interaction, which for two point charges is simply

$$v(|\mathbf{r}_i - \mathbf{r}_j|) = \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|}$$

To prevent singularities as $|\mathbf{r}_i - \mathbf{r}_j| \to 0$, we replace this with a screened pair interaction:

$$v(|\mathbf{r}_i - \mathbf{r}_j|) = \frac{q_i q_j}{|\mathbf{r}_i - \mathbf{r}_j|} \mathrm{erfc}\left(\frac{|\mathbf{r}_i - \mathbf{r}_j|}{r_c}\right)$$

This is equivalent to replacing one of the point charges with a Gaussian charge distribution of width r_c. The appropriate value of r_c may vary for different pair interactions. For now, we assign values of r_c by MM species. While this assumes that a single value is appropriate for all QM species which interact with each MM atom type, this is expected to be sufficiently accurate; the authors of Eichinger *et al*. found it acceptable to use a fixed value of r_c =0.8 Å for all QM-MM ion-ion Coulomb interactions.

The interaction between the electron density and the classical atoms is handled similarly. A screened Coulomb interaction, in which the classical atoms are represented by Gaussian charge distribution instead of point charges, is needed to prevent the electron density from unphysically localizing on the classical atoms. Thus, the Coulomb interaction between the classical atoms and the electron density can be written as

$$E_{coul}^{ion\text{-}rho}(\rho;R_{MX}) = \sum_{j\in R_{MX}} \int d\mathbf{r}\, \rho(\mathbf{r}) \frac{q_j}{|\mathbf{r} - \mathbf{r}_j|} \mathrm{erfc}\left(\frac{|\mathbf{r} - \mathbf{r}_j|}{r_c}\right)$$

The computational cost to calculate this term is proportional to the number of classical atoms times the number of real-space grid points in the \emph{ab initio}

calculation, which will quickly become prohibitive for large systems and/or large plane wave energy cutoffs. Further approximations to this term, in the form of particle-mesh Ewald sums, are therefore necessary to reduce the cost of this term.

b. Calculating E_{vdw}

Because classical mechanics does not implicitly include the effects of Pauli exclusion, an empirical van der Waals interaction between the classical atoms and quantum ions must be added to prevent e.g. oppositely charged atoms from unphysically collapsing onto each other.

$$E_{vdw}(R_{QX};R_{MX}) = \sum_{i \in R_{QX}} \sum_{j \in R_{MX}} v_{vdw}(|\mathbf{r}_i - \mathbf{r}_j|)$$

where v_{vdw} is typically a Lennard-Jones interaction:

$$v_{vdw}(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

It should be noted that some classical codes, including NAMD, use a slightly different form for this interaction:

$$v_{vdw}(r) = \varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - 2\left(\frac{\sigma}{r}\right)^{6}\right]$$

The values of σ and ε are specified by the user for each QM/MM pair interaction. The choice of Lennard-Jones interaction between the above two forms can be specified with the VDWFORM keyword.

c. Calculating E_{bond}

With the exception of small molecules and nanostructures in solution, covalent bonding between the quantum and classical regions will be unavoidable in most biological systems. Therefore, a general QM/MM method must realistically reproduce the effects of chemical bonding on both the quantum and classical atoms involved in the bond. On the classical side, this involves the application of empirical two-, three- and/or higher-body terms. On the quantum side, dangling bonds must be saturated so that the electronic structure of the rest of the system realistically reproduces a bonded fragment.

This has traditionally been accomplished with *link atoms*, constrained atoms in the quantum system which mediate bonded interactions. The simplest choice of link atom is hydrogen, although more specialized one-electron pseudopotentials have been developed for biological systems. To prevent additional degrees of freedom from being introduced into the system, the position of the link atom is constrained along the vector between the quantum and classical atoms in the bond. Eichinger *et al.* developed the scaled-position link atom method (SPLAM) in which the third degree of freedom, the bond distance between the link atom and the quantum atom, is set to be a fraction of

the QM/MM bond distance. This allows one to use the *ab initio* bond between the quantum atom and the link atom to determine the stretch component of the force between the quantum and classical atoms, with some rescaling correction to account for the different bond types. We propose starting with a simpler approach, wherein the bond distance between the link atom and the quantum atom is fixed at a constant value, and the stretch force of the QM/MM bond is handled empirically. Considering the approximations inherent in QM/MM bonding, we believe this should be sufficient to realistically model covalent bond forces between the two systems without introducing significant additional error.

