General transition-state force field for cytochrome p450 Hydroxylation

Rydberg, Patrik; Olsen, Lars; Norrby, Per-Ola; Ryde, Ulf

Published in:
Journal of Chemical Theory and Computation

DOI:
10.1021/ct700110f

2007

Document Version:
Peer reviewed version (aka post-print)

A general transition-state force field for cytochrome P450 hydroxylation

Patrik Rydberg [a], Lars Olsen [b], Per-Ola Norrby [c] and Ulf Ryde [a]*

[a] Department of Theoretical Chemistry, Lund University, Chemical Centre, P. O. Box 124, SE-221 00 Lund, Sweden

[b] Biostructural Research Group, Faculty of Pharmaceutical Sciences, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen Ø, Denmark

[c] Department of Chemistry, Göteborg University, Kemigården 4, SE-412 96 Göteborg, Sweden

* Correspondence to Ulf Ryde; E-mail: Ulf.Ryde@teokem.lu.se; Phone: +46 – 46 2224502; Fax: +46 – 46 2224543

2017-04-10

Running title: P450 transition-state force field

Abstract

We have developed force-field parameters for the hydrogen-abstraction transition state of aliphatic hydroxylation by cytochrome P450 using the Q2MM approach. The parameterisation is based on quantum chemical (B3LYP) transition-state structures and Hessian matrices for 24 diverse substrate models (14 in the training set and 10 in the test set). The force field is intended to be applicable to any drug-like molecule by the use of the general Amber force field (GAFF) for the substrates. The parameters reproduce the geometries within 0.1 Å and 1.2º for bond lengths and angles, respectively, with no significant differences between the training and test sets. The Hessian matrix is also well reproduced with a correlation coefficient of 0.99. The parameterisation is performed by the ideal iterative approach of Norrby and Liljefors, which we have implemented for the Amber software.

Introduction

The cytochrome P450 enzymes (CYPs) are a superfamily of mono-oxygenases found in all types of organisms from bacteria to mammals. In the human genome, there are almost 60 genes for CYPs. They take part in the synthesis of important endogenous compounds such as steroids, prostaglandin, and fatty acids. However, they also contribute to the degradation of exogenous compounds. They affect both activation of prodrugs as well as the bioavailability and degradation of drugs. It has been estimated that the CYPs are responsible for 75% of the phase I drug metabolism.[i,ii] Therefore, they have attracted much attention in pharmaceutical research.

Almost 200 crystal structures of CYPs have been published. They show a highly conserved active site, which consists of a haem group with an iron ion in the centre of the porphyrin ring. In contrast to most other haem enzymes, the iron ion coordinates to the sulphur atom of a cysteine residue. This negatively charged ligand is believed to favour the formation of high-valent reactive iron intermediates of the CYPs.[iii] The sixth coordination site of the iron ion, opposite to the cysteine ligand, is open to the binding of small extraneous ligands during the reaction cycle of the enzyme.

In the resting state, it is occupied by a water molecule, and the iron ion is in a low-spin Fe(III) state (cf. Figure 1). Binding of a substrate triggers the release of the water ligand, leading to a switch to the high-spin Fe(III) state, but the substrate does not directly coordinate to the iron ion. After a one-electron reduction, the Fe(II) ion binds O_2. This complex is reduced a second time, which triggers a heterolytic cleavage of the O–O bond, giving rise to a Fe(V)=O complex (formally), called compound I. This state is highly reactive, and can perform many different reactions, such as hydroxylation, epoxidation, dealkylation, N, S, and SO oxidation.[i,ii]

Much effort has been devoted to the understanding of the reactivity of various CYPs. For example, the intrinsic reactivity of the CYP active site has been extensively studied, especially with density functional theory (DFT) and for the hydroxylation reaction.[iv,v,vi] Attempts have also been made to reproduce the DFT results with cheaper methods, allowing for the scanning of many compounds in a short time.[vii,viii,ix] Moreover, the accessibility and binding of various substrates to the active site of human CYPs have been studied with pharmacophore models, docking, (3D-)QSAR, etc.[x,xi,xii]

Standard docking algorithms consider only the binding of ground-state molecules. However, for enzyme substrates, it could be more favourable to dock in the reactive transition state instead, because it must form if the substrate is to react. Such an approach would make the docking discriminative, because the conformational space available to the substrate is more restricted in the transition state (it must bind to the oxoferryl group). In fact, it has been recently shown that docking of high-energy intermediates to enzymes can improve the prediction of the reactivity of an enzyme significantly.[xiii]

Unfortunately, docking of transition states is not straight forward, because standard methods of molecular mechanics cannot be directly employed (because transition states are not equilibrium states, but first-order saddle points on the potential energy surface).[xiv] Therefore, special software is normally needed for their optimisation and available algorithms are quite time-consuming and cannot guarantee that a transition state is obtained from any starting point. Several methods are available for reproducing entire potential energy surfaces of reactions using empirical force fields.[xiv] Already in 1980, Warshel introduced the empirical valence bond method,[xv] whereby the entire potential energy surface could be obtained by mixing of the reactant and product states, each modelled by an appropriate force field. This methodology has been used extensively, in particular for describing reactions in enzymes.[xvi] More recent variations on the same theme include the Rappé's reactive force field[xvii] and Truhlar's multi-configurational molecular mechanics[xviii] methods. In an alternative approach, Goddard has recently developed a force field that allows direct bond breaking.[xix] Each of these methods generate a full potential energy surface, requiring a saddle point search for locating a transition state, a task that is not easily automated and combined with a conformational search method. The SEAM method by Jensen[xx] circumvents this problem by directly locating the intersection between the reactant and product force fields, a method that is very robust, even for poor starting geometries. Finally, the Q2MM method,[xxi] or more generally, transition state force fields,[xxii] defines a new force field that treat transition state structures as energy minima. A severe drawback with this method is that it no longer allows comparison with reactants, and therefore cannot yield absolute activation energies; only relative barriers can be calculated. However, this is sufficient for investigating selectivities. Moreover, the method can be directly implemented in standard

molecular mechanics software, and it is robust enough to allow full conformational searching. For these reasons we decided to proceed with the Q2MM method.

In this paper, we take the first step in this direction by developing a general transition-state force field for the hydroxylation reaction of the CYPs. A haem group has been parameterised before, both in the CYPs[xxiii] and in other proteins (i.e. with different axial ligands),[xxiv,xxv,xxvi,xxvii,xxviii] but never a transition state. Our force field is based on 14 transition-state structures obtained at the DFT level on a diverse set of substrates.[ix] In order to make the force field more general, we do not attempt to parameterise the substrate, but instead take those parameters directly from the general Amber force field (GAFF), which is a general and diverse force field, designed for drug-like molecules.[xxix] Thus, our effort is concentrated around the haem group and the reactive oxoferryl group. We evaluate the force field by comparing structures and energies with results obtained at the DFT level for a test set of 10 compounds.

