Compound I Reactivity Defines Alkene Oxidation Selectivity in Cytochrome P450cam

Richard Lonsdale, Jeremy N. Harvey,* and Adrian J. Mulholland*

Centre for Computational Chemistry, School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, United Kingdom

Received: October 23, 2009

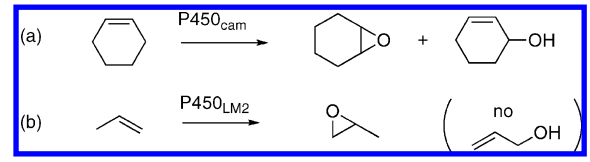
Prediction of the chemoselectivity of drug oxidation by the human cytochrome P450 enzymes will aid in the avoidance of adverse drug reactions. The chemoselectivity of alkene oxidation is an important problem to address, as it can result in the formation of epoxides, which can have toxic effects. In this paper the epoxidation and hydroxylation of cyclohexene and propene by the bacterial $P450_{cam}$ isoform are modeled with hybrid quantum mechanical/molecular mechanical (QM/MM) methods. Snapshots for QM/MM modeling are chosen from molecular dynamics trajectories, to sample the different conformations of the enzyme−substrate complex. The energy barriers obtained for these processes are in qualitative agreement with experimental work, supporting the use of QM/MM methods in the study of selectivity for this class of enzyme. This work highlights the complexity involved in modeling these systems with QM/MM and the importance in the selection of starting geometries.

Introduction

The cytochromes P450 (CYPs) are an important family of enzymes involved in the metabolism of xenobiotic molecules in the liver.[1] They catalyze a wide variety of reactions and are capable of oxidizing many different functional groups.[2] Knowledge of the regio- and chemoselectivity of CYP-mediated drug oxidation will aid in the prediction of the formation of reactive intermediates, such as epoxides. The latter are known to be responsible for the toxicological effects of some drugs.[3] Most models that are currently used to predict the selectivity of metabolism by CYPs are based on the relative accessibility of the different parts of the substrate molecule to the active species, rather than their intrinsic reactivity.[4] Examples exist of where substrates can adopt multiple binding positions in the active sites of CYPs.[5] In these cases the relative reactivity of the accessible functional groups of the substrate to the active species must be taken into account. This type of approach has been applied previously to CYP-catalyzed aromatic hydroxylation, using both QM and QM/MM to predict relative oxidation barriers for substituted benzenes[6] and several aromatic drug molecules.[7] The work in this paper serves to test whether hybrid quantum mechanical/molecular mechanical (QM/MM) calculations are capable of providing insight into the origin of chemoselectivity of alkene oxidation where accessibility is unlikely to be the sole factor governing the products observed.

The oxidations of cyclohexene and propene by $P450_{cam}$ from the soil bacterium *Pseudomonas putida* were chosen as model systems as they are relatively simple, with extensive existing experimental and theoretical data.[1,8] The oxidation of cyclohexene results in formation of C=C epoxidation and allylic hydroxylation products with an approximate 2:1 preference for epoxidation.[9] In contrast, the oxidation of propene by $P450_{LM2}$ results in formation of propene-1-oxide, with no hydroxylation products observed, as shown in Scheme 1.[10] Although no experimental data are available for propene oxidation by $P450_{cam}$, this is nevertheless a useful system to study by comparison with the cyclohexene case. As both of the substrates studied are small and nonpolar, it is expected that free tumbling of these substrate molecules occurs in the active site and hence the product distribution should be dependent on the relative reactivity of the functional groups relative to the active species of the enzyme, Compound I. Compound I is a high-valent iron−oxo porphyrin moiety that is widely accepted to be responsible for the majority of the oxidation chemistry catalyzed by CYPs.[11]

SCHEME 1: Oxidation of (a) Cyclohexene by $P450_{cam}$[9] and (b) Propene by $P450_{LM2}$[10]



Cases such as this, where several different products are formed from a single substrate, provide an excellent test for the performance of QM/MM in the study of CYP selectivity.

A previous attempt to model the oxidation of cyclohexene and propene with QM/MM methods was unable to account for the experimentally observed selectivity using calculated QM/MM energy barriers.[12] Cohen et al. suggested that relative energy barriers for oxidation of propene and cyclohexene obtained by QM/MM are not good predictors of the observed selectivity. Energy barriers were reported that favored the hydroxylation of cyclohexene by 6.8 kcal mol^{-1}, respectively (for propene, data reported in the same paper favored hydroxylation by 2.7 kcal mol^{-1}). An alternative method for predicting the selectivity was presented in which barriers to epoxidation and hydroxylation were calculated for the entire catalytic cycle. Each barrier was calculated as the sum of the energies of the enzyme−product complex, the barrier to the rate-limiting step (the second electron-transfer step), and the reaction free energy. Using this expression, the energy barrier is dominated by the energy of the enzyme−product complex, which reflects how tightly the product binds to the enzyme active site. Cyclohexen-3-ol and cyclohexene oxide products were calculated to bind to the heme

* Corresponding authors. E-mail: J.N.H., Jeremy.Harvey@bris.ac.uk; A.J.M., Adrian.Mulholland@bris.ac.uk.