The total energy contribution from the QM/MM bonds is:

$$E_{bond} = \sum_{k \in \text{QM/MM bonds}} E_{stretch}(\mathbf{r}^k_{QA}, \mathbf{r}^k_{MA}) + E_{angle}(\mathbf{r}^k_{QA}, \mathbf{r}^k_{QB}, \mathbf{r}^k_{MA}) + E_{angle}(\mathbf{r}^k_{QA}, \mathbf{r}^k_{MA}, \mathbf{r}^k_{MB}) + E_{torsion}(\mathbf{r}^k_{QA}, \mathbf{r}^k_{QB}, \mathbf{r}^k_{MA}, \mathbf{r}^k_{MB}) - \Delta E_{dipole}(\mathbf{r}^k_{QA}, \mathbf{r}^k_{L}, \mathbf{r}^k_{MA}) - \Delta E_{vdw}(\mathbf{r}^k_{L}, R_{QM})$$

The first four terms take the same form as bonded interactions for these species within a fully classical simulation:

$$E_{stretch}(\mathbf{r}_1, \mathbf{r}_2) = \frac{1}{2} k_{12} (|\mathbf{r}_1 - \mathbf{r}_2| - r^0_{12})^2$$

$$E_{angle}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3) = \frac{1}{2} k_{123} (\theta_{123} - \theta^0_{123})^2$$

$$E_{torsion}(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{r}_4) = \frac{1}{2} k_{1234} [1 - \cos(\chi_{1234} - \chi^0_{1234})]$$

where the parameters k, r^0, θ^0, χ^0 are given by the same empirical values which would be employed in a fully classical simulation of the same system, i.e. one in which the quantum atoms were treated classically.

The last two terms in the definition of E_{bond} are correction terms designed to remove the unphysical contributions from the link atom to the total energy. The dipole energy term accounts for the unphysical effects of the artificial bond between the QA atom and the link atom being polar, when the QM/MM bond itself is not. The van der Waals correction term is designed to approximate the nonbonded interactions between the link atom and the rest of the quantum system, using the same empirical form used in the classical calculation, and remove it.

Once all the energy terms have been defined, the ionic forces can be calculated from the derivatives with respect to the ionic coordinates and added to the equations of motion of each system, namely:

$$F_{QM} = -\nabla_{QM} E_{tot} = -\nabla_{QM} E_{QM} - \nabla_{QM} E_{QM/MM}$$

and

$$F_{MM} = -\nabla_{MM} E_{tot}$$
$$= -\nabla_{MM} E_{MM} - \nabla_{MM} E_{QM/MM}$$

$-\nabla_{QM} E_{QM}$ and $-\nabla_{MM} E_{MM}$ are already calculated within the quantum and classical codes, respectively. Calculating the derivatives of $E_{QM/MM}$ with respect to the atomic positions is straightforward. The force on the electronic density of the quantum system must also be calculated. This is done by treating the Coulomb interaction with the classical point charges as an external potential in the self-consistent loop. This is given by a functional derivative of the interaction energy with respect to the charge density:

$$V_{ext}(\mathbf{r}) = \frac{\delta E_{coul}^{ion-rho}}{\delta \rho} = \sum_{j \in R_{MX}} \frac{q_j}{|\mathbf{r} - \mathbf{r}_j|} \mathrm{erfc}\left(\frac{|\mathbf{r} - \mathbf{r}_j|}{r_c} \right)$$

As mentioned, care must be taken to choose r_c appropriately in order to avoid unphysical charge localization.

Computational Implementation

Considering the significant amount of effort that has been invested in developing powerful, fully-featured codes in both the first-principles and classical molecular dynamics fields, it would be inefficient to attempt to create a QM/MM code by adding MM (or QM) functionality to an existing QM (or MM) code. A far better approach would be to couple two stand-alone QM and MM codes, ideally in such a way that new versions of either code could be plugged in without extensive modification. As such, our programming model centered around putting as much QM/MM-specific code as possible in a separate library, which would then interface with both individual QM and MM codes. Although both QM and MM codes must still be modified to interface with the QM/MM library and apply the QM/MM interactions to the respective systems, these modifications can be confined to the main iteration loop of each code.

To this end, we chose to couple the first-principles molecular dynamics code GP, developed here at Lawrence Livermore National Laboratory, with the classical molecular dynamics code NAMD, developed at the University of Illinois, Urbana-Champaign. Both fully-featured and massively parallel, NAMD was an ideal candidate for QM/MM. More importantly, NAMD was designed with a socket interface which allows it to exchange forces and positions in real-time with the visualization program VMD, an indication that the code's interface would be well-suited for use within a QM/MM framework.