Methods

Substrate models

According to the consensus mechanism, the hydroxylation of substrates by the CYPs takes place in two steps:[iv,v,vi] First, the oxoferryl group of compound I abstracts a hydrogen atom from the substrate (Figure 1), yielding a Fe(IV)–OH intermediate and a substrate radical. In the next step, the substrate radical rebounds to the OH group, forming an iron-bound hydroxylated substrate. The DFT calculations show that the first step is rate limiting, and therefore, we have restricted our investigation to the transition state of that step.

Our transition-state force field for the CYP hydroxylation reaction is based on DFT optimised structures of 24 small but diverse organic substrates.[ix] This set was divided into a training set of 14 compounds and a test set with 10 compounds (to get a complete force field, the division of the two sets is slightly different from that in the original study[ix]). These two sets are shown in Figures 2 and 3.

Quantum mechanical calculations

Reference data for the parameterisations were obtained with quantum mechanical calculations. We used data from our previous work,[ix] which were obtained using DFT

calculations with the B3LYP functional[xxx,xxxi] and the 6-31G* basis set for all atoms except iron, for which we used the double-ξ basis set of Schäfer et al.[xxxii] enhanced with a *p* function with the exponent 0.134915 (DZP). The calculations were performed with the Gaussian03 software package.[xxxiii] The haem model was studied in the quartet state. We employed the geometries, energies, and the Hessian matrix from a frequency calculation of the transition state. All new B3LYP calculations were performed in the same way as the previous ones. The transition-state structures are insensitive to the choice of the model, theoretical method, and basis set used. Therefore, the employed structures are closely similar[ix] to those obtained with other DFT methods.[v,vi,vii,viii]

Parameterisation

The Amber force field has the following functional form:

$$E_{MM} = \sum_{bonds} C_i(b_i - b_{i0})^2 + \sum_{angles} D_i(\alpha_i - \alpha_{i0})^2 + \tfrac{1}{2} \sum_{dihedrals} \sum_{j=1} E_{ij}(1 + \cos(j\varphi_i + \delta_{ij})) + \sum_{atom\_pairs} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \quad (1)$$

In this equation, the energy of a bond depends harmonically on the actual bond length (b_i; b_{i0} is the corresponding ideal bond length and C_i is the corresponding force constant). The same applies to the angles (with the actual and ideal angles α_i and α_{i0}, and the force constant D_i), whereas the dihedral angles (the torsions) are assumed to be described by a cosine function with a periodicity (j) of 1, 2, 3 or 4. E_{ij} is the corresponding force constant and δ_{ij} is a phase factor, which determines the position of the minimum. Non-bonded interactions are described by Coulomb's law between partial charges on each atom (q_i) and a Lennard–Jones potential between each pair of atoms that are more than two bonds apart, using the constants A_{ij} and B_{ij}. Non-bonded interactions between atoms separated by three bonds are scaled down by a factor of 1.2 (electrostatics) or 2.0 (van der Waals), whereas those between atoms separate by one or two bonds are ignored.

To make the force field as general as possible, we decided to use the standard GAFF force field for the substrates.[xxix] Thereby, we avoid the need of reparameterising the force field every time a new substrate will be studied. The atom types of the substrates were determined according to the philosophy of the GAFF

force field[xxix] using the antechamber[xxxiv] module in Amber 8.[xxxv] The only parameters not available in the GAFF force field were for an angle, and they were obtained according to the GAFF analogy rules (atom types c3–c2–f, force constant 66.0 kcal/mole/Å and ideal angle 113.06º). Missing improper torsions were determined with the parmchk module of Amber.[xxxv]

Likewise, we decided to let the carbon and hydrogen atoms of the iron ligand SCH_3^- have the same atom types as cysteine in the Amber 1999 force field.[xxxvi] Thus, we took parameters for its C–H bonds, H–C–H angle, and H–C–S–Fe dihedral from this force field.

Therefore, our parameterisation of the transition state for the CYP hydroxylation is restricted to the haem group and the reactive hydrogen atom of the substrate. Nine new atom types were defined, as is shown in Figure 4. However, all parameters involving the NO and NP atom types were constrained to be identical (these two atom types are needed only to differ between the two types of N–Fe–N angles, 90 or 180º). Atom names and atom types of the full haem group with axial ligands are shown in Figures S1 and S2 in the supplementary material.

The Lennard–Jones parameters for all atoms were taken from the Amber 1999 force field (except for HQ, for which the parameters were taken from the GAFF force field, because this atom is part of the substrate).[xxxvi] Thus, the new van der Waals parameters were taken from the following old atom types: HQ = h3, OQ = OH, NO = NP = N, SQ = S, Cb = Cc = Cd = C*, whereas those for iron were taken from the haem parameters supplied with Amber 8 (r = 1.2 Å and $\sum$ = 0.05 kcal/mole).[xxxv]

Atomic charges for the isolated substrates were obtained by optimising their geometry and calculating electrostatic potential around the molecule in points sampled with the Merz–Kollman scheme[xxxvii] using quantum mechanical calculations at the Hartree–Fock level, following the philosophy of the GAFF force field.[xxix] Charges were fitted to these electrostatic potentials using the RESP method,[xxxviii] as implemented in the antechamber software.[xxxiv]

For the haem group (with substrates), we instead used DFT calculations with the B3LYP functional and the DZP/6-31G* basis set (because the complicated electronic structure of the metal-containing transition states cannot be properly described at the Hartree–Fock level). Charges for all atoms were calculated using the RESP method for all 14 models in the training set. The net charge of the full model, except the

substrate (which varies between the various models), was on average –0.2687 *e* with a mean absolute deviation of 0.05 *e*. The individual charges varied by up to 0.47 *e*, but large differences were observed only for the central iron and nitrogen atoms, which shows that they are caused mainly by the buried-charge problem of the RESP fit (atoms that are buried by other atoms in the structure have no ESP points close to them and therefore are less well-determined than more exposed atoms).[xxxviii] In fact, the well-defined hydrogen atoms varied by less than 0.05 *e* and the moderately buried carbon atoms varied by 0.15 *e* on average.

Our goal was to obtain a force field that can be combined with the GAFF force field for any substrate. Therefore, we need to have charges for all atoms in the haem group and its ligands that are independent of the substrate. We selected the model that had a net charge (excluding the substrate) closest to the mean charge of all the 14 models (–0.2687 *e*) and also had the smallest mean absolute deviation of all atoms compared to the average. This was the model with isobutane. We used the charges of this model for all atoms in the haem group. For the substrates, the GAFF charges were used, but a charge of +0.2679 *e* was added to the reactive hydrogen atom to make the full complex neutral.

In the DFT calculations,[ix] the side chains of the haem group were replaced by hydrogen atoms to reduce the calculation times. In order to get parameters for a complete haem group, parameters are needed also for the side chains. It is unlikely that these groups should significantly influence the parameters around reactive centre. Therefore, we simply added bonded parameters for the side chains from the GAFF force field by analogy. Charges for the side chains were determined by first running a geometry optimisation of the isobutane model with the side chains added (the orientation of haem ring was selected so that the Cys ligand and the two propionate side chains point in the same direction, as is observed in all available crystal structures[xxxix]). During this optimisation, the coordinates of the central core were fixed to those of the model without side chains and the dihedrals of the side chains relative to the porphyrin ring were fixed to those observed in the crystal structure of human CYP 2C9[xl] (to avoid that the charged propionate groups curl inwards and form unphysical interactions with the haem group). With these coordinates, the charges were calculated with the RESP method as described above, keeping the charges of the core model fixed to their previous values. The resulting charges of the full haem

group are shown in Table S1 in the supplementary material. The force field can, through the parametrisation from DFT data, reproduce the average oxidation state of a hydrogen abstraction transition state, but it cannot model major changes in the electronic structure, e.g. if the transition state becomes much more product- or reactant-like than in the average case, or if the oxidation state of the transition state changes significantly.

