

Computer-based studies on enzyme catalysis

From structure to activity

Promotor: dr. ir. I.M.C.M. Rietjens
Persoonlijk hoogleraar, Laboratorium voor Biochemie

Co-promotor: dr. ir. J. Vervoort
Universitair Hoofddocent, Laboratorium voor Biochemie

NH02701, 2242

Lars-Olaf Ridder

Computer-based studies on enzyme catalysis

From structure to activity

Proefschrift

ter verkrijging van de graad van doctor
op gezag van de rector magnificus
van Wageningen Universiteit,
dr. ir. L. Speelman,
in het openbaar te verdedigen
op dinsdag 3 oktober 2000
des namiddags te vier uur in de Aula.

2 0 0 0 1 0 1 0

The research described in this thesis was carried out at the Laboratory of Biochemistry, Wageningen University, The Netherlands. This work was supported by the Netherlands Organisation for Scientific Research (NWO)

ISBN 90-5808-252-0

BIBLIOTHEEK
LANDBOUWUNIVERSITEIT
WAGENINGEN

Stellingen

- 1 Ondanks significante verschillen in de structuur van de katalytische centra van phenol hydroxylase en *p*-hydroxybenzooat hydroxylase, komt het mechanisme van substraat-activering in beide enzymen overeen.
Dit proefschrift, hoofdstuk 8
- 2 QSARs die gebaseerd zijn op gecombineerde quantummechanische / molecuair mechanische berekeningen kunnen gebruikt worden voor sterker uiteenlopende reeksen substraten dan QSARs die gebaseerd zijn op berekeningen in vacuüm.
Dit proefschrift, hoofdstuk 5 en 7
- 3 Het hangt van de temperatuur af welke stap in het reactie-mechanisme van catechol-1,2-dioxygenase snelheidsbeperkend is.
Dit proefschrift, hoofdstuk 4
Walsh en Ballou (1983) J. Biol. Chem. 258, 14422-14427
- 4 Het verschil in substraatspecificiteit tussen type I en type II catechol-dioxygenasen wordt voor een belangrijk deel veroorzaakt door een verschil in het effect van substraatsubstituenten op de zuurstofbinding.
Dit proefschrift, hoofdstuk 4,
Dorn en Knackmuss (1978) Biochem. J. 174, 85-94
- 5 Wanneer intermediären in enzymatische reacties te instabiel zijn om experimenteel waargenomen te kunnen worden, zijn gecombineerde quantummechanische / molecuair mechanische modellen het beste middel tot hun karakterisering.
Dit proefschrift
- 6 Experimentele data zijn onmisbaar in theoretisch onderzoek.

- 7 De toenemende snelheid van computers zal in het quantumchemisch onderzoek niet leiden tot een kortere duur van de uitgevoerde berekeningen.
- 8 In het onderzoek aan toxische stoffen dient optimaal gebruik te worden gemaakt van computersimulaties, opdat schade aan proefdieren, onderzoekers en milieu geminimaliseerd wordt.
- 9 Een betere implementatie van nieuwe technologie in de samenleving vraagt eerder een extra inspanning op politiek dan op wetenschappelijk vlak.
- 10 De toegenomen populariteit van het lange-afstandswandelen in Nederland is zowel een negatieve reactie op, als een positief gevolg van de grote mobiliteit in dit land.
- 11 Succes in de voetbalpool is niet gecorreleerd aan kennis op dit vlak.
- 12 Wetenschap is falsificeerbare kunst.

Stellingen behorende bij het proefschrift "Computer-based studies on enzyme catalysis; from structure to activity" van Lars Ridder, Wageningen Universiteit, 3 oktober 2000

Contents

1	Introduction: computer-based QSAR studies on enzymes involved in degradation of aromatic compounds	1
2	Introduction to the theoretical principles of molecular orbital theory and its combination with molecular mechanics	19
3	Computational methods in flavin research	31
4	Quantitative structure activity relationship for the rate of conversion of C4-substituted catechols by catechol-1,2-dioxygenase from <i>Pseudomonas putida</i> (arvilla) C1	53
5	Correlation of calculated activation energies with experimental rate constants for an enzyme catalysed aromatic hydroxylation	71
6	Combined quantum mechanical and molecular mechanical reaction pathway calculation for aromatic hydroxylation by p-hydroxybenzoate-3-hydroxylase	79
7	Modelling flavin- and substrate substituent effects on the activation barrier and rate of oxygen transfer by p-hydroxybenzoate hydroxylase	103
8	A quantum mechanical/molecular mechanical study of the hydroxylation of phenol and halogenated derivatives by phenol hydroxylase	115
9	Summary and conclusions	145
	Samenvatting en conclusies	151
	Symbols and abbreviations	158
	References	159
	Samenvatting voor niet-vakgenoten	169
	Dankwoord	173
	List of publications	175
	Curriculum Vitae	177

1

Introduction: computer-based QSAR studies on enzymes involved in degradation of aromatic compounds.