Figure 3 shows the details of our QM/MM implementation. The shared library is dynamically linked to GP, allowing the user to substitute any library with a compatible interface (e.g. a QM/MM library which connects to a different MM code) at runtime. The shared library was written to connect to NAMD via its VMD socket interface. Modifications within NAMD were required to force the code to wait to receive information from the QM/MM library before every iteration, instead of simply polling the sockets and continuing to run, as is appropriate with interactive visualization. In addition, modifications were required to enforce synchronization between program threads. Interactive visualization takes place on a much coarser time scale than a single molecular dynamics iteration, and thus the socket thread was allowed to lag one

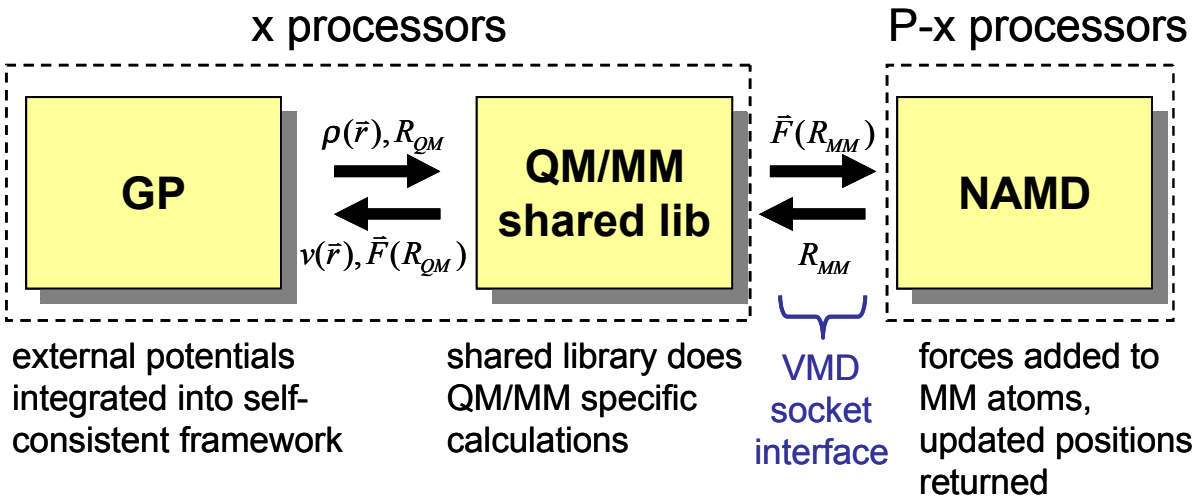
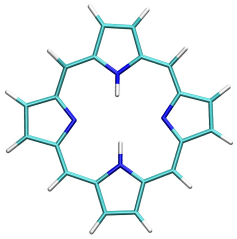


Figure 3: Schematic representation of QM/MM implementation.

step behind the main iteration thread. For visualization, this improves performance as both threads can run simultaneously. For QM/MM dynamics, however, this causes a small but constant energy leak as forces calculated from the positions at a given step are not applied until the subsequent one. Working with NAMD code author Jim Phillips, we were able to identify and resolve this subtle problem. Energy conservation of the full QM/MM system was found to hold within 10^{-8} eV/iteration for a small test system.

Results, Discussion, and Methods
Porphine Systems
Results

A quantum simulation method such as FPMD is required to accurately model the Fe^{2+}-porphyrin, which is the essential catalytic component of P450 CYP1A2. Before modeling the complete Fe^{2+}-heme, the porphine system (no Fe^{2+}) was modeled to simplify the calculations. We performed geometry optimizations (no molecular dynamics) on porphine and obtained structural properties which are in good agreement with previous quantum chemistry studies (Ghosh, 2000), see Figure 4. Before adding the Fe^{2+} to the porphine system, several iron pseudopotentials were thoroughly tested. We identified an iron pseudopotential that is well-suited for the Fe^{2+}-porphyrin system, see Figure 5. Calculations of Fe^{+2}-porphyrin in a variety of spin states using this potential are in excellent agreement with both experimental and previous theoretical results (Collman et al., 1975; Rovira et al., 1997).

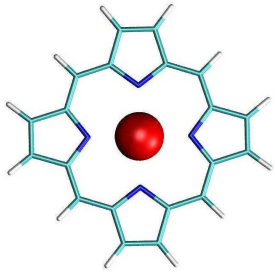


Method	*RHF/TZDP*	*MP2/SVP*	*LDF/DZP*	*NLDFT/TZDP*	*GP*
N-H (Å)	0.99	0.99	1.03	1.0195	1.01
N-C (Å)	1.36	1.36	1.36	1.3709	1.35
C-C (Å)	1.43	1.37	1.38	1.3928	1.40
H-N-C (°)	124.20	124.40	124.80	124.638	125.60
N-C-C (°)	124.90	125.50	125.30	125.607	125.20

Figure 4. Porphine molecule used in the GP calibration and convergence calculations (top). Selected porphine geometry parameters from other high level ab initio calculations (Ghosh, 2000) compared with our results from GP.