Finally, the bonded parameters (bond, angle, and dihedral terms in Eqn. 1) for all interactions involving any of the new atom types were determined by the Q2MM method,[xxxix] as is described in the next section. This method requires start values for all parameters. These should be as realistic as possible, to make sure that we do not end up in non-physical local minima during the optimisation. We obtained start values in the following manner: Parameters in the haem group were taken from the set of haem parameters supplied with Amber 8.[xxxv] The parameters for the SCH_3^- group were taken from the cysteine parameters of the Amber 1999 force field.[xxxvi] Force constants of the angles including both the reacting hydrogen and carbon (atom types HQ and c3) were taken from the GAFF force field by analogy. The force constants of the iron–sulphur, iron–oxygen and oxygen–hydrogen bonds and related angles were started from reasonable guesses.

A few parameters were excluded from the parameterisation: First, all dihedrals containing an angle larger than 150º were excluded by zeroing the force constant (i.e. the dihedral A–B–C–D was excluded if either of the angles A–B–C or B–C–D was larger than 150º). These included the OQ–HQ–c3–X dihedral in the reactive bond, because the OQ-HQ-c3 angle is normally close to straight, and all dihedrals involving the two straight N–Fe–N angles of haem (atom types NO–FE–NO and NP–FE–NP). This is necessary, because the dihedral becomes undefined when one of the angles becomes straight, and close to that point, the dihedral can vary extensively for small changes in the position of the atoms. The excluded dihedral angles were selected at the start of the parameterisation and were then kept fixed.

Second, the X–NO/NP–Fe–X dihedrals were also excluded by zeroing the force constants, because they approached zero during the optimisation and often caused instabilities.

Third, some data were excluded from the parameterisation because they provided too much noise. This applies to the dihedrals N–Fe–O–H and N–Fe–S–C, because the

variation in these two dihedrals is large among the 14 substrates. This would give a low force constant in a parameterisation of those dihedrals, and therefore, the optimisation may randomly end up in different local rotational minima, which gives problems in the parameterisation.

Q2MM implementation with Amber

The parameterisation was performed with the Q2MM approach,[xxi] which we have interfaced with Amber. This method tries to minimise the deviation of geometries and Hessian elements between the DFT and MM data using a penalty function that gives different weights to different kinds of data. The geometries were described as lists of all bonds, angles, and dihedral angles, rather than by absolute positions. The weight factors of the various data types were 100 $Å^{-1}$ for bonds, 2 $degree^{-1}$ for angles, 1 $degree^{-1}$ for torsions, and 0.01–0.1 mole $Å^2$/kcal for Hessian elements (0.01 for elements involving interactions of an atom with itself, 0.02 for atoms bound to each other, 0.04 for atoms connected by two bonds, 0.1 for atoms connected by three bonds, and 0.01 for all other elements).[xxi] In the following, we will describe the changes made compared to the original implementation.

To generate structural data and start files for MM minimisation, we read PDB files of the DFT structures into the Amber module tleap and generated coordinate and topology files. The coordinate files were used as reference structures. Using local programs, we extracted all bond lengths, angles, and dihedrals using these coordinate and topology files. Reference Hessians were taken from the Gaussian03 output files and the negative eigenvalue was adjusted to a large positive value (1.0 a.u.) before use, to enable us to model the transition state as a minimum.[xxi,xli]

The structures were then minimised, either by the conjugate gradient method in the sander module or by Newton–Raphson method in the nmode module of Amber.[xxxv] In this application, we have used nmode, because it is more robust and very seldom fails, but it is free to the user to decide what program to use. Finally, the Hessian matrix of the minimised structure was calculated by the nmode module. It was necessary to modify nmode to write out the forces and the mass-weighted Hessian matrix to a file.

The implementation of Q2MM for Amber is available from the authors on request. A detailed description of its use is available in

http://www.teokem.lu.se/~ulf/Methods/ponparm.html.

Results and Discussion

Molecular mechanics parameters

The final transition-state force field parameters are reported in the supplementary material (Tables S1–S4). Figure 5 shows the relation between the DFT and MM Hessian elements. It is worth noting that there are over 180 000 data points in this figure, and the great majority of these are found along the diagonal. The correlation coefficient is 0.986. Considering that the force field is a compromise of 14 different structures, the fit is impressive and of a quality similar to previous parameterisations with the Q2MM method.[xlii,xliii,xliv] However, there is a hint of a straight line along the *x* axis (i.e. with the MM Hessian = 0). These points represent interactions that cannot be described by the MM force field in Eqn. 1 (e.g. trans-effects[xlv] or torsions across the metal), and therefore no parameter modifications can improve their fit.

Of the optimised parameters, two force constants for angles are quite high. These are the constants for the angles NO–FE–NP and FE–SQ–CT, which are 608 and 348 kcal mole^{-1} rad^{-2} respectively. One could argue that these are too high and force them to have a lower value, but this gave much larger errors for the N–Fe–N and N–Fe–S angles. Therefore, we have chosen to keep them at these optimised values. Most probably, they compensate for differences in the description of non-bonded forces between MM and QM methods (remember that the Amber van der Waals parameters and the RESP charges were *not* varied in the parameterisation).

Reproducing geometries

Geometries obtained after minimisation with our optimised force field reproduce the DFT geometries quite well, disregarding parts of the structures that are described only by the GAFF force field (the parameters determining these parts were not optimised). When including all data (except dihedrals with an angle larger than 150º), the root-mean-squared deviation (RMSD) of bonds, angles and dihedrals are 0.010 Å, 1.17°, 7.1° for the training set and 0.012 Å, 1.25°, 8.2° for the test set (the full data set is shown in Table 1). If we also exclude the dihedrals with a force constant of zero and those that are described by the GAFF force field, the average RMSD of the

dihedrals drops to 2.1º for both the training and test sets. This shows that the parameterised part of the structures is excellently described. An overlay of the structures with the highest and lowest RMSD (of the atomic positions, disregarding the substrate) are shown in Figure 6. It can be seen that the haem ring is almost perfectly reproduced, whereas the dihedrals to the substrate and to the cysteine ligand, show larger discrepancies.

The reactive centre around the oxoferryl group and the reactive hydrogen atom of the substrate is the most important part of the transition-state structure, but it is also the part that is hardest to describe with the force field, because it is a transition state and the DFT structures show quite extensive variations for the bond lengths and angles around the reactive hydrogen atom. In Table 2 we illustrate how well the crucial bonds and angles around the reactive hydrogen are reproduced by the force field. It can be seen that the errors of the O–H and H–C bonds, as well as the O–H–C angle are somewhat larger than the average errors for all bonds and angles (in Table 1), but they are still of the same size as the variation among the DFT structures, which is as good as one can expect to get when using many different structures in a parameterisation.