10.1021/jp910127j
Published on Web 12/16/2009

by similar amounts hence the observed product ratio of C=C/C−H ≈ 1. For propene, the alcohol was calculated to bind more strongly to the heme than the epoxide and hence the strong preference for propene-1-oxide.

The problem with this theory is that it implicitly assumes that the two reaction cycles involving hydroxylation and epoxidation are disconnected, as the enzyme is required to have memory of the previous reaction turnover (see Supporting Information for an explanatory diagram). This would require that a CYP molecule releasing an epoxide can then only go on to form another epoxide (and likewise for hydroxylation). This is not a valid assumption, because after the product leaves the active site cavity, a common resting state is restored before the entrance of a subsequent substrate molecule. Oxidation of alkenes by Compound I is not a reversible process and if the barrier to hydroxylation was lower than that of epoxidation, even if the alcohol product was released more slowly than the epoxide, hydroxylation products would be observed. The enzyme must pass through the resting state to oxidize the next molecule, in which it is devoid of substrate and the heme group contains the low spin Fe(III) species, which is bound to a water molecule. The enzyme can hence have no memory of the previous cycle. The widely accepted view is that epoxidation and hydroxylation occur via a common intermediate when the substrate enters the protein active site.[13]

Previous QM/MM studies of other enzyme reactions have suggested that the choice of starting geometry can have a large effect on the energy barriers obtained in adiabatic mapping profiles.[14] Given the complexity in modeling enzymatic reactions with QM/MM, this type of system can benefit from study by more than one group. We hence decided to investigate the same system independently. Hydroxylation and epoxidation of cyclohexene and propene have been modeled in $P450_{cam}$ using QM/MM adiabatic mapping with multiple starting structures taken from molecular dynamics (MD) trajectories. Starting structures were selected on the basis of geometric criteria to model the most reactive geometries. An averaging procedure based on a Boltzmann distribution is employed, which reduces the contributions of unreasonably high energy barriers when comparing relative energies.

Computational Methods

Model Setup. Many of the intermediates in the hydroxylation of camphor by $P450_{cam}$ have been characterized by X-ray crystallography. One of these crystal structures was suggested to contain the elusive active oxygenating species Compound I but was later disproved by an EPR/ENDOR study by Davydov et al.[15] This X-ray crystal structure was obtained from the Protein Databank (code 1DZ9[16]) and used as a starting point for MD simulations for both systems studied here. Camphor was deleted from the original crystal structure and replaced by hand with propene in one case and cyclohexene in the other. Each MD simulation was run in CHARMM version 30b2[17] with the CHARMM27 force field.[19] The MD simulations had production phases of 4 and 1 ns for cyclohexene and propene, respectively. Further details of the setup for the MD simulations are contained in the Supporting Information.

QM/MM Setup. QM/MM calculations were performed using density functional theory (DFT) with the B3LYP[18] functional to describe the QM region and the CHARMM27 force field[19] to describe the MM region. The QM part of the calculations was carried out in Jaguar 5.5[20] and included the electric field generated by the charges on the MM atoms. Three different basis sets were used. For geometry optimization, the LACVP basis set[21] was used on the heme iron atom and the 6-31G** basis set on all other atoms (BSI). Single point calculations using the B3LYP and BP86 functionals were also performed on the optimized reactant complex and transition state geometries, both using the LACV3P basis set for iron, along with the 6-311G* and 6-311G** basis sets on all other atoms, denoted herein as BSII and BSIII, respectively. Previous calculations have shown the closely lying quartet and doublet spin states of Compound I to display similar energetics in the bond activation steps of hydroxylation and epoxidation reactions.[8] For this reason, only the quartet spin state was modeled in detail in this study. The QM region consisted of the substrate molecule, the methyl thiolate side chain of the heme-coordinating cysteine, and the unsubstituted porphyrin ring (Figure 1).

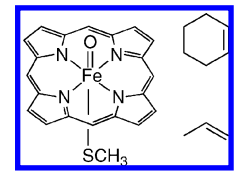


Figure 1. QM region used in all QM/MM and gas-phase QM-only calculations.

The MM part of the QM/MM calculations was carried out using Tinker 4.2.[22] Coupling between the QM and MM regions, and energy minimization of the QM region, was carried out with the QoMMMa program.[23] No cutoffs for electrostatic or van der Waals nonbonded interactions were used in the QM/MM calculations. The charges of the surface residues were not neutralized in the QM/MM calculations, resulting in an overall charge of $-14.95\ e^-$ for each system. This large negative charge was deemed to reside far enough away from the active site not to have an appreciable effect on the relative energies obtained along the reaction profiles. A similar setup was used in the study of camphor hydroxylation by Zurek et al. in which the energies obtained were consistent with those conducted on a whole enzyme system.[24]

Reaction profiles were generated using the adiabatic mapping procedure by carrying out a series of QM/MM minimizations with the distance between the Compound I oxygen and reacting atom of the substrate fixed at different values.