Methods

Our plane-wave DFT calculations used the local-density approximation (LDA) and Perdew-Burke-Ernzerhof (PBE) generalized-gradient-approximation exchange-correlation functionals. Specific parameters that were used during the geometry optimizations are shown below in Fig 6. Each system started with semi-empirically determined structures and an initial randomized wave function. Once the wave function was generated, it was updated after 100 steps of electronic minimization. The wave functions were minimized using preconditioned steepest descents with Anderson acceleration (PSDA). The electronic preconditioning was set conservatively to 5 and the Car-Parinello electron mass was set to 340 for all calculations described in this study.



	Previous Calculations		***Expt***	***GP***	
Fe spin	S=1/2	S=3/2	S=3/2	S=1/2	S=3/2
Fe-N (Å)	1.97	1.98	1.97	2.00	2.00
N-C (Å)	1.39	1.39	1.38	1.37	1.37
C-C (Å)	1.38	1.38	1.38	1.38	1.38
Fe-N-C (°)	127.80	127.60	127.3	127.25	127.35
N-C-C (°)	125.10	125.20	125.1	125.00	125.10

Figure 5. Iron-porphine molecule used in iron psuedopotential calibration calculations (top). The iron is shown as a red ball in the center. Structural geometry parameters from previous theoretical work (Rovira et al., 1997) and experimental work (Collman et al., 1975) are compared to parameters derived from our GP calculations.

```
singlet.sys
# exchange-correlation potential to use
set xc PBE
set netcharge -2
set spin ON
set delta_spin 0
# print out the coordinates every iprint iteration
set iprint 1
# planewave cutoff (size of basis set)
set ecut 90
# mass of electrons for Car-Parrinello dynamics
set emass 340
# wavefunction dynamics
set wf_dyn PSDA
set ecutprec 5.0
#randomize_wf
load singlet.wf
status
run 1 100
save singlet.wf
sysfile singlet.sys
```

Figure 6 A sample of Jeep input for singlet Fe(+2)-porphine wavefunction minimization.

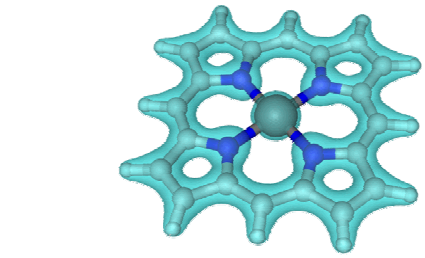
Table 1 Iron Porphine Energies

System	Spin	Final Energy (au)
Fe-Porphine	**Triplet**	**-281.77335685**
Fe(+2)-Porphine	**Singlet**	**-281.2996544**
Fe(+2)-Porphine	**Triplet**	**-281.30787780**
Fe(+3)-Porphine	**Triplet**	**-281.27172372**

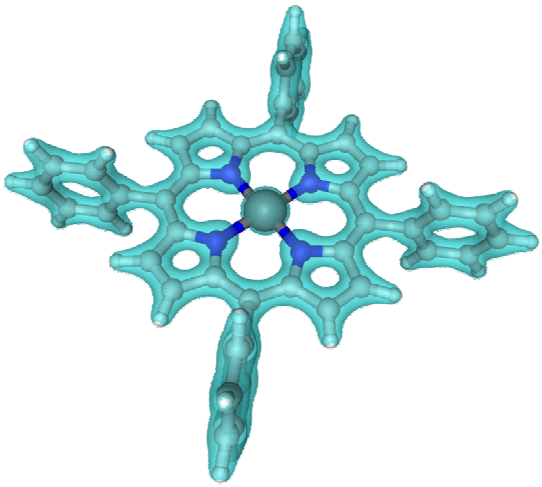
The Fe-porphine calculations were divided into three charge states each with an initial 2 spin states. In the case of charged Fe-porphine systems, several special-case settings were needed. Firstly, the +2 and +3 charge states required setting a net charge of -2 and -3 respectively, over the entire system. Secondly, spin and delta spin were set. Testing of the Fe pseudopotentials showed that the minimum ecut needed for proper convergence was 90 with a box size of 21 x 21 x 7 au. The geometry of each system was considered converged when atomic forces were less than 10^{-5} au. This system size required all the memory capacity of 1 frost node (16 cpus), calculated 1.11 steps/min, and converged in ~10K steps. Sixteen cpus of MCR were used for the geometry optimization of Fe(+3)-porphine which ran at 3.33 steps/min and converged in ~22k steps. Final energies are shown in Table 1. During our calculations it was discovered that the singlet state could not be modeled in the Fe(+3)-porphine system which has unoccupied states below the HOMO. Finally, time did not permit calculations of the quartet spin states of any charge species.