Specificity versus generality

In order to test how much the force field is deteriorated by the use of structures from many different substrates, we made a separate force field, based on a single DFT structure (model 4, isobutane). This gave RMSDs of 0.008 Å (bonds), 1.08° (angles) and 3.10° (dihedrals, 1.98° when excluding the substrate dihedrals). This is actually not significantly better than for the general force field for the bonds and angles (0.007 Å, 1.09º, and 3.37°/1.78º, cf. Table 1). Looking at the specific bonds and angles around the reaction centre, we get the following deviations; Fe–O +0.003 Å, O–H +0.014 Å, H–C +0.032 Å, Fe–O–H +0.96°, and O–H–C +0.29°. These errors are similar to those in the full parameterisation except for the Fe–O–H, angle which for this system had an error of –3.35° in the full parameterisation. Thus, there is no significant gain in performing a separate parameterisation for each substrate.

Energies

Next, we tested if the MM optimised structures can be used to calculate DFT

energies, without any further optimisation. Unfortunately, this was not possible; single-point DFT energy calculations on the MM optimised structures gave 8–41 kJ/mole higher energies than those obtained from the DFT optimised structures (Table 1, column $\otimes E$). This shows that even if the force field gives quite good geometries, one still needs to run a DFT geometry optimisation to get reliable DFT energies. Interestingly, the major difference in energy comes from the substrates, which are not parameterised: If the DFT energies were calculated only for the haem groups, with the substrate converted to methane for models 4, 10, 19, and 23, $\otimes E$ was reduced from 14–37 kJ/mole to 7–15 kJ/mole (note also that the smallest value for $\otimes E$, 8 kJ/mole. is obtained with the smallest substrate, methane).

To make sure that there is not a significant contribution to $\otimes E$ from dispersion interactions (which are present in MM, but poorly described by DFT), we also did a parameterisation of model 19 (propen-2-ol), in which we also included all substrate parameters and dihedrals into the parameterisation. The resulting $\otimes E$ was only 9 kJ/mole (37 kJ/mole for the general parameterisation), which verifies that the substrate parameters cause the major part of $\otimes E$.

Conclusions

In this paper, we describe an implementation of the ideal and automatic parameterisation method by Norrby and Liljefors[xli] for the widely used Amber software.[xxxv] This method takes into account all interactions (bonded, as well as non-bonded) in a self-consistent way during the iterative parameterisation and therefore provides the best possible structure, given the functional form of the force field (Eqn. 1). In particular, it provides appreciably better structures than methods that try to extract the bonded parameters in the force field directly from the Hessian matrix[xlvi] (without taking into account that the Hessian elements also contain contributions from all non-bonded interactions). The method minimises a penalty function consisting of the squared deviation of the optimised force-field structures from the reference values (in our case obtained by DFT calculations) for all bonds, angles and dihedrals, as well as the elements of the Hessian matrix, all properly weighted according to the acceptable error of each type of data.[xli] The method is fully automatic, but it must be carefully checked that one does not end up in an unphysical local minimum. Of

course, it is more time-consuming than simpler methods – for the present complicated application, a full optimisation of all the parameters typically took about one week.

Using this method, we have constructed a general force field for the transition state of the hydrogen abstraction from sp^3 carbons in the cytochromes P450. Transition states cannot be treated with standard MM methods.[xiv] Therefore, we have used the Q2MM approach,[xli] in which the transition state is treated as a normal minimum by switching the negative eigenvalue to a large positive value. Thereby, we can use a standard MM program to obtain structures of the transition state and essentially all starting structures will converge to the transition state.

In order to obtain a force field that is as general as possible, we have used structures of transition states for 14 different substrates (Figure 2), which is an unusually large set of structures for a force-field parameterisation. Still, the results are impressive: The MM optimised bond lengths and angles reproduce those obtained by DFT with an average RMS deviation of only 0.01 Å and 1.2º. The dihedrals of the porphyrin ring are equally well reproduced with an average RMS error of 2.1º. In most cases, our force field gives errors of the same size as the variations in the input data, which is as good as possible. However, for the dihedrals between the haem group and the substrate, the result is worse, because the 14 structures in the training set show large variations for these torsions. In real applications of the force field in proteins, this problem is less serious, because the torsions are low-energy modes and their actual values are typically dictated by the surrounding protein.

No attempt has been made to optimise the force field for the substrates, because we intended to keep the parameters as general as possible, to allow simulations of any drug-like molecule. Therefore, we have used the standard GAFF force field for the substrates.[xxix] Of course, this leads to worse results for the substrates, but the results are not worse than in a normal use of the GAFF force field.

The use of an unusually large number of structures in the training set has ensured that a versatile force field is obtained. In fact, there is no significant difference in the performance of the force field for the training and test sets (Tables 1 and 2). We have not observed any conflicts between the various structures, except for the above mentioned dihedrals. For example, the RESP charges for the 14 structures differed by less than 0.05 *e* for the exposed (i.e. not buried) hydrogen atoms. Most importantly, a separate parameterisation for a single structure did not give any significantly

improved force field.

Thus, the new force field gives excellent structures of the transition state for aliphatic hydrogen abstraction. Unfortunately, the structures are still not good enough to give accurate energies – single-point DFT calculations on the MM structures give ~24 kJ/mole higher energies than on the DFT structures. The majority of this difference comes from the substrate, which is treated with the original GAFF force field. However, even if the substrate is also parameterised, the difference is still 9 kJ/mole, showing that it is very hard to obtain accurate DFT energies from MM structures, obtained with a force field with the simple functional form in Eqn. 1.

We see many uses of our transition state force field. First, it can be used to rapidly obtain starting structures to DFT optimisations of transition states for other substrates. There is a great interest of predicting the reactivity of drug candidates with the CYPs and the intrinsic reactivity of the drugs are most accurately determined by DFT methods.[ix] Unfortunately, such calculations are very time-consuming (about one week for the small substrates in Figures 2–3). Good starting structures can make such calculations much faster.

Second, the force field can be used for molecular dynamics simulations of the transition state, e.g. to study how it may be stabilised by the protein. The Q2MM force field shares many properties with the empirical valence bond model potential.[xvi] For example, it will be possible to explore variations in the transition state through molecular dynamics or conformational searching. This will not in itself give free energies of activation, but since the transition-state cross-section is faithfully reproduced, the vibrational and conformational contributions to the free energy should be obtainable from the force field.

Third, the new force field can be used to dock various molecules into a CYP enzyme. Such docking studies will show if a drug candidate sterically fits into the active site and therefore may be a substrate of the enzyme (to be a substrate, the drug must pass the transition state). Such a transition-state docking would provide a more restrictive and therefore more discriminative test than a standard docking of the ground state of the drug candidate to an arbitrary state of the enzyme.