QM-Only Model Calculations. Gas-phase QM calculations were carried out using Jaguar 6.0[25] with the same density functionals and basis sets as those used for the QM/MM calculations. Transition states (TSs) were optimized in two steps. An energy profile was first generated by optimizing the geometry at a large number of fixed distances between the ferryl Compound I oxygen and the atom to which a bond is being formed. A full TS optimization was then carried out using the quadratic synchronous transit (QST) method implemented in Jaguar.[25] Frequency calculations carried out on each TS revealed a single imaginary frequency corresponding to movement along the expected reaction coordinate. All gas-phase calculations modeled Compound I in the quartet spin state. Gas-phase energy barriers were calculated by subtracting the sum of the separate energies of Compound I and the substrate from the transition state energy. Zero-point energies (ZPEs) obtained from the QM-only calculations were used in the calculation of ZPE corrections for the QM/MM barriers.

Boltzmann-Weighted Averaging of Energy Barriers. For the QM/MM reaction barriers, two different methods have been used to calculate the average value over all of the chosen MD snapshots. In the first instance, simple arithmetic averages were calculated. Another method was also chosen in which a Boltzmann-weighted average reaction barrier was calculated for

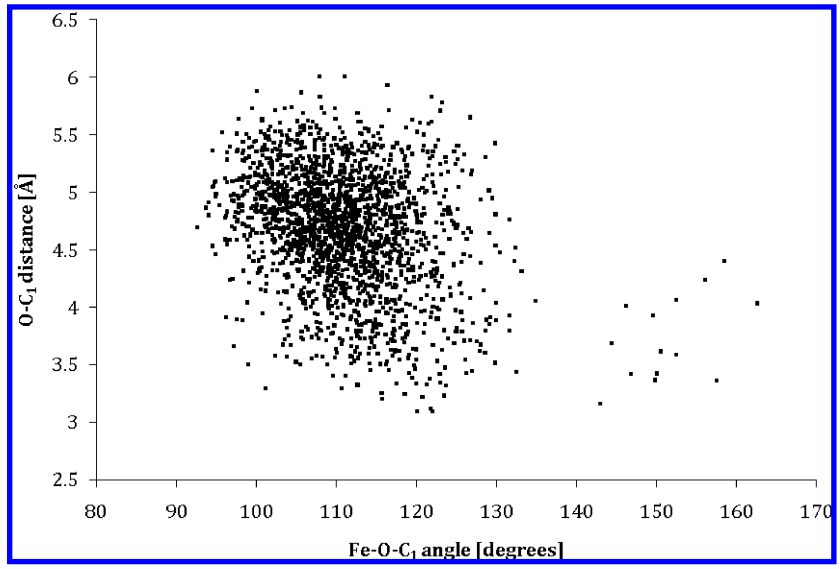


Figure 2. Plot of the distance between the ferryl oxygen of Compound I and C_1 of propene against the angle formed between the Fe−O bond and C_1 carbon. Each point on the graph corresponds to an individual snapshot from the MD simulation of propene with $P450_{cam}$.

each process. The expression used to calculate these averages is given in eq 1.

$$\Delta E^{\ddagger}_{ave} = -RT \ln\left\{\frac{1}{n}\sum_{i=1}^{n} \exp\left(\frac{-\Delta E_i^{\ddagger}}{RT}\right)\right\} \qquad (1)$$

where $\Delta E^{\ddagger}_{ave}$ is the average barrier height, R is the gas constant, n is the number of energy profiles considered, $\Delta E_i^{\ddagger}$ is the barrier height computed for pathway i, and T is the temperature (300 K). This averaging method relies on the assumption that each starting geometry sampled from the MD trajectory occurs with equal probability. It also assumes that the overall reactivity is obtained from the arithmetic average of the "local" rate constants k_i associated with each starting geometry, with the values k_i determined by the Eyring equation:

$$k_i = \frac{k_B T}{h} \exp\left(\frac{-\Delta G_i^{\ddagger}}{RT}\right)$$

where k_B is the Boltzmann constant, h is the Planck constant, and $\Delta G_i^{\ddagger}$ is the free energy barrier for the reaction. A further assumption is that $\Delta G_i^{\ddagger}$ and $\Delta E_i^{\ddagger}$ are similar, which is reasonable as a large change in entropy is not expected.[26]

This latter approach could be considered to be more sophisticated as it largely eliminates the contributions of barriers that are unreasonably high, often due to less-favorable starting geometries. It hence focuses on the lower, more relevant barriers that are more representative of the actual reactions taking place. In the limit of very large n, one should therefore get quite accurate results. However, for small n, if the set of starting geometries happens to include one with an anomalously low value of $\Delta E_i^{\ddagger}$, this will have a disproportionate effect on the overall barrier. In our results presented below, n is fairly small, so care is needed when interpreting results from both averaging methods. For the Boltzmann-weighted averaging method, we will consider, in particular, the effect of neglecting the pathway with the lowest barrier value $\Delta E_i^{\ddagger}$ for each process considered, as this gives some insight into the degree of robustness of our conclusions with respect to the addition of further pathways.