A major challenge in biochemistry is to understand how enzymes catalyse the specific conversion of their substrates, on the basis of the structures of the enzyme and the substrates involved. For a full understanding of how enzymes work, computational methods are becoming increasingly important. Computer calculations can be used to derive essential properties of the enzyme and substrate molecules from their 3-dimensional structures. The simulation of the actual reaction of the enzyme and substrate molecules has become feasible using recently developed methods.

By comparing the results of these calculations with results of experiments, the computational models and the hypothesised mechanisms behind these models can be validated. Furthermore, the use of theoretical models can provide valuable information to supplement experimental results, for example in cases of processes that occur too fast to be investigated experimentally. The development and testing of theoretical models will ultimately lead to methods to predict if, and how fast, specific biochemical processes occur.

1.1 Deriving biochemical reaction rates from molecular models

The general objective of this thesis is to investigate and validate computer models of molecular structures (i.e. of substrates and enzymes) which provide information about mechanisms and rates of enzymatic reactions. The general approach chosen, is to correlate experimental rates of biochemical reactions quantitatively to specific calculated properties for series of substrate (and cofactor) molecules. These so-called quantitative structure-activity relationships (QSARs) demonstrate the ability of various computational models to translate structural data on enzymes and substrates into information about relative reaction rates and reaction mechanisms.

In this thesis, three approaches have been used to obtain relative parameters for the rate of enzymatic reactions. These approaches are based on the physical relationship

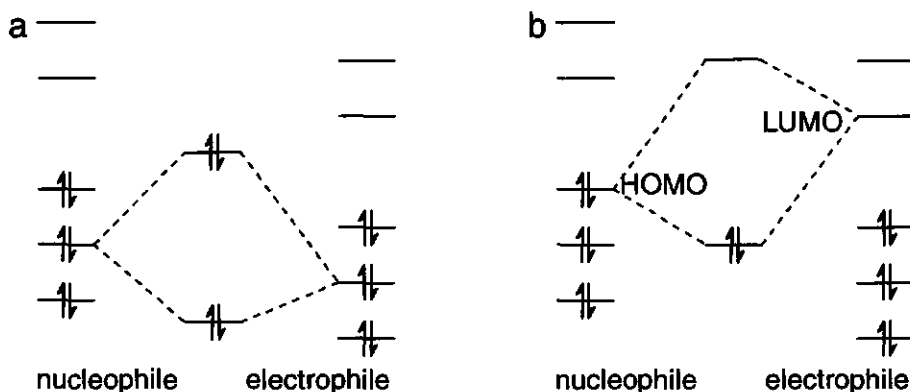


Figure 1.1 Orbital energy diagrams for the transition state interactions between a) two occupied orbitals of the reactants and b) the highest occupied molecular orbital (HOMO) of the nucleophile and the lowest unoccupied molecular orbital (LUMO) of the electrophile. The (non-bonding) interactions between occupied orbitals lead to a raise in energy (due to electron repulsion), whereas the (bonding) interaction between the HOMO and LUMO leads to a decrease in energy. This energy gain strongly depends on the energy difference between the HOMO and LUMO: A smaller energy difference results in a larger orbital splitting upon interaction, which yields a larger gain in energy.

between the rate constant of a reaction and the activation barrier of that reaction, which was first described by the Arrhenius equation: $k_r = Ae^{-E_{act}/RT}$. The activation energy in this equation is an empirical quantity, which can be derived from experiment. (A thermodynamic formulation of this equation is given on the basis of absolute rate theory: $k_r = \frac{kT}{h} e^{-\Delta G^\ddagger/RT}$ (Chang, 1981).) From this equation it becomes clear that useful parameters for (enzymatic) reaction rates should represent the activation energy of the reaction, at least in a relative way.