As with other MM methods, energies obtained with the present force field are comparable only if the models contain exactly the same bonded and non-bonded interactions. Therefore, it is in general not possible to compare the reactivity of

different atoms on a substrate (i.e. the regioselectivity). However, for the special case of hydrogen atoms bound to the same carbon atoms in a substrate, the interactions are the same and the energies are comparable. Therefore, our Q2MM force field can directly be applied to study this special case of regioselectivity. Moreover, it should be possible to study differences in the regioselectivity in different CYP isoforms by studying proper energy differences. However, to compare reactions at different sites in a substrate molecule, the docking energies must be combined with estimates of the intrinsic reactivity of each site, obtained by DFT or other quantum mechanical methods.[ix] Investigations along these lines are currently performed in our laboratory.[xlvii]

Acknowledgements

This investigation has been supported by funding from the research school in pharmaceutical science (FLÄK), the Swedish Research Council, the Carlsberg Foundation, and by computer resources of LUNARC at Lund University.

Supporting Information Available

The final transition-state force field for the aliphatic hydroxylation in cytochrome P450 (charges, bond, angle, and dihedral parameters), as well as figures of the atom names and atom types used. In addition, Amber topology and parameter files of the force field are included. This information is available free of charge via the Internet at http://pubs.acs.org.

Table 1. Root-mean-squared deviation for all bonds, angles, and dihedral angles (in Å and degrees, respectively) between the structures optimised with DFT and with the force field. Data for both the training and test sets are given. Dihedrals with zeroed force constants, any angle larger than 150º, or only consisting of substrate atoms are not included. $\otimes E$ is the energy difference (in kJ/mole) of the DFT and MM structures calculated at the B3LYP/DZP/6-31G* level. The model numbers refer to Figures 2 and 3.

Training set					**Test set**				
Model	Bonds	Angles	Dihedrals	ΔE	Model	Bonds	Angles	Dihedrals	ΔE
3	0.008	1.17	1.93	13.2	1	0.017	1.13	2.28	7.8
4	0.007	1.09	1.78	13.7	2	0.010	1.16	2.03	12.2
5	0.010	1.30	2.07	22.2	9	0.009	1.20	1.70	22.3
6	0.013	1.21	2.03	29.2	12	0.010	1.36	2.16	41.0
7	0.009	1.13	2.06	18.3	13	0.016	1.15	2.03	22.7
8	0.010	1.14	1.95	25.9	15	0.017	1.25	2.11	16.1
10	0.012	1.19	2.51	16.6	18	0.010	1.70	2.29	25.1
11	0.009	1.07	2.49	18.3	21	0.011	1.24	1.98	36.6
14	0.015	1.13	2.26	25.9	22	0.010	1.34	2.36	36.1
16	0.011	1.25	2.29	16.6	24	0.011	0.98	2.30	30.1
17	0.010	1.00	2.22	12.3					
19	0.011	1.61	2.10	36.9					
20	0.009	1.09	1.98	34.1					
23	0.011	1.03	2.22	35.4					
Average	**0.010**	**1.17**	**2.14**	**22.8**		**0.012**	**1.25**	**2.12**	**25.0**

Table 2. Geometry around the reaction centre compared to DFT data. The bonds are in Å and the angles are i degrees.

	Training set				Test set			
	DFT av[a]	MAD[b]	MM err[c]	MM MAD[d]	DFT av[a]	MAD[b]	MM err[c]	MM MAD[d]
Fe–O	1.743	0.010	0.009	0.011	1.745	0.011	0.007	0.012
O–H	1.238	0.025	0.006	0.026	1.219	0.040	0.024	0.046
H–C	1.316	0.017	0.012	0.022	1.333	0.027	-0.008	0.033
Fe–O–H	122.43	1.51	0.82	1.93	122.09	1.64	1.03	1.90
O–H–C	168.49	3.83	-0.84	2.87	168.69	2.99	-1.93	3.56

[a] Average of all DFT structures

[b] Mean absolute deviation of the DFT structures.

[c] Average deviation of the MM optimised structures compared to the corresponding DFT structure.

[d] Mean absolute deviation of the MM optimised structures compared to the corresponding DFT structure.

Figure 1. The CYP reaction cycle for a hydroxylation reaction, including the studied transition state and the intermediate state after it. The N quadrant represents the porphyrin ring.