Results and Discussion

Selection of Prereactive Geometries. The MD simulations described above were used to sample the movement of the substrates propene and cyclohexene around the active site of $P450_{cam}$ and obtain starting snapshots for QM/MM calculations. Both simulations showed very low root-mean-square deviations of the protein backbone atoms from their positions in the crystal structure (less than 0.7 Å), indicating that both systems are well-equilibrated and stable at 300 K. Propene and cyclohexene both move relatively freely around the active site of the enzyme, unlike the natural substrate camphor.[24] This is due to the lack of hydrogen-bonding available to propene and cyclohexene with the residues surrounding the active site pocket. As a result of this, a wide range of substrate orientations was observed during the MD simulations, with the C=C bond and the allylic C−H bonds not always in an optimal position for oxidation to occur. We assume that the MD simulation generates a reasonably accurate description of the distribution of geometries sampled through rapid equilibria by the Compound I−substrate system, so that the many local minima obtained upon minimization of MD snapshots are roughly equally likely to occur. Note that it is well-known that these local minima can vary in potential energy by more than 100 kcal mol^{-1},[27] so their total energy does not provide a meaningful way to identify the most favored structure. Figure 2 displays the large variation in both Fe−O−C_1 angle and O−C_1 distance throughout the course of the MD simulation of $P450_{cam}$ with propene. For this reason extensive analysis of the MD trajectories was necessary as well as careful screening of the snapshots used for QM/MM calculations.

The slowest step in the active-site oxygen-transfer process leading to epoxidation of propene is the bond formation between C1 of propene and the ferryl oxygen of Compound I. The energy profile for the oxygen transfer as established in work carried out by Shaik et al. using a gas-phase model clearly shows that formation of the second C−O bond is more facile than the first step.[28] It is possible that other steps in the overall catalytic cycle, such as electron transfer or product release, are overall turnover-limiting but this is not important for assessing selectivity. Addition to C_1 rather than C_2 is preferred, resulting in formation of a secondary radical on propene as opposed to the less stable primary radical. Hydroxylation of alkenes by Compound I proceeds via the same mechanism as that of alkanes, i.e., by slower hydrogen abstraction followed by rapid radical-oxygen rebound leading to alcohol formation.[29]

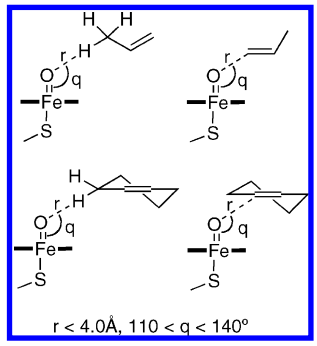


Figure 3. Selection criteria used in the prescreening of MD snapshots for QM/MM adiabatic mapping.

Because the substrates move relatively freely around the active site, they are expected to spend a significant amount of the time in positions that are not optimal for oxidation. In computational terms, this would translate into initial structures with high activation barriers along the adiabatic pathway to oxidation. The nature of eq 1 is such that individual pathways with high activation barriers barely contribute to the net average barrier. Hence for the QM/MM calculations of the reaction barriers, the starting snapshots were prescreened on the basis of knowledge of the transition state geometries from QM calculations.[28] For epoxidation snapshots the distance between the activated carbon and the Compound I oxygen was required to be less than 4.0 Å, and the angle formed between these two atoms and the iron atom was required to be between 110 and 140°. Similarly, for hydroxylation, the distance between the abstracted hydrogen and the Compound I oxygen was required to be less than 4.0 Å, and an angle criterion analogous to that for epoxidation was also used (see Figure 3).

If this prescreening had resulted in considering only very few starting geometries, and especially if the number of geometries considered suitable for epoxidation and hydroxylation had been very different, this would have introduced a bias in the results. While it is true that *individual* pathways with high activation barriers do not contribute to the average barrier, if most of the possible reactant complexes are unfavorable; i.e., if the alkene resides in unreactive geometries with respect to Compound I for most of the time, then only a small number of the exponential terms will contribute significantly to the summation in eq 1, yet n will be large. This is equivalent to an additional contribution to the free energy of activation, resulting from the free energy difference between the unreactive and reactive configurations. In fact, the time spent in geometries suitable for epoxidation and hydroxylation of both substrates represents a substantial fraction of the total MD simulation times (>5% in all cases; see Supporting Information for further details). It can therefore be assumed that this accessibility free energy term, which would need to be included in a fully quantitative study, does not have a major effect on the selectivity in this case, and so can be ignored in the present, semiquantitative approach.

We note that the selection criteria defining "reactive" geometries for epoxidation and hydroxylation are different. Indeed, the difference was sufficiently large that it was not possible to locate MD snapshots that satisfied the criteria for both processes. Hence the QM/MM modeling of epoxidation and hydroxylation of the substrate relied upon different ensembles of starting geometries. Enforcing a unique set of geometries to use as starting structures to model both processes is not necessary and, with a small number of pathways, might bias the results in favor of one or the other process. In the present work, epoxidation and hydroxylation barriers were calculated from different sets of MD snapshots.

QM/MM Energy Profiles: Propene. QM/MM energy barriers for hydroxylation of propene were calculated from adiabatic mapping profiles originating from eight different MD snapshots. Profiles were generated by fixing the distance between the Compound I ferryl oxygen and the closest of the three hydrogen atoms attached to the saturated carbon atom of propene at a range of values, while optimizing the rest of the system. The structures for the Compound I species in the reactant complex and for the TSs for hydrogen abstraction or addition to the double bond are rather similar to those found in earlier QM or QM/MM studies.[12,24,31] In particular, the key Fe−S, Fe−O, O−H, and H−C distances, as well as the relevant angles, in the abstraction TSs are similar to those found previously. Likewise for the epoxidation TS, the O−C distance at the addition TS is similar to that described in other studies. These structural parameters are listed in the Supporting Information. Notably, the variability in these structural parameters within the QM region is quite small from one pathway to another, so it does not appear that this alone could account for differences in relative energies obtained with respect to the previous QM/MM study.[12] Also, the electronic structure, especially the distribution of unpaired electrons, obtained for the Compound I species and the TSs (see calculated spin densities in the Supporting Information) is in line with that found in previous studies.