The first approach used in this thesis is based on the so-called frontier orbital theory. It provides (relative) parameters for electrophilic or nucleophilic reactivity on the basis of calculations on the reactant(s) only. The frontier orbital theory explains that, when comparing the reactions of series of (soft) nucleophiles and (soft) electrophiles, the energy difference between the highest occupied molecular orbital (HOMO) of the nucleophile and the lowest unoccupied molecular orbital (LUMO) of the electrophile is an important factor determining the variation in the activation energies (Figure 1.1) (Fleming, 1976). The HOMO and LUMO energies can therefore be used as relative parameters for the nucleophilic and electrophilic reactivity within a homologous series of reactant(s). Furthermore, the activation barrier for this type of reaction also depends on the densities of the frontier orbitals on the reacting atom centres (Fukui *et al.*, 1952, Klopman, 1968, Salem, 1968). The frontier orbital densities can therefore be used as relative parameters for reactivity of different atoms within a reactant molecule. The

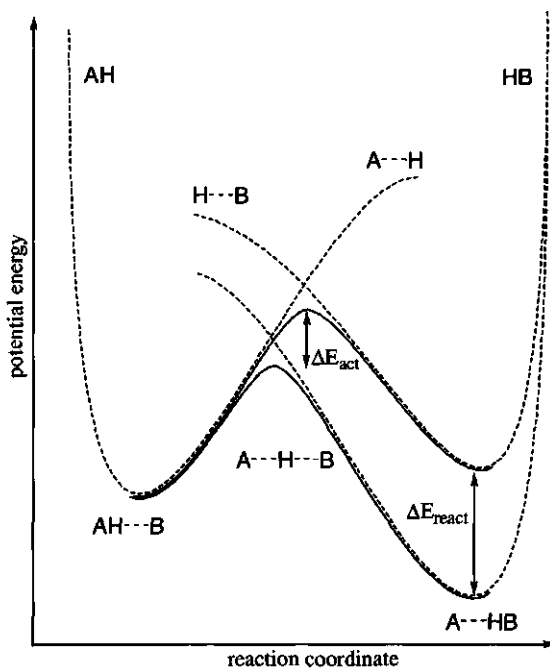


Figure 1.2 Qualitative illustration of the Brønsted relationship between the activation energy and the energy change from reactant to product for proton transfer. The dotted lines represent the potential curves for the proton dissociation of the acid and two different bases respectively. The solid lines represent the energy profiles for the two proton transfer reactions.

frontier orbital parameters for reactivity, based on calculations on isolated reactants in vacuum, have been successfully applied in QSAR studies on enzyme reaction mechanisms (Cnubben *et al.*, 1994, Peelen *et al.*, 1995, Soffers *et al.*, 1996, Van Haandel *et al.*, 1996, Vervoort *et al.*, 1992). In this thesis, frontier orbital parameters have been used in chapters 3, 4 and 7.

The second approach is to calculate the energy differences between the reactant(s) and product(s) of a reaction and to use this reaction energy as a relative parameter for the activation barrier. The relationship between the reaction energy and the activation energy of a reaction was first described by Brønsted for proton transfer reactions and is illustrated in Figure 1.2. The energy difference between reactant and product energies in vacuum have also been used successfully as parameters for enzyme activity (Rietjens *et al.*, 1995, Soffers *et al.*, 1996, Vervoort and Rietjens, 1996). In this thesis, reaction energy parameters have been used in chapters 3 and 7.

The third and most important approach of the present thesis is to actually calculate the activation energy. The activation energy, i.e. the energy difference the reactant and

transition states, can in principle be obtained from a calculation of the reaction pathway, which describes how the reactants transform into the products. As the reaction pathway is approximate and does not include entropic contributions, the resulting energy barriers should still be used as relative parameters (see Chapter 2). For the reaction pathway calculations in this thesis, a so-called combined quantum mechanical/molecular mechanical (QM/MM) method is used. The advantage of this type of method is that it enables the inclusion of the protein atoms surrounding the reacting substrate and cofactor molecules in an explicit way. This implies that a reaction pathway for the reacting substrate and cofactor molecules can be calculated, explicitly taking into account the relevant steric restrictions and electrostatic influences of the enzyme active site. The QM/MM reaction pathway calculations thus provide a sophisticated parameter for enzymatic activity. More importantly, the QM/MM reaction pathway approach gives detailed insight into the enzyme mechanism and provides an explicit model for unstable reaction intermediates. This QM/MM approach forms the basis of chapters 5, 6, 7 and 8 of this thesis.