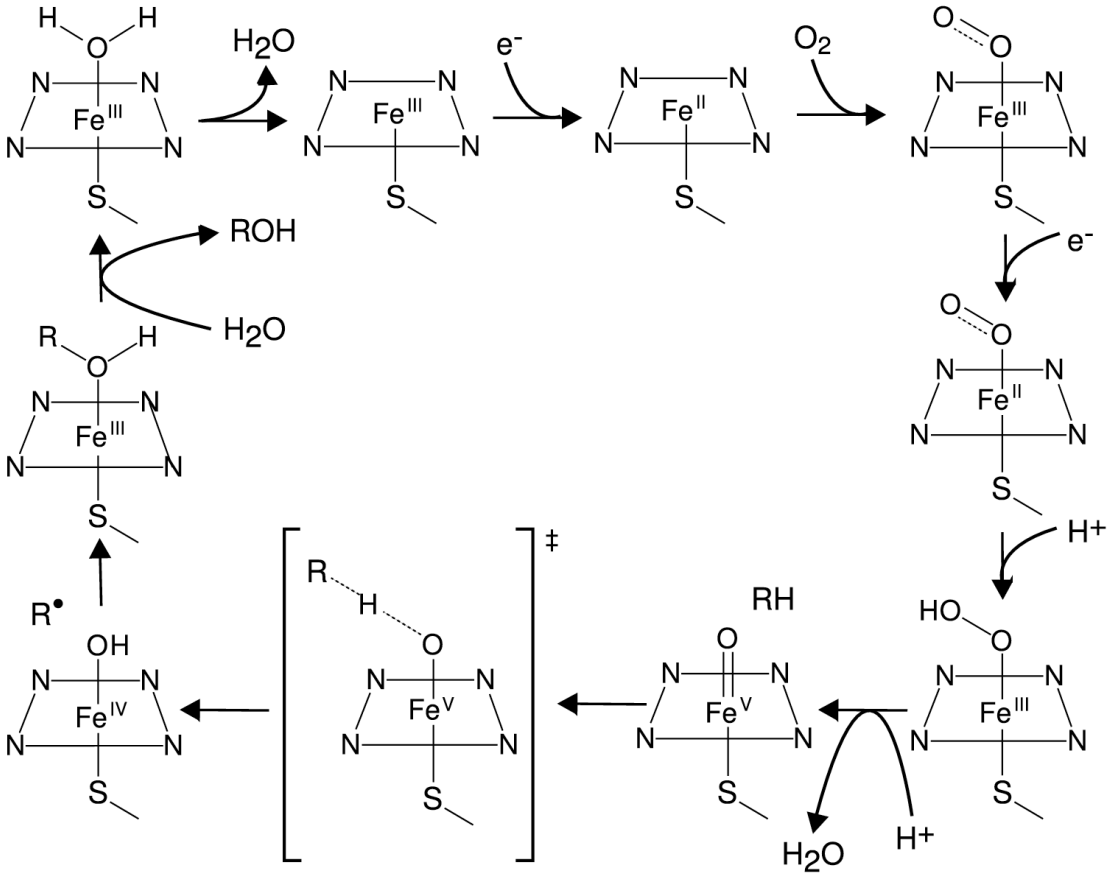


Figure 2. Substrate models used in the training set

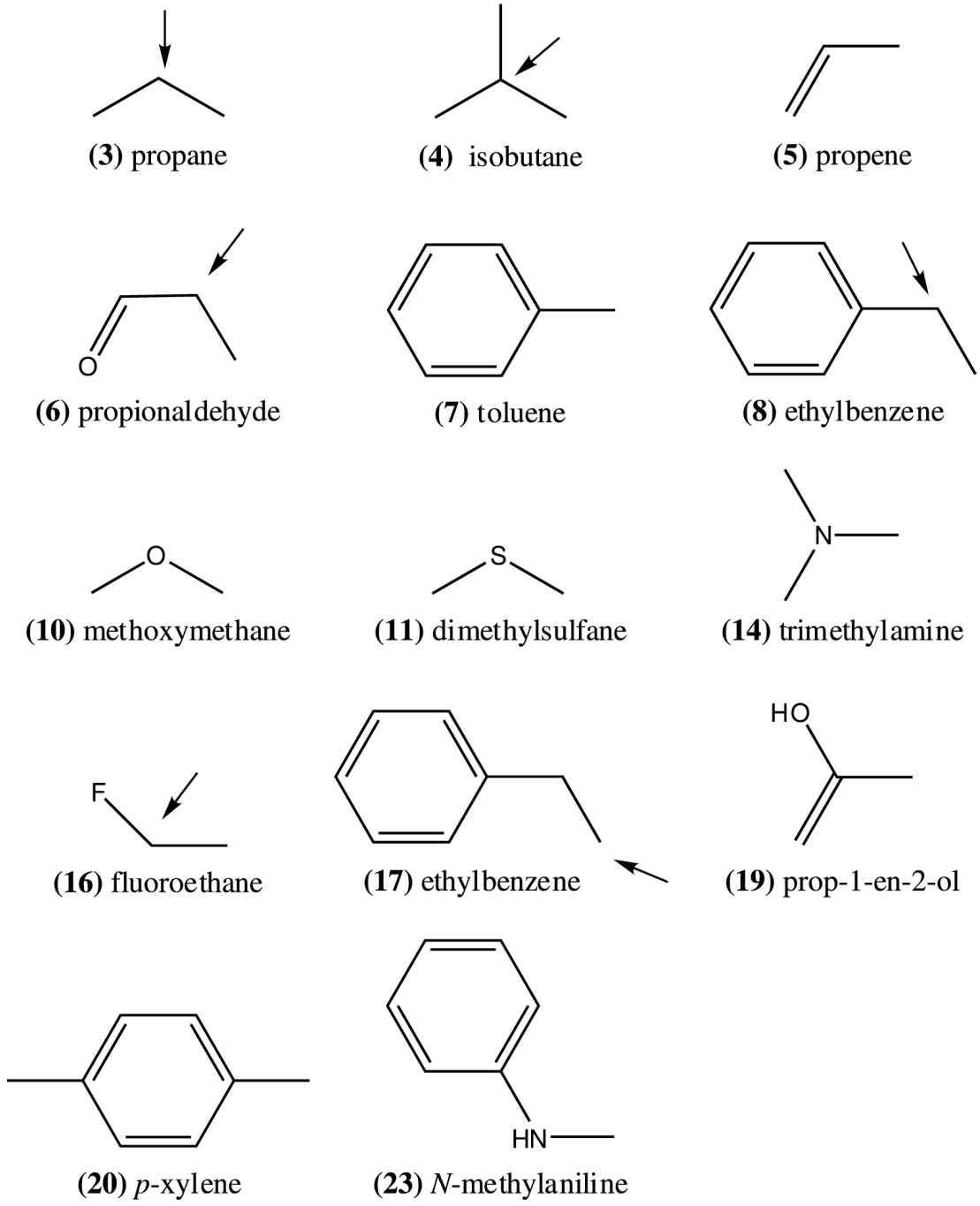
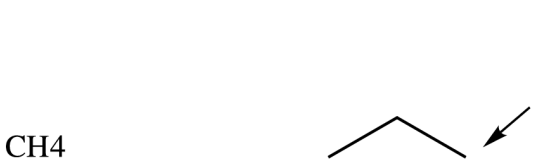
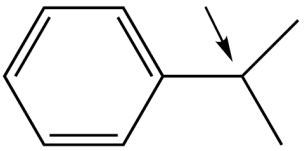


Figure 3. Substrate models used in the test set.

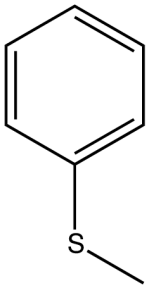


(1) methane

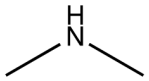
(2) propane



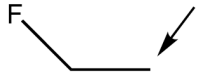
(9) 1-methylethylbenzene



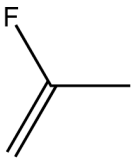
(12) methyl(phenyl)sulfane



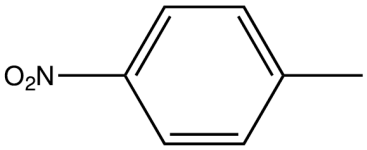
(13) dimethylamine



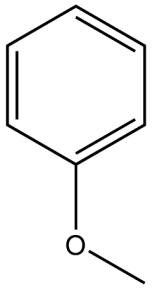
(15) fluoroethane



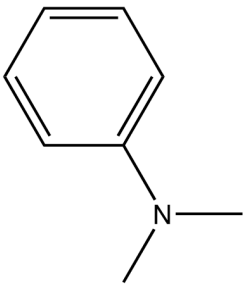
(18) 2-fluoroprop-1-ene



(21) 1-methyl-4-nitrobenzene



(22) methoxybenzene



(24) *N*,*N*-dimethylaniline

Figure 4. The new atom types.