The ZPE-corrected barriers for hydroxylation of propene calculated with BSIII range between 5.0 and 22.1 kcal mol^{-1}, giving an average barrier of 11.4 kcal mol^{-1}. Zero-point energy (ZPE) corrections were obtained from frequency calculations performed on the gas-phase transition states. The more sophisticated Boltzmann-weighted averaging procedure yielded a value of 6.2 kcal mol^{-1}. The gas-phase barrier was calculated using a starting geometry of the porphyrin and propene from the reactant complex derived from the 854.5 ps snapshot. The calculated barrier of 14.6 kcal mol^{-1} is higher than the value of 10.3 kcal mol^{-1} that was calculated in the enzyme with QM/MM. It is presently not possible to use the QST algorithm that is used to optimize the gas-phase TSs on a full protein system due to size limitations. Hence the QM/MM TSs are approximate but their geometries closely resemble those observed in the gas phase.

Seven QM/MM energy barriers have been obtained for the epoxidation of propene. These profiles were generated by modeling the distance between the Compound I ferryl oxygen and the terminal unsaturated carbon of propene as the reaction coordinate. The ZPE-corrected barriers calculated with BSIII range between 8.1 and 13.4 kcal mol^{-1}, with an average value of 10.9 kcal mol^{-1}. As with all of the QM/MM barriers calculated in this work, there is no correlation between the energy barrier obtained and the time of the snapshot used. The Boltzmann-weighted average in this case is 9.2 kcal mol^{-1}. The gas-phase propene epoxidation barrier was calculated using the reactant complex from the 292.5 ps snapshot as a starting geometry. The ZPE-corrected gas-phase barrier calculated with BSIII lies 6.8 kcal mol^{-1} higher than the Boltzmann-weighted average QM/MM barrier.

For propene, the Boltzmann-weighted average barrier to epoxidation, 9.2 kcal mol^{-1}, is 3.0 kcal mol^{-1} higher than that of hydroxylation, 6.2 kcal mol^{-1}. Upon leaving out the lowest computed barrier for both processes, the Boltzmann average still favors hydroxylation over epoxidation by 2.5 kcal/mol. The experimental work[10] led to exclusive formation of epoxide, albeit with a different isoform, $P450_{LM2}$. Assuming this is equivalent to an epoxide to alcohol product ratio of 98:2 or greater, the

epoxidation reaction should be favored by more than 2.3 kcal/mol, leading to an error of ca. 5 kcal/mol in the computational results if one further assumes that the selectivity would be similar in $P450_{cam}$. While this second assumption is not necessarily completely justified, the small size of the substrate suggests that steric effects from the protein environment should be fairly small, and a very different selectivity would be unexpected. In any case, caution is required in comparing our results with experiment, with a conservative conclusion being that the present QM/MM procedure probably slightly overestimates the ease of hydroxylation versus epoxidation for alkenes. Disagreement between experiment and theory of the magnitude found here is not uncommon, especially when one considers that in the present case, as well as possible errors in the DFT methodology,[30] errors in the QM/MM method and in the sampling of starting geometries can also contribute. In particular, we note that a relatively small number of barriers have been sampled here.

The calculated QM/MM activation barriers are similar to the values obtained in QM studies of iron−porphine reacting with propene,[28] showing that the protein environment does not completely change the reactivity, as expected given that the electronic structure of Compound I does not vary significantly between models and in QM/MM calculations.[31] Finally, we note that the use of simple arithmetic averaging of the individual epoxidation barriers yields an energy barrier of 10.9 kcal mol^{-1}, which is less than the corresponding value of 11.4 kcal mol^{-1} observed for hydroxylation. Although this represents a superior agreement with experiment than with the Boltzmann-weighting procedure, this is probably somewhat fortuitous. The propene hydroxylation barriers show a greater degree of variation than those of epoxidation with the highest barrier being 22.1 kcal mol^{-1}, compared to 13.4 kcal mol^{-1} for epoxidation. Hence the average hydroxylation barrier is unreasonably elevated by the highest barrier, which is unlikely to be chemically relevant.