1.2 The enzymes studied in this thesis

Three enzymes, involved in the microbial degradation of aromatic compounds, have been chosen as model enzymes in this thesis: catechol-1,2-dioxygenase (1,2-CTD), *para*-hydroxybenzoate hydroxylase (PHBH) and phenol hydroxylase (PH). These enzymes have been chosen for several reasons.

The first reason is that, despite extensive experimental research on these enzymes, some mechanistic questions concerning specific steps in their reaction cycles remain unsolved, due to the lifetime of the intermediates involved. Thus, computer models are useful to simulate and test proposed mechanisms for these reaction steps and to characterise the unstable intermediates.

The second reason is the particular substrate specificity of these enzymes. On the one hand, their substrate-specificity is not too narrow: they are capable of converting various substituted substrate analogues in addition to the native substrate, which makes them suitable model enzymes for the QSAR-type studies of this thesis. On the other hand, they are also not highly substrate-*unspecific*: previous QSAR studies were focused on enzymes with large active sites and broad substrate specificity, such as mammalian biotransformation enzymes (Cnubben *et al.*, 1994, Soffers *et al.*, 1996). The enzymes studied in the present thesis have smaller active sites with more strictly defined substrate specificity, which is in principle a further challenge in the development of QSARs. It may require computer methods that take into account the specific effects of the active site surrounding the reacting molecules. The development

of such models is one of the objectives of this thesis and may ultimately lead to QSARs for broader ranges of substrates and/or enzymes.

In addition, there are some technical reasons for choosing the model enzymes. The possibilities to perform accurate calculations on systems containing transition metals have increased over the last decade. Due to developments of QM methods based on density functional theory, it has (only recently) become feasible to perform reasonably accurate calculations on a large transition metal system, such as the iron–ligand–substrate complex of the catechol-1,2-dioxygenase studied in this thesis. With respect to the pathway simulation studies of this thesis, the flavin enzymes, PHBH and PH, are challenging model systems for the QM/MM approach. The simulation of flavin enzymes is currently feasible within the capabilities of the QM/MM method used. The quantum mechanical part of the applied QM/MM method is based on semiempirical molecular orbital theory. The flavin cofactor and the substrates in the model enzymes chosen for the QM/MM studies only consist of the bio-organic elements H, C, N, O, S, F and Cl, for which this semiempirical treatment is reasonably accurate (Dewar and Yuan, 1990, Dewar and Zuebisch, 1988, Dewar *et al.*, 1985). The flavin ring is a large and highly polarisable aromatic system, which is involved in a large variety of biological reactions catalysed by flavoproteins (Palfey and Massey, 1998). A question of general scientific interest is, therefore, how various catalytic functions of the flavin cofactor are obtained through differences in the protein environment. This is a typical example of the kind of questions that can be studied very usefully by QM/MM simulation.

Another technical motive for choosing the model enzymes is the availability of crystal structures. Especially, the simulation of reaction pathways within the environment of the active site requires accurate information about the 3-dimensional structure of the proteins. The crystal structure of the enzymes studied in this thesis, available from the protein databank, provide this information at a reasonably high resolution.

Finally, an additional reason for choosing the model enzymes of this thesis is their involvement in the degradation of xenobiotic compounds. Insight into factors that determine the rates of conversion of substrate analogues by these enzymes is of interest for the assessment of the biodegradability of aromatic xenobiotics.

1.3 Outline of this thesis

In the present thesis, the reaction mechanisms of specific steps in the catalytic cycles of the model enzymes described above are studied by means of three different QSAR approaches mentioned in section 1.1. Chapter 2 will introduce the theoretical principles of the different computational techniques applied. Then, some practical aspects of the application of quantum mechanical models to enzymes, with special

reference to flavin enzymes, are discussed in chapter 3. The subsequent chapters present the results obtained with the specific model enzymes of this thesis.

Chapter 4 describes a frontier orbital-based QSAR study which provides insight into the mechanism of the rate-limiting step of the reaction cycle of catechol-1,2-dioxygenase. The QSARs presented in this study are based on the first two approaches mentioned in section 1.1, using gas-phase quantum mechanical calculation and experimentally determined rate constants for the overall conversion of a series of substrate derivatives.