Iron Tetraphenyl Porphyrin Systems
Results

After we were confident that the heavy atom pseudopotentials could reproduce previous theoretical and experimental results we extended the calculations to larger systems. The current literature is divided in regards to the amount of charge the porphyrin group can accept and distribute (Guallar et al., 2003; Guallar and Friesner, 2004; Schoneboom et al., 2004). In addition, we needed to know what the minimum size of the QM region was before we started our QM/MM calculations. A smaller QM region would allow for easier segregation from the MM region and faster calculations, however, if the QM region is too small, the modeling of the active site chemistry might be wrong. Building on previous calculations of Fe-porphine, we added four phenyl groups to our original Fe-porphine molecule (TPP) (Fig. 7).



Fe-Porphine



Fe-TetraphenylPorphyrin

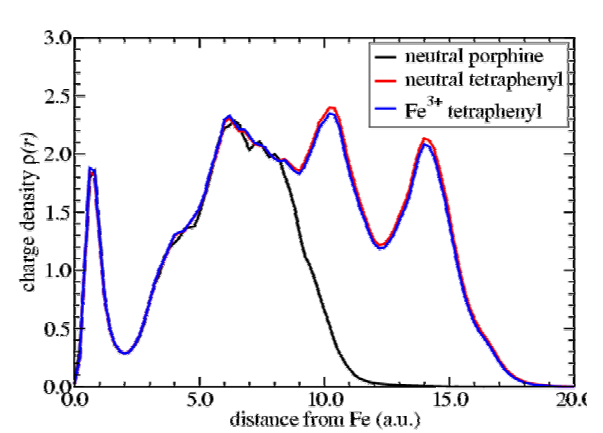
Figure 7 Electron-spin density iso-surfaces of Iron pophine and tetraphenylPorphyrin.

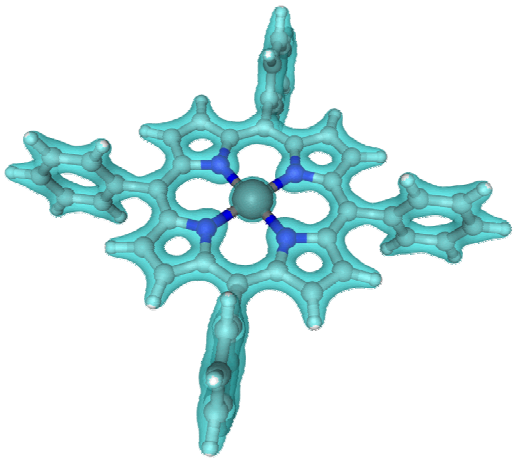
These additional substituents allowed us to determine how charge was distributed throughout the molecule. If significant spin density was shifted to the outer phenyl rings upon addition of one and two electrons, then this would be an indication that the QM region needed to be larger than the Fe-porphine studied previously. The following systems had the wave function minimized, geometry optimized and FPMD simulations started: Fe-TPP, Fe(+2)-TPP, Fe(+3)-TPP. As shown in Figure 8 charge density does not significantly differ within in the porphine core between neutral

porphine, Fe-TPP and Fe(+3)-TPP. However, significant differences exist when we compared Fe-TPP and Fe(+3)-TPP. Extra charge density was present on the phenyl rings in Fe-TPP compared to Fe(+3)-TPP which suggests that the porphine core is not adequate for a QM region. In addition, other increasingly polar moieties on porphyrin rings, such as the glutamic acids chains found in protoporphyrin IX, will indeed affect charge density and active site chemistry as electrons are added or removed within the QM region. Preliminary results from FPMD simulations of each system show that the molecules are stable. These simulations need to be continued for greater statistical sampling.