QM/MM Energy Profiles: Cyclohexene. A total of eight adiabatic mapping profiles were calculated from selected MD snapshots for the hydroxylation of cyclohexene. As cyclohexen-3-ol is one of the major products observed in oxidation by $P450_{cam}$,[9] it is assumed that the rate limiting step is hydrogen atom abstraction from a carbon atom adjacent to the double bond. Hence the reaction coordinate was selected as the distance between the Compound I ferryl oxygen and the closest of the four hydrogen atoms attached to the adjacent carbon atoms. The barriers range between 9.7 and 18.2 kcal mol^{-1}, with an average value of 13.9 kcal mol^{-1}. The use of the Boltzmann-weighting procedure gives an average value of 10.9 kcal mol^{-1}. The ZPE-corrected gas-phase barrier of 13.8 kcal mol^{-1}, calculated with BSIII is comparable to the average QM/MM barrier of 13.9 kcal mol^{-1}. Calculations using the BSII basis set on the QM/MM geometries obtained with BSI raise the barriers by ca. 1.7 kcal mol^{-1} relative to the values obtained with BSI; however, barriers calculated with BSIII, which (like BSI) contains polarization functions for hydrogen atoms, are mostly within 0.1 kcal mol^{-1} of the energies calculated with BSI. Energies of all pathways are presented in the Supporting Information.

Eleven QM/MM adiabatic mapping profiles were obtained for the epoxidation of cyclohexene. The reaction coordinate was chosen as the distance between the Compound I ferryl oxygen and the closest of the two unsaturated carbon atoms of cyclohexene. The QM/MM energy barriers vary between 12.6 and 25.4 kcal mol^{-1}, with an average value of 17.9 kcal mol^{-1}. The Boltzmann-weighted average in this case is 14.0 kcal mol^{-1}. The starting geometry for the modeling of hydroxylation of cyclohexene in the gas phase originated from the 928.5 ps snapshot. The gas-phase barrier is 18.2 kcal mol^{-1}, 1.7 kcal mol^{-1} higher than the corresponding barrier height calculated using the QM/MM method. In the case of epoxidation of cyclohexene, BSII lowers the reaction barriers of the QM/MM adiabatic mapping profiles by ca. 2.0 kcal mol^{-1} compared with the values obtained with BSI (see the Supporting Information for more details). The barriers calculated with BSIII are in this case mostly within 0.1 kcal mol^{-1} of those calculated with BSII. Further single point energies were calculated using the pure BP86 density functional with BSIII to compare with the energies obtained using B3LYP. These energies are presented in the Supporting Information. The BP86 energy barriers were, on average, 5 kcal mol^{-1} lower for epoxidation and 6 kcal mol^{-1} lower for hydroxylation, compared with the B3LYP pathways calculated with the same basis set, showing that the energetics in this system are quite sensitive to the DFT functional used.[30] Some error in the energies reported in this work must hence be allowed for.

The Boltzmann-weighted average barrier for hydroxylation of cyclohexene is 10.9 kcal mol^{-1}, compared to the value of 14.0 kcal mol^{-1} calculated for epoxidation. On first inspection, this difference of 3.1 kcal mol^{-1} appears not to agree with the experimentally observed 2:1 ratio of cyclohexene-1-oxide and cyclohexen-3-ol formed in the oxidation of cyclohexene by $P450_{cam}$. However, as with propene oxidation, the errors of the QM/MM and DFT methods must be considered, and these could easily lead to an error margin of up to 3 kcal mol^{-1}. Relatively few pathways have been sampled and using this method with several dozen more pathways might yield a better agreement but at great computational expense. Omitting the lowest barriers in both cases leads to a difference between the Boltzmann-averaged barriers for epoxidation and hydroxylation of 3.5 kcal/mol, similar to the difference of 3.1 kcal/mol obtained when including all barriers. The QM/MM calculations in the enzyme show slightly better agreement with experiment than those of the gas-phase QM-only model, where hydroxylation is favored by 4.4 kcal mol^{-1}. The average barrier calculated with the simpler arithmetic scheme for hydroxylation is 13.9 kcal mol^{-1} compared to 17.9 kcal mol^{-1} for epoxidation. Unlike with propene, this does not represent a better agreement with the experimentally observed product ratios, but as mentioned above, there is no reason to expect better results with arithmetic averaging; indeed this procedure is expected to yield less meaningful results. A key point is that as for propene, Boltzmann-weighted averaging and arithmetic averaging yield differences between epoxidation and hydroxylation that are consistent with experiment, within the expected computational accuracy.

A summary of the relative energies for hydroxylation and epoxidation of propene and cyclohexene, along with the experimentally observed product ratios and previous computational work (by other authors[12]) is displayed in Table 1. To make the most meaningful comparison between this work and previous work, the energies presented in Table 1 are those calculated with both BSII and BSIII. BSII most closely resembles the basis set used to calculate the energies in the previous work.[11] In Table 1 it can be shown that in the present work, hydroxylation of cyclohexene appears to be favored by 1.5 kcal mol^{-1} with BSII, compared to the previous work,[12] where cyclohexene was favored to a much larger extent (7.6 kcal mol^{-1}). Likewise, for propene, hydroxylation is favored to a smaller extent in this work than previously. This suggests that the hypothesis whereby

TABLE 1: Relative Difference in Activation Energies, $\Delta\Delta E^{\ddagger}$ (kcal mol^{-1}) for Oxidation of Cyclohexene and Propene by P450$_{cam}$

$\Delta\Delta E^{\ddagger} = \Delta E_E{}^{\ddagger} - \Delta E_H{}^{\ddagger}$	cyclohexene	propene
experiment	−0.5	−2.3 (upper bound estimate)
this work (BSII)	+1.5[a]	+0.3[a]
this work (BSIII)	+3.1[a]	+3.0[a]
ref 10	+7.6[b]	+1.4[b]

[a] Taken from the Boltzmann-weighted average barriers, calculated with BSII (BSIII energies are provided in parentheses). [b] Calculated with LACVP (Fe), 6-31G (all other atoms).

oxidation chemistry is not selectivity-determining is not necessary. The difference between the present study and the previous work is presumably due to minor differences in methodology (QM basis sets, MM setup, size of system), as well as differences in sampling geometries.