Chapter 5 presents a QSAR for the conversion of fluorinated *p*-hydroxybenzoates by *p*-hydroxybenzoate hydroxylase. This QSAR is based on a combined quantum mechanical and molecular mechanical reaction pathway model, being an example of the third approach in section 1.1. Further analysis of the QM/MM model is presented in Chapter 6 and provides detailed insight into the reaction mechanism of the simulated reaction pathway for hydroxylation and into the role of specific active site residues. Chapter 7 presents an extended QSAR study on PHBH, in which the various computational approaches/parameters are compared. Furthermore, the QSARs presented in this chapter include substitutions on the flavin cofactor as well as on the substrate.

Finally, chapter 8 presents a QM/MM study on phenol hydroxylase. This chapter further establishes the QM/MM energy barriers as useful QSAR-parameters. Furthermore, this study provides valuable insight in the reaction mechanism of PH, which has not been studied as extensively as PHBH. This chapter also makes a detailed comparison of these two enzymes, which yields insight into their structural and functional relationship.

The remainder of the present chapter will now give more specific information on the model enzymes, relevant for the investigations in this thesis. First, their function in the degradation of aromatic compounds is discussed. Then, a brief overview on the cofactors and reaction mechanisms of these enzymes is given.

1.4 The β -ketoadipate pathway

The enzymes studied in this thesis are all part of the β -ketoadipate pathway, one of the first studied pathways of aerobic microbial degradation of aromatic compounds (Ornston and Stanier, 1964). This microbial degradation of aromatic compounds is an important part of the carbon cycle on earth. Among the most abundant substances produced in photosynthetic processes is lignin, the main constituent of wood. Lignin makes wood strong and rigid. About 25% of the biomass on land consists of this relatively stable (and difficult to degrade) compound (Harwood and Parales, 1996). It has a very irregular, polymeric structure composed of phenylic building blocks. After aerobic splitting of lignin, aromatic acids are the remaining products, which need to be

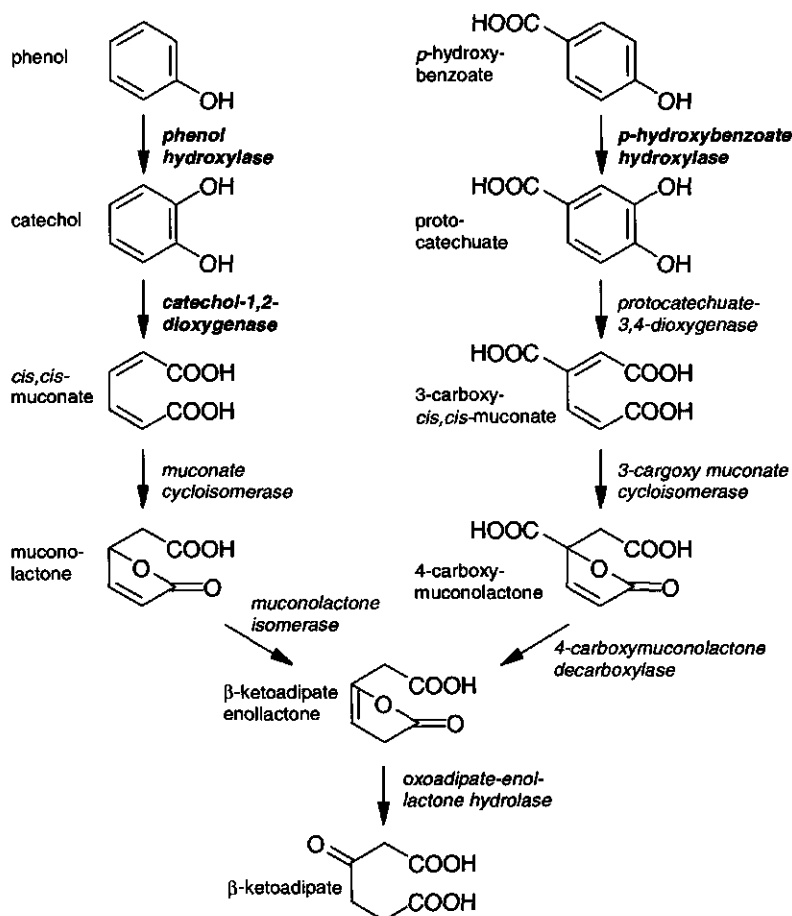


Figure 1.3 The prokaryotic β -ketoadipate pathway. Two branches involving ortho cleavage of the aromatic ring are presented, which proceed via the catechol or protocatechuate key intermediates. In bold the enzymes investigated in this thesis.

further degraded. The degradation of these aromatic compounds occurs most efficiently under aerobic conditions, although it is known to proceed under several anaerobic conditions as well (Reineke and Knackmuss, 1988).