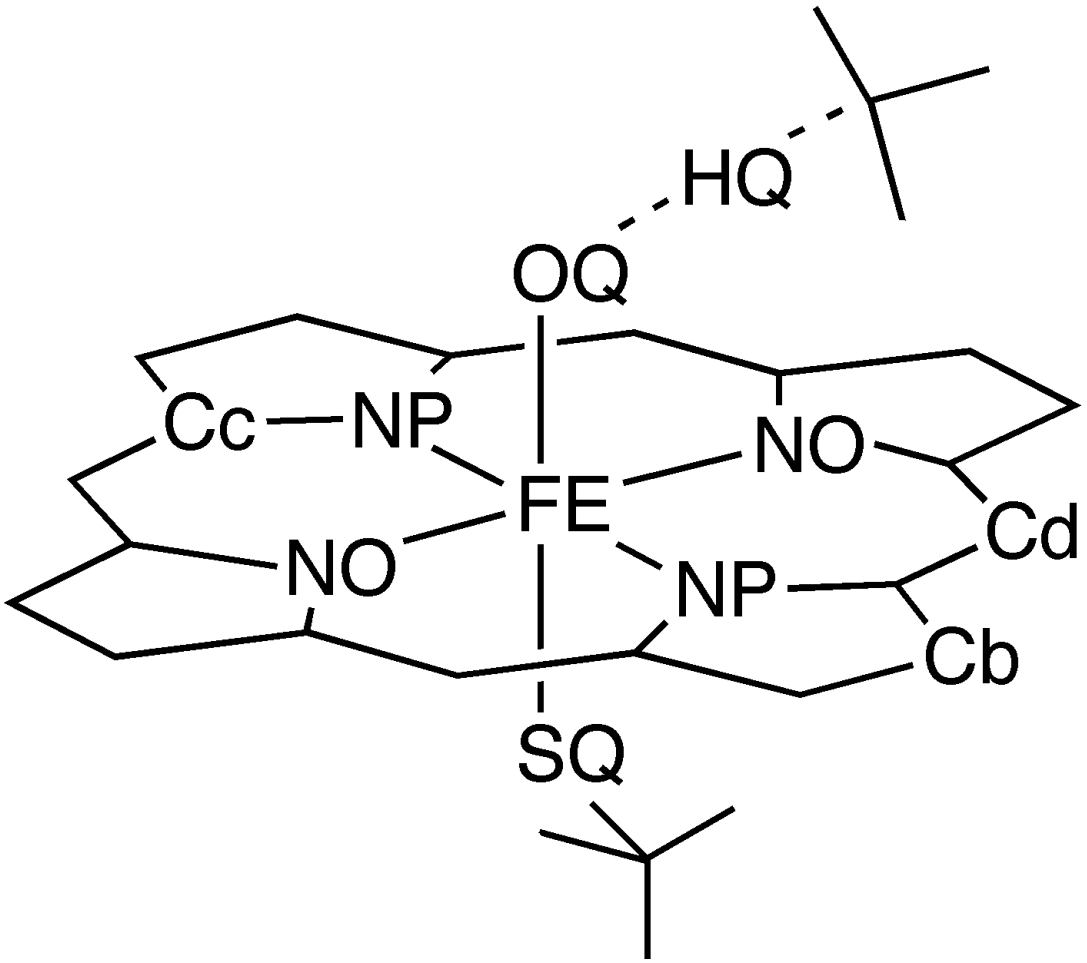


Figure 5. A comparison of the DFT and MM mass-weighted Hessian elements (in kcal mole^{-1} Å^{-2} amu^{-1}).

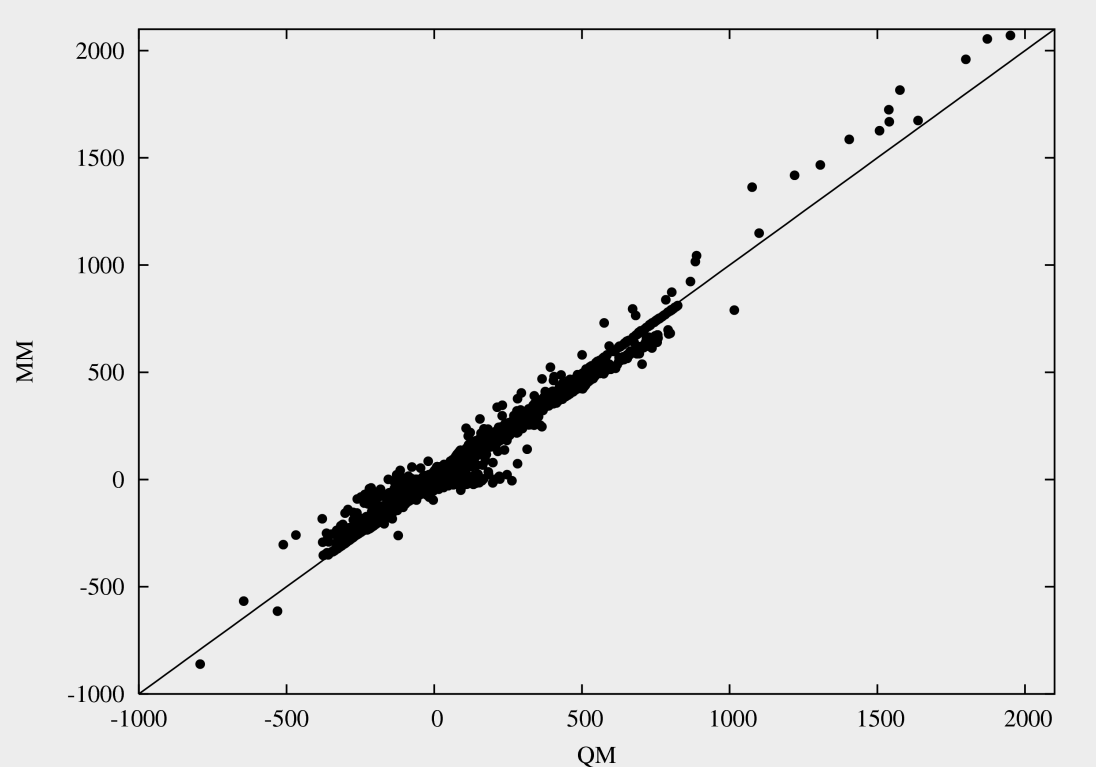
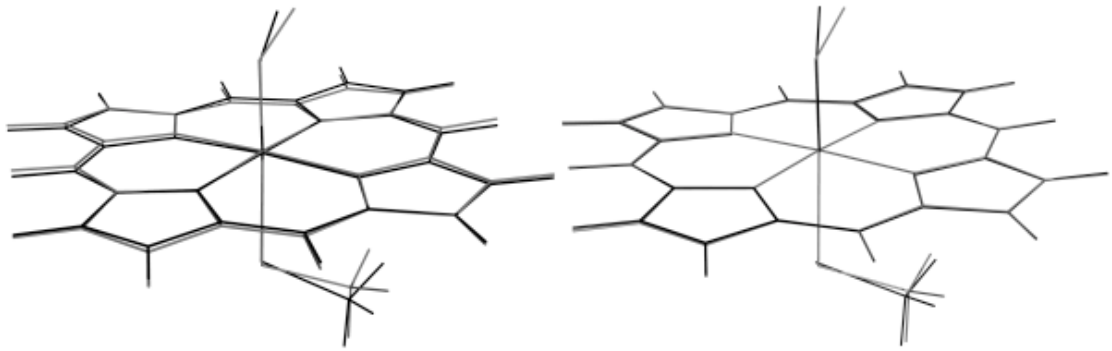


Figure 6. Overlay of the parameterised part of structures 12 and 24, which have RMSDs of 0.10 Å and 0.05 Å, respectively, for this part of the structure (the highest and lowest RMSD of all 24 structures). The DFT structures are black and the MM structures are grey.