Methods

The TPP systems required over 1GB of system memory per node which excluded all LC machines but MCR. These systems required extra settings to aid convergence of the wave function and geometry. Firstly, nempty was set at 5 to allow extra states. Secondly, the Fermi temperature was set at 2000 and finally the preconditioning ecut was set to 20 instead of 5. The box size for each system was 38 x 18.5 x 38 au which included 77 atoms, 240 electrons (neutral system), 250 states (126 up and 124 down i.e. the triplet configuration). The wave function was minimized until





8 Fe tetraphenylPorphyrin with plots of normalized (top) and cumulative (bottom) charge density. Charge differences exist between neutral and Fe3+ system delocalized across the phenyl rings. Phenyl rings and other biologically relevant foliage are thus needed in QM region to accurately model different charge states.

the total energy was converged to 10^{-8} au, which cost~2600 steps at 1 step/hr. Next, each system was geometry optimized to the wave function using 1 ionic step to 100 electronic steps. Preconditioned steepest descents were used for the electronic minimization while blocked molecular dynamics was used to relax the nuclei. The time step was set to 10 au. Five nuclear with 20 electronic steps each required 6.5 hours. Convergence was attained in ~300 steps when the atomic forces were less than 10^{-4} au. The same convergence criteria were used in the Fe(+2)-TPP and Fe(+3)-TPP calculations. Final energies for each system are shown in Table 2.

Table 2 Iron-Tetraphenylporphyrin Geometry Optimized Energies

System	Spin	Final Energy (au)
Fe-TPP	Triplet	-427.79492292
Fe(+2)-TPP	Triplet	-427.46921675
Fe(+3)-TPP	Triplet	-427.23407251

Fe-TPP FPMD Simulations

Results

We started FPMD simulations for each charge state to understand what effects the nuclei dynamics would have on electronic spin/charge distribution. Unfortunately the simulations have not progressed enough to make substantial conclusions. However, preliminary data from the Fe-TPP simulation indicate stable energetics and structure. We also see indications of slight asymmetric ruffling of the porphryin as seen in the iron out-of-plane distance, bonds and angles (Table 3). These observations are compatible with experimental results (Collman et al., 1975) in that major ruffling was observed for the charged Fe-TPP molecules. These calculations need to be propagated further in all charge states for more meaningful comparison.

Table 3 Geometric Observables From Iron-TPP (Triplet) FPMD Simulations

	Out-of-Plane (Å)				
Fe	0.017				
	Bonds (Å)			Core-Size	
Fe-N1	Fe-N2	Fe-N3	Fe-N4	N1-N3	N2-N4
2.0172	2.0086	2.0070	2.0077	4.0162	4.0197
	Angles				
Fe-N1-C	Fe-N2-C	Fe-N3-C	Fe-N4-C		
127.372	127.353	127.213	127.103		

Methods

Final structures from each of the geometry optimizations were used start the FPMD simulations. Additional parameters were needed for the FPMD simulations; the temperature was set to 300 K and temperature coupling via the thermostat was set to 4000. Each MD calculation consisted of 20 steps of wave function minimization for each ionic step. The simulation time step was set to 10 au and structures saved every 100 electronic steps for later analysis. As of this writing 324 ionic steps have been calculated for Fe-TPP on MCR with 128 nodes in 360 hrs. FPMD simulations of Fe(+2)-TPP and Fe(+3)-TPP are currently underway.

Ligated Porphyrin Systems

In biological systems porphyrin molecules are non-covalently ligated by two chemical entities above and below the plane of the molecule. In addition, certain proteins also form covalent bonds with the porphyrin through peripheral methinyl goups. Proteins involved in electron transfer coordinate the porphyrin molecule (heme) with one or two electron donating amino-acid side-chains, histidine or cysteine for example. However, P450 enyzmes involved in oxidative metabolism are singly ligated by protein and contain water molecules that form the second interaction. Substrate molecules for these enzymes displace this water molecule thus beginning the cycle described in scheme 1.

Several groups have published theoretical models of ligated porphyrins (Harris et al., 1998; Rovira et al., 2001; Scherlis and Estrin, 2002; Guallar and Friesner, 2004; Schoneboom et al., 2004; Schoneboom and Thiel, 2004; Altun and Thiel, 2005; Hackett et al., 2005; Kumar et al., 2005). Therefore as a benchmark to our QM/MM development, we started calculations that would allow us to build on or explain inconsistencies in the current literature. Two models were built. The first model was an approximation of the hemoglobin oxygen binding site and the second model portrayed the P450 active site (Fig. 9). The ligating amino-acid side chains were approximated by