In summary, an improved agreement has been observed here between the relative QM/MM energy barriers for oxidation of cyclohexene and the experimentally observed product ratio. This provides evidence to support the suitability of using QM/MM to provide qualitative understanding of chemoselectivity for cytochrome P450 oxidation of other, more complex, substrates. Indeed, the results are good enough to provide qualitative prediction of relative reactivity.

Conclusions

In this work it has been shown that energy barriers obtained from QM/MM calculations are capable of qualitative agreement with experimental selectivities for the oxidation of alkenes by P450$_{cam}$, at least to within the range of errors expected from the present energy-based QM/MM calculations.

Molecular mechanics-based molecular dynamics simulations have been used to show that cyclohexene and propene both move relatively freely around the active site of P450$_{cam}$, exposing both hydroxylation and epoxidation sites to the ferryl oxygen of Compound I. From these simulations it can hence be assumed that accessibility is not a major factor in determining the selectivity of oxidation in these systems. It is more likely that the relative reactivity of Compound I toward the particular sites is the dominant factor in determining the relative ratios of hydroxylation and epoxidation products for these substrates.

Using prescreened snapshots from the MD simulations, multiple QM/MM adiabatic mapping profiles have been obtained for the hydroxylation and epoxidation of both cyclohexene and propene. The relative energy barriers for the processes are consistent with the selectivity that is observed experimentally, given the expected errors in the calculations. The relative Boltzmann-weighted average barriers for epoxidation and hydroxylation of propene suggest that these processes are competitive. Experimentally, propene-1-oxide is the sole product of propene oxidation by P450$_{LM2}$,[10] and assuming that the selectivity in P450$_{cam}$ is similar, this suggests that the present methodology somewhat (2−3 kcal/mol) overestimates the ease of hydroxylation. However, our results are closer to the observed experimental behavior than those from a previous study.[12] The Boltzmann-weighted average barriers obtained for epoxidation and hydroxylation of cyclohexene are of 14.0 and 10.9 kcal/mol. Experimentally, cyclohexene-1-oxide and cyclohexen-3-ol are formed with a product ratio of 2:1 in the oxidation of cyclohexene by P450$_{cam}$,[9] corresponding to a difference in free energy of activation of ca. 0.5 kcal/mol in favor of epoxidation. Hence for this case, where the computations and experiment address the same system, it again appears that the QM/MM methodology is biased in favor of the hydroxylation by a small amount. The error is, however, comparable to that expected for problems of this type. In fact, even pure QM calculations can lead to errors of the magnitude observed.[30] Importantly, the error is small enough that the calculations provide a model of the selectivity that is consistent with the experimental observations. In previous work, a much larger difference in energy barrier (7.6 kcal/mol) in favor of hydroxylation was found, and it was suggested therefore that selectivity must arise from some other step in the catalytic cycle.[12] The present results suggest that there is no strong need to consider a novel type of explanation for the selectivity, as the computations are consistent with the a priori most likely and most simple interpretation, whereby it is determined during the oxidation step by the relative reactivity of the Compound I species.

The individual energy barriers obtained for these processes vary very significantly depending on the starting geometries chosen. This occurs despite the fact that starting geometries were obtained after prescreening of the structures found in the MD simulation, to select only structures that were expected to lead to reasonably low barriers. It is well-known that QM/MM energy barriers derived from adiabatic mapping are very sensitive to small differences in the starting geometry. The observed variation in barriers stresses the need for extensive conformational sampling to generate enzyme−substrate complex structures that are truly representative of reactive conformations, especially for an enzymatic system of the present type, where the structure of the enzyme−substrate complex is not well-defined due to the absence of specific interactions between the substrate and the binding pocket. Careful selection of a range of starting geometries is important for getting accurate results and may explain part of the difference between the present results and those obtained previously by others. However, given the complexity of the system, it is likely that other factors contribute also.[12] In summary, this work highlights the complexity of QM/MM calculations yet shows them to be a promising approach to modeling selectivity in CYPs.

Acknowledgment. We acknowledge support from the BBSRC and EPSRC.

Supporting Information Available: Energy cycle diagram for calculating effective barries for epoxidation and hyroxylation; disconnected reaction pathway diagram; discussion of computational methods; list of force field parameters; epoxidation and hydroxylation figures; tables of criteria for epoxidation and hydroxylation, energy barriers for bond activation, hydroxylation and epoxidation profiles and pathways, bond lengths and angles, and absolute energies; discussion of the influence of including the proximal water molcules in the QM region; and Cartesian coordinates. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) *Cytochrome P450: Structure, Mechanism and Biochemistry*, 3rd ed.; Ortiz de Montellano, P. R., Ed.; Kluwer Academic/Plenum Publishers: New York, 2004.