Although lignin is the main source of aromatic compounds to be degraded, there are other sources of aromatic compounds as well. In chemical industries, man produces significant amounts of xenobiotic compounds, which are not necessarily degradable via the natural mineralisation pathways. An important group of these xenobiotics are the halogenated aromatics, such as the well-known PCBs (polychlorinated biphenyls). The various reaction steps of the β -ketoadipate pathway and the enzymes catalysing these reactions are presented in Figure 1.3. Two general phases in the aerobic

degradation of aromatics can be distinguished. First, the aromatic ring is prepared for cleavage by modifying it to specific dihydroxylated benzene intermediates. In the case of the β -ketoadipate pathway the intermediates are catechol and protocatechuate. These key intermediates are cleaved and further degraded to common metabolic intermediates. In the β -ketoadipate pathway, the aromatic ring in catechol or protocatechuate is cleaved between the two adjacent carbons bearing the hydroxyl moieties. This reaction is referred to as ortho-cleavage, as distinct from meta-cleavage (occurring adjacent to the two hydroxyl substituted carbons), which is the entrance to another pathway. The product of ortho-cleavage, (carboxy)muconate, is further degraded via (carboxy)muconolactone and β -ketoadipate-enollactone to β -ketoadipate (3-oxoadipate). The latter is converted into succinyl-CoA and acetyl-CoA, two central intermediates in the cell metabolism.

The eukaryotic pathway differs from the bacterial pathway with respect to the protocatechuate branch in that β -carboxymuconolactone is formed, rather than γ -carboxymuconolactone, which is converted directly to β -ketoadipate. Although a great diversity exists with respect to induction, regulation and gene organisation of the β -ketoadipate pathway, the functional activities involved are highly conserved throughout various bacterial strains (Harwood and Parales, 1996).

Degradation of substituted aromatics via the β -ketoadipate pathway

An important step in the aerobic detoxification and degradation of the halogenated aromatics is the removal of the halogen substituent(s). Dehalogenation can occur in an initial phase, such as hydrolytic dehalogenation of 4-chlorobenzoate (Kobayashi *et al.*, 1997, Marks *et al.*, 1984) or oxidative dehalogenation (Husain *et al.*, 1980, Peelen *et al.*, 1995), after which the product may be degraded via the native pathways. Alternatively, the halogen substituents may be (spontaneously) removed from the non-aromatic pathway intermediates after ring-cleavage has taken place (Reineke and Knackmuss, 1988). Recent ^{19}F -NMR experiments demonstrated that various fluorinated phenols can enter the β -ketoadipate pathway and may lose their fluoro substituents later (Boersma *et al.*, 1998, Cass *et al.*, 1987). This spontaneous dehalogenation may yield products which are not further degraded by the normal β -keto-adipate pathway (Boersma *et al.*, 1998, Schlömann *et al.*, 1990). For chlorinated (Reineke and Knackmuss, 1988) and methyl substituted phenols (Powlowski and Dagley, 1985) so-called modified ortho-pathways exists, which have specialised enzymes (Dorn and Knackmuss, 1978). The chlorocatechol dioxygenase (type II) (Dorn and Knackmuss, 1978) and chloromuconate cycloisomerase of this modified pathway are closely related to catechol dioxygenase (type I) and muconate cycloisomerase enzymes of the normal β -keto-adipate pathway. The subsequent part