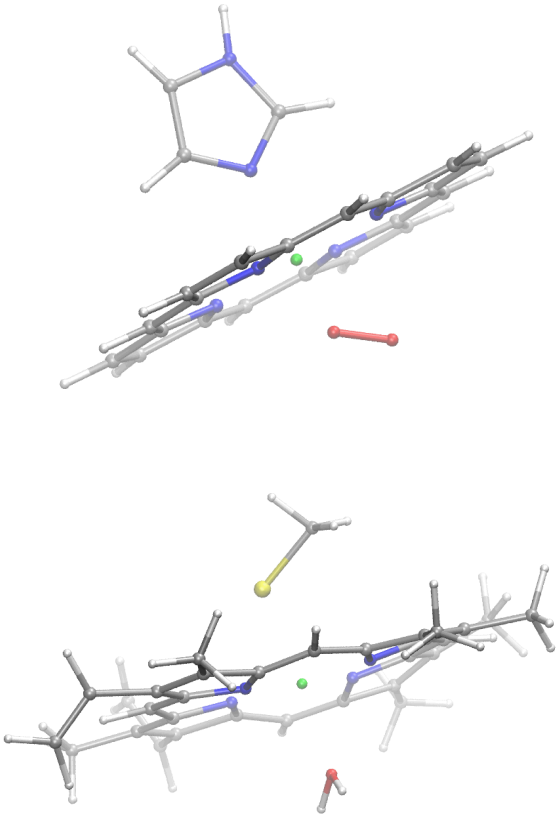


Figure 9 Ligated porphyrin models under study. Imidazole-porphine-oxygen (top) and water-diethyl-pophine-methylsufide (bottom).

imidazole (histidine) and methylsulfide (cysteine). The imidazole-porphine-oxygen model was calculated as a singlet and the water-porphine-sulfide model as a singlet and triplet. The later model is described as an open shell singlet by (Rovira et al., 1997), and (Harris et al., 1998) found that the energy difference between the ferrous singlet and triplet to be small, 0.8 kcal.

To our knowledge no simulations contradicting or confirmating these results have been published. Furthermore, experiment (Scherlis and Estrin, 2002) shows that this system should be a closed shell singlet, in contrast to Roviras' work which states that the ground state is an open shell singlet. These calculations are currently underway. The imidazo-porphine-water model has received 2200 and 10000 steps of wave function minimization. The water-porphine-sulfide model has had 2900 steps of minimization. Neither wave function has converged at the time of this writing.

CONCLUSIONS

During the course of this project, we coupled the GP first-principles molecular dynamics code to the NAMD classical molecular dynamics code to create a powerful new dynamic QM/MM capability for computational biology here at LLNL. Through a series of careful first-principles simulations, we demonstrated that atomic pseudo potentials for oxygen, carbon, hydrogen, and nitrogen were suitable for modeling of biological molecules. We also developed a stable iron pseudo potential which when coupled with the other organic atom pseudo potentials in the porphine molecule gave results that were in agreement with previously published theoretical and experimental data. Furthermore, we demonstrated that charge density distribution is in fact dependent on the size and substitution of the porphyrin macrocycle. This is important for determining the size and makeup of the QM region in our QM/MM hybrid simulations.

References

Altun A and Thiel W (2005) Combined quantum mechanical/molecular mechanical study on the pentacoordinated ferric and ferrous cytochrome P450(cam) complexes. *Journal of Physical Chemistry B* **109:**1268-1280.

Car R and Parrinello M (1985) Unified Approach for Molecular Dynamics and Density-Functional Theory. *Physical Review Letters* **55:**2471-2474.

Carloni P and Alber F (1998) Density-Functional Theory Investigations of Enzyme-Substrate Interactions. *Perspectives in Drug Discovery and Design* **9/11:**169-179.

Collman JP, Hoard JL, Kim N, Lang G and Reed CA (1975) Synthesis, Stereochemistry, and Structure-Related Properties of Alpha, Beta, Gamma, "Delta-Tetraphenylporphinatoiron(Ii). *Journal of the American Chemical Society* **97:**2676-2681.

Eichinger M, Tavan P, Hutter J, Parrinello M (1999) A Hybrid Method for Solutes in Complex Solvents: Density Functional Theory Combined with Empirical Force Fields. *Journal of Chemical Physics* **110**:10452-10467.

Galli G and Parrinello M (1991) Ab-Initio Molecular Dynamics: Principles and Practical Implementations, in: *Computer Simulation in Materials Science*, pp 283-304, Kluwer Academic Publishers, The Netherlands.

Galli G and Pasquarello A (1993) First Principles Molecular Dynamics, in: *Computer Simulation in Chemical Physics* (Allen MP and Tildesley DJ eds), pp 261-313, Kluwer, Dordrecht.

Ghosh A (2000) Quantum Chemical Studies of Molecular Structures and Protential Energy Surfaces, in: *The Porphyrin Handbook* (Kadish KM, Smith, Kevin M., Guilard, Roger ed), pp 1-75, Academic Press, San Diego.