(2) Guengerich, F. P. *Chem. Res. Toxicol.* **2001**, *14*, 611–650.

(3) Guengerich, F.; Arneson, K.; Williams, K.; Deng, Z.; Harris, T. *Chem. Res. Toxicol.* **2002**, *15*, 780–792.

(4) See, e.g.: Cruciani, G.; Carosati, E.; De Boeck, B.; Ethirajulu, K.; Mackie, C.; Howe, T.; Vianlello, R. *J. Med. Chem.* **2005**, *48*, 6970–6979.

(5) (a) Strickler, M.; Goldstein, B. M.; Maxfield, K.; Shireman, L.; Kim, G.; Matteson, D. S.; Jones, J. P. *Biochemistry* **2003**, *42*, 11943–11950. (b) Modi, S.; Gilham, D. E.; Sutcliffe, M. J.; Lian, L. Y.; Primrose, W. U.; Wolf, C. R.; Roberts, G. C. K. *Biochemistry* **1997**, *36*, 4461–4470.

(6) (a) Bathelt, C. M.; Ridder, L.; Mulholland, A. J.; Harvey, J. N. *J. Am. Chem. Soc.* **2003**, *125*, 15004–15005. (b) Bathelt, C. M.; Ridder, L.; Mulholland, A. J.; Harvey, J. N. *Org. Biomol. Chem.* **2004**, *2*, 2998–3005.

(7) Rydberg, P.; Ryde, U.; Olsen, L. *J. Phys. Chem. A* **2008**, *112*, 13058–13065.

(8) Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. *Chem. Rev.* **2005**, *105*, 2279–2328.

(9) (a) Yoshioka, S.; Takahashi, S.; Ishimori, K.; Morishima, I. *J. Inorg. Biochem.* **2000**, *81*, 141–151. (b) White, R. E.; Groves, J. T.; McClusky, G. A. *Acta Biol. Med. Germ.* **1979**, *38*, 475–482.

(10) Groves, J. T.; Avaria-Neisser, G. E.; Fish, K. M.; Imachi, M.; Kuczkowski, R. L. *J. Am. Chem. Soc.* **1986**, *108*, 3837–3838.

(11) (a) Egawa, T.; Shimada, H.; Ishimura, Y. *Biochem. Biophys. Res. Commun.* **1994**, *201*, 1464–1469. (b) Guengerich, F. P.; Vaz, A. D. N.; Raner, G. N.; Pernecky, S. J.; Coon, M. J. *Mol. Pharmacol.* **1997**, *51*, 147–151.

(12) Cohen, S.; Kozuch, S.; Hazan, C.; Shaik, S. *J. Am. Chem. Soc.* **2006**, *128*, 11028–11029.

(13) Denisov, I.; Makris, T.; Sligar, S.; Schlichting, I. *Chem. Rev.* **2005**, *105*, 2253–2277.

(14) (a) Lodola, A.; Mor, M.; Zurek, J.; Tarzia, G.; Piomelli, D.; Harvey, J.; Mulholland, A. *Biophys. J.* **2007**, *92*, L20–L22. (b) Claeysens, F.; Ranaghan, K.; Manby, F.; Harvey, J.; Mulholland, A. *Chem. Commun.* **2005**, 5068–5070.

(15) Davydov, R.; Makris, T. M.; Kofman, V.; Werst, D. E.; Sligar, S. G.; Hoffmann, B. M. *J. Am. Chem. Soc.* **2001**, *123*, 1403–1415.

(16) Schlichting, I.; Berendzen, J.; Chu, K.; Stock, A.; Maves, S.; Benson, D.; Sweet, R.; Ringe, D.; Petsko, G.; Sligar, S. *Science* **2000**, *287*, 1615–1622.

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

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

(19) 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, X.; 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.

(20) *Jaguar 5.5*; Schrödinger LLC: Portland, OR, 2005.

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

(22) Ponder, J. W. Tinker - Software Tools for Molecular Design; 2004.

(23) Harvey, J. N. *Faraday Discuss.* **2004**, *127*, 165–177.

(24) Zurek, J.; Foloppe, N.; Harvey, J. N.; Mulholland, A. J. *Org. Biomol. Chem.* **2006**, *4*, 3931–3937.

(25) *Jaguar 6.0*; Schrödinger LLC: Portland, OR, 2005.

(26) 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., Int. Ed. Engl.* **2006**, *45*, 6856–6859.

(27) Elber, R.; Karplus, M. *Science* **1987**, *235*, 318–321.

(28) de Visser, S. P.; Ogliaro, F.; Sharma, P. K.; Shaik, S. *J. Am. Chem. Soc.* **2002**, *124*, 11809–11826.

(29) Groves, J.; McClusky, G. *J. Am. Chem. Soc.* **1976**, *98*, 859–861.

(30) (a) Harvey, J. N. *Annu. Rep. Prog. Chem., Sect. C: Phys. Chem.* **2006**, *102*, 203–226. (b) Tian, L.; Friesner, R. A. *J. Chem. Theory Comput.* **2009**, *5*, 1421–1431.

(31) Bathelt, C. M.; Zurek, J.; Mulholland, A. J.; Harvey, J. N. *J. Am. Chem. Soc.* **2005**, *127*, 12900–12908.

JP910127J