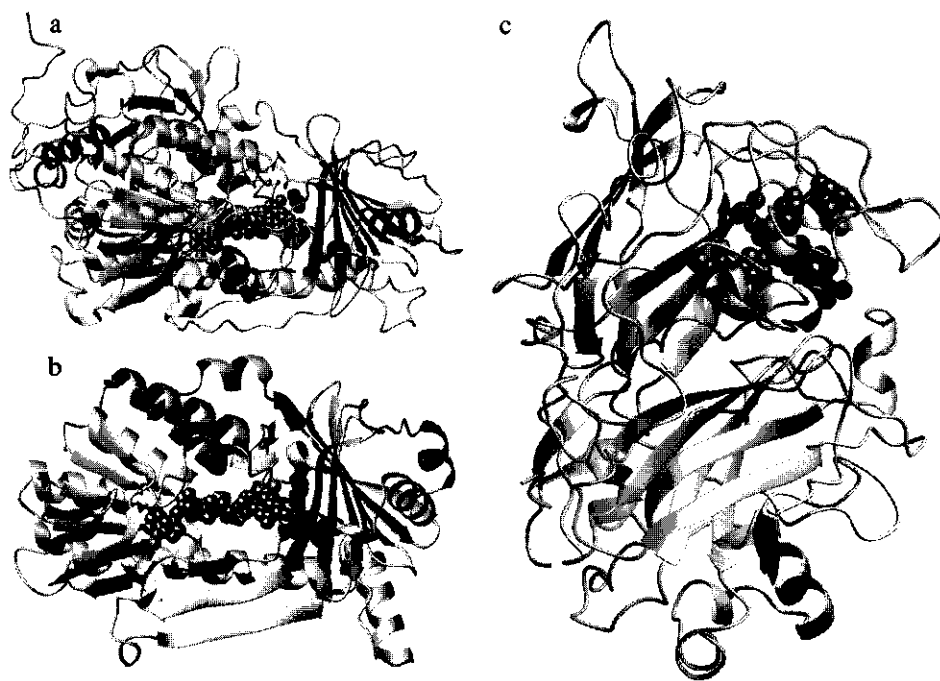


Figure 1.4 Crystal structures of the catalytic units of a) PH (1FOH.PDB), b) PHBH (1PBE.PDB) and c) 3,4-PCD (3PCA.PDB), displayed using MOLMOL. The cofactors and substrates are represented in spheres. The substrates are colored darker than the cofactors. The overall folds of PH and PHBH are similar, except for an extra domain in PH (left top) with an unknown function (Enroth *et al.*, 1998). The iron in 3,4-PCD is represented as a larger sphere. The two subunits of the catalytic $\alpha\beta\text{Fe}$ unit of 3,4-PCD show considerable sequence homology with the subunits in 1,2-CTD and especially the residues involved in iron binding are conserved (Nakai *et al.*, 1990). This indicates that, although no crystal structure is available for 1,2-CTD yet, the structure of the dimeric 1,2-CTD may be very similar to the catalytic unit of 3,4-PCD.

of the modified ortho-pathways, during which dehalogenation occurs, are somewhat more different (Schlömman, 1994), in order to degrade the intermediates formed upon dehalogenation.

Oxygenases involved in the β -ketoadipate pathway

The investigations presented in this thesis concern three oxygenase enzymes of the β -ketoadipate pathway: phenol hydroxylase (PH), *p*-hydroxybenzoate hydroxylase (PHBH), and catechol-1,2-dioxygenase (1,2-CTD) (Figure 1.3). Figure 1.4 schematically presents the structures of these enzymes. Oxygenases catalyse the incorporation of either one (monooxygenases) or both (dioxygenases) atoms of molecular oxygen into their substrate. Molecular oxygen is a highly oxidative species,

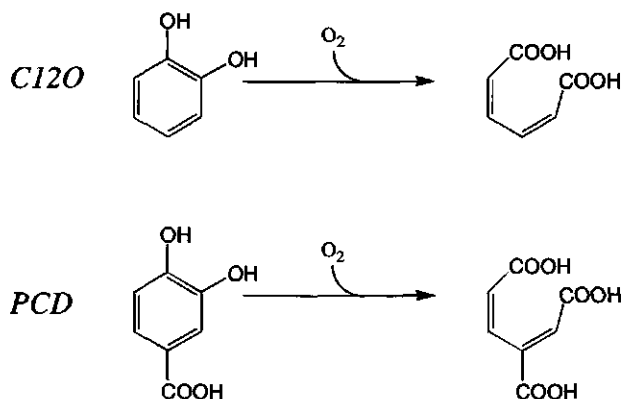


Figure 1.5 The reaction catalysed by catechol-1,2-dioxygenase (1,2-CTD) and protocatechuate-3,4-dioxygenase (3,4-PCD)

which generally reacts very exothermic with organic material. It is, in that respect, a suitable reactant to attack the relatively stable aromatic nucleus. However, due to the fact that oxygen has a triplet groundstate, its reaction with organic compounds, which generally have a singlet groundstate, is spin-forbidden. Thus, the reactions of oxygen with organic compounds have a high activation barrier and, consequently, are very slow. The major task of the oxygenases is, therefore, to somehow circumvent this (quantum mechanical) restriction. This is done by means of different cofactors: the monooxygenases studied in this thesis use a flavin cofactor, whereas the catechol-dioxygenase uses a non-heme iron, to facilitate the reaction of molecular oxygen with the aromatic substrate.