Guallar V, Baik MH, Lippard SJ and Friesner RA (2003) Peripheral heme substituents control the hydrogen-atom abstraction chemistry in cytochromes P450. *Proceedings of the National Academy of Sciences of the United States of America* **100:**6998-7002.

Guallar V and Friesner RA (2004) Cytochrome P450CAM enzymatic catalysis cycle: A quantum mechanics molecular mechanics study. *Journal of the American Chemical Society* **126:**8501-8508.

Guallar V, Gherman BF, Miller WH, Lippard SJ and Friesner RA (2002) Dynamics of alkane hydroxylation at the non-heme diiron center in methane monooxygenase. *Journal of the American Chemical Society* **124:**3377-3384.

Guengerich FP (2001) Common and uncommon cytochrome P450 reactions related to metabolism and chemical toxicity. *Chemical Research in Toxicology* **14:**611-650.

Hackett JC, Brueggemeier RW and Hadad CM (2005) The final catalytic step of cytochrome P450 aromatase: A density functional theory study. *Journal of the American Chemical Society* **127:**5224-5237.

Harris D, Loew G and Waskell L (1998) Structure and spectra of ferrous dioxygen and reduced ferrous dioxygen model cytochrome P450. *Journal of the American Chemical Society* **120:**4308-4318.

Kohn W and Sham LJ (1965) Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review A* **140:**1133-1138.

Kumar D, Hirao H, de Visser SP, Zheng JJ, Wang DQ, Thiel W and Shaik S (2005) New features in the catalytic cycle of cytochrome P450 during the formation of compound I from compound 0. *Journal of Physical Chemistry B* **109:**19946-19951.

Laio A, VandeVondele J, Rothlisberger U (2002) A Hamiltonian Electrostatic Coupling Scheme for Hybrid Car-Parinello Molecular Dynamics Simulations. *Journal of Chemial Physics* **116**:6941-6947.

Lewis DFV (1996) *Cytochromes P540 Structure, Function, and Mechanism*. Taylor and Francis, Bristol, PA.

Pantano S, Alber F and Carloni P (2000) Proton Dynamics in an Enzyme Model Substrate: An Ab Initio Molecular Dynamics Study. *Journal of Molecular Structure (Theochem)* **530:**177-181.

Rovira C, Kunc K, Hutter J, Ballone P and Parrinello M (1997) Equilibrium geometries and electronic structure of iron-porphyrin complexes: A density functional study. *Journal of Physical Chemistry A* **101:**8914-8925.

Rovira C and Parrinello M (2000) First-principles molecular dynamics simulations of models for the myoglobin active center. *International Journal of Quantum Chemistry* **80:**1172-1180.

Rovira C, Schulze B, Eichinger M, Evanseck JD and Parrinello M (2001) Influence of the heme pocket conformation on the structure and vibrations of the Fe-CO bond in myoglobin: A QM/MM density functional study. *Biophysical Journal* **81:**435-445.

Scherlis DA and Estrin DA (2002) Structure and spin-state energetics of an iron porphyrin model: An assessment of theoretical methods. *International Journal of Quantum Chemistry* **87:**158-166.

Schoneboom JC, Cohen S, Lin H, Shaik S and Thiel W (2004) Quantum mechanical/molecular mechanical investigation of the mechanism of C-H hydroxylation of camphor by cytochrome P450(cam): Theory supports a two-state rebound mechanism. *Journal of the American Chemical Society* **126:**4017-4034.

Schoneboom JC, Lin H, Reuter N, Thiel W, Cohen S, Ogliaro F and Shaik S (2002) The elusive oxidant species of cytochrome P450 enzymes: Characterization by combined quantum mechanical/molecular mechanical (QM/MM) calculations. *Journal of the American Chemical Society* **124:**8142-8151.

Schoneboom JC and Thiel W (2004) The resting state of p450(cam): A QM/MM study. *Journal of Physical Chemistry B* **108:**7468-7478.

Stich I, Car R and Parrinello M (1989) Bonding and Disorder in Liquid Silicon. *Physical Review Letters* **63:**2240-2243.

Williams PA, Cosme J, Sridhar V, Johnson EF and McRee DE (2000) Mammalian microsomal cytochrome P450 monooxygenase: Structural adaptations for membrane binding and functional diversity. *Molecular Cell* **5:**121-131.

Yin MT and Cohen ML (1982) Theory of Ab Initio Pseudopotential Calculations. *Physical Review B (Condensed Matter)* **25:**7403-7412.