[i] Bertz, R. J.; Granneman G. R. *Clin. Pharmacokinet.* **1997**, *32*, 210–258.
[ii] Evans, W. E.; Relling, M. V. *Science* **1999**, *286*, 487–491.
[iii] Rydberg, P; Sigfridsson, E.;Ryde, U. *J. Biol. Inorg. Chem.* **2004,** *9*, 203-223.
[iv] Meunier, B.; de Visser, S. P.; Shaik, S. *Chem. Rev.* **2004**, *104*, 3947-3980.
[v] Shaik, S.; Kumar, D.; de Visser, S. P.; Altun, A.; Thiel, W. *Chem. Rev.* **2005**, *105*, 2279-2328.
[vi] Hirao, H.; Kumar, D.; Thiel, W.; Shaik, S. *J. Am. Chem. Soc.* **2005**, *127*, 13007-13018.
[vii] de Visser, S. P.; Kumar, D.; Cohen, S.; Shacham, R.; Shaik, S. A. *J. Am. Chem. Soc.* **2004**, *126*, 8362-8363.
[viii] Park, J. Y.; Harris, D. J. *Med. Chem.* **2003**, *46*, 1645-1660.
[ix] Olsen, L.; Rydberg, P.; Rod, T. H.; Ryde, U. *J. Med. Chem.* **2006**, *49*, 6489-6499.
[x] de Graaf, C.; Vermeulen, N. P. E.; and Feenstra, K. A. *J. Med. Chem.* **2005,** *48*, 2725-2755.
[xi] Jones, J. P.; Korzekwa, K. R. *Methods Enzymol.* **1996,** *272*, 326-335.
[xii] Vermeulen, N. P. E. *Current Topics in Medicinal Chemistry* **2003,** *3*, 1227-1239.
[xiii] Hermann, J. C.; Ghanem, E.; Li, Y.; Raushel, F. M.; Irwin, J. J.; Shoichet, B. K. *J. Am. Chem. Soc.* **2006**, *128*, 15882-15891.
[xiv] Jensen, F.; Norrby, P.-O *Theor. Chem. Acc.* **2003,** *109*, 1-7.
[xv] Warshel, A. J. Am. Chem. Soc. **1980**, 102, 6218
[xvi] Åqvist, J.; Warshel, A. Chem. Rev. **1993**, 93, 2523
[xvii] Rappé, A.K.; Pietsch, M.A.; Wiser, D.C.; Hart, J.R.; Bormann, L.M.; Skiff, W.M. Mol. Eng. **1997**, 7, 385
[xviii] Kim, Y.; Corchado, J.C.; Villa, J.; Xing, J.; Truhlar, D.G. J. Chem. Phys. **2000**, 112, 2718
[xix] van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; Goddard, W. A. J. Phys. Chem. A **2001**, 105, 9396
[xx] Jensen, F. J. Comput. Chem. **1994**, 15, 1199

[xxi] Norrby, P.-O. J. Mol. Struct. (Theochem) **2000**, 506, 9
[xxii] Eksterowicz, J. E.; Houk, K. N. *Chem. Rev.* **1993**, *93*, 2439.
[xxiii] Oda, A.; Yamaotsu, N.; Hirono, S. *J. Comput. Chem.* **2005**, *26*, 818-826.
[xxiv] Kuczera, K.; Kuriyan, J.;Karplus, M. *J. Mol. Biol.* **1990**, *213*, 351-373.
[xxv] Banci, L.; Bertini, I.; Bren, K. L.; Gray, H. B.; Sompornpisut, P.; Turano, P. *Biochemistry* **1995**, *34*, 11385-11398.
[xxvi] Autenrieth, F.; Tajkhorshid, E.; Baudry, J.; Luthey-Schulten, Z. *J. Comp. Chem.* **2004**, *25*, 1613-1622.
[xxvii] Ma, J.-G.; Zhang, J.; Franco, R.; Jia, S.-L.; Moura, I.; Moura, J. J. G.; Kroneck, P. M. H.; Shelnutt, J. A. *Biochemistry* **1998**, *37*, 12431-12442.
[xxviii] Park, H.; Lee, S. *J. Comput. Aid. Mol. Des.* **2005**, *19*, 17-31.
[xxix] Wang, J., Wolf, R. M.; Caldwell, J. W.;Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157-1174.
[xxx] Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648-5652.
[xxxi] Stephens, P. J.; Devlin, F. J.; Chabalowski, C. F.; Frisch, M. J. *J. Phys. Chem.* **1994**, *98*, 11623-11627.
[xxxii] Schafer, A.; Horn, H.; Ahlrichs, R. *J. Chem. Phys.* **1992**, *97*, 2571-2577.
[xxxiii] Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. Gaussian 03, Revision C.02; Gaussian, Inc.: Wallingford, CT, 2003.
[xxxiv] Wang, J., Wang, W., Kollman P. A.; Case, D. A. *J. Mol. Graph. Model.* **2006**, *25*, 247-260.
[xxxv] Case, D. A.; Darden, T. A.; Cheatham, III, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M.; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C.; Caldwell, J. W.; Ross, W. S.;Kollman, P. A. *AMBER 8*, **2004**, University of California, San Francisco.
[xxxvi] Wang, J.; Cieplak, P.; Kollman. P. A. *J. Comput. Chem.* **2000**, *21*, 1049-1074.
[xxxvii] Besler, B. H.; Merz, K. M.; Kollman, P. A. *J. Comput. Chem.* **1990**, *11*, 431-439.
[xxxviii] Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269-10280.
[xxxix] Li, H, Cytochrome P450 In: Messerschmidt A, Huber R, Poulos T, Wieghart K (eds) Handbook of metalloproteins. Wiley, Chichester, pp. 267-282.
[xl] Yano, J. K.; Wester, M. R.; Schoch, G. A.; Griffin, K. J.; Stout, C. D.; Johnson. E. F. *J. Biol. Chem.* **2004**, *279*, 38091-38094.

[xli] The Q2MM has an incorrect response to increase of steric strain along the reaction coordinate.[xxi] The problem can be somewhat alleviated by minimising distortions along the reaction coordinate. This is accomplished by selection of a large value for the replacement eigenvalue in the Hessian modification step, in effect freezing movement along the reaction coordinate, and thus minimising the systematic error inherent in the method.[xxi] However, for cases where the systematic error is expected to influence the results significantly, the Q2MM force field should only be used as a conformational search tool, with final results obtained from a method with a correct response to changes in exothermicity, like the SEAM method.[xx]

[xlii] Norrby, P.-O.; Brandt, P.; Rein, T. *J. Org. Chem.* **1999**, *64*, 5845-5852.

[xliii] Norrby, P.-O.; Rasmussen, T.; Haller, J.; Strassner, T.; Houk, K. N. *J. Am. Chem. Soc.* **1999**, *121*, 10186-10192.

[xliv] Rasmussen, T.; Norrby, P.-O. *J. Am. Chem. Soc.* **2003**, *125*, 5130-5138.

[xlv] Brandt, P.; Norrby, T., Åkermark, B.; Norrby, P.-O. Inorg Chem **1998,** 37, 4120-4127.

[xlvi] Seminario, J. M. *Int. J. Quantum Chem.* **1996**, 6*0*, 1271-1277.

[xlvii] Rydberg, P. Theoretical studies of cytochrome P450, Ph.D. Thesis, Lund University, 2007.