1.5 The intradiol catechol dioxygenases

The intradiol catechol dioxygenases of the β -ketoadipate pathway, catechol-1,2-dioxygenase (1,2-CTD) and protocatechuate-3,4-dioxygenase (3,4-PCD), are related enzymes, which catalyse a similar reaction: the cleavage of the aromatic ring between two adjacent hydroxyl-substituted carbon atoms (Figure 1.5). These two enzymes share the essential mechanistic features described in the following paragraphs.

The catechol dioxygenases contain non-heme iron as a cofactor. Instead of being part of a heme cofactor, or an iron-sulphur cluster, non-heme iron is coordinated by amino-acid side chains. In the resting state of the intradiol dioxygenases, the iron is bound to two anionic (deprotonated) tyrosine side-chains (Que, 1989, Que and Epstein, 1981) and two neutral histidine side-chains (Felton *et al.*, 1982) (Figure 1.6). A fifth ligand in the trigonal bipyramid iron site is formed by a hydroxide coming from the solvent

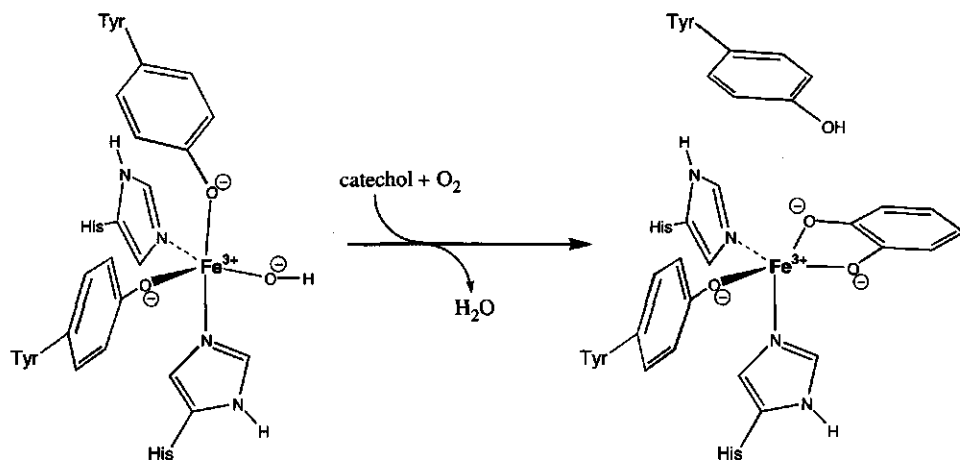


Figure 1.6 The catalytic site of the intradiol dioxygenases in the free enzyme and in the enzyme substrate complex (Ohlendorf *et al.*, 1994, Orville *et al.*, 1997). After substrate is bound, oxygen enters the active site to form a ternary complex. Comparison of the crystal structures of free enzyme and the enzyme-substrate complex provides insight into this sequential binding of substrate and oxygen. Upon substrate binding, structural changes occur at the α/β subunit interface. This may affect the accessibility of the active site for oxygen (Orville *et al.*, 1997). Furthermore, the bidentate binding of catechol to the active site iron, thereby replacing a tyrosine and hydroxide ligand, appears to change the iron coordination site such that an oxygen binding site is created adjacent to the substrate (indicated by a dashed circle) (Orville *et al.*, 1997).

(Whittaker and Lipscomb, 1984a). The iron is in a ferric (i.e. 3+) high-spin state throughout the catalytic cycle (Lipscomb and Orville, 1992).

Overall reaction mechanism

Upon binding to the enzyme, the substrate catechol (or protocatechuate) coordinates bidentately to the iron(III) (Orville and Lipscomb, 1989) (Figure 1.6). The two phenolic moieties of the substrate become deprotonated and replace the hydroxide and one of the tyrosine ligands (True *et al.*, 1990). These structural characteristics of the iron containing catalytic site and the changes upon binding of the substrate, derived from spectroscopic studies, were recently confirmed by crystal structures of free 3,4-PCD (Ohlendorf *et al.*, 1994) and of 3,4-PCD in complex with substrate and various substrate analogues (Orville *et al.*, 1997). It is proposed that this binding of catechol to iron induces radical density on the aromatic ring, which would enhance the direct reaction with triplet oxygen. However, the precise nature of the iron-substrate complex is still subject to discussion. On the one hand spectroscopic studies indicate that the iron remains high-spin ferric, i.e. Fe^{3+} (Lipscomb and Orville, 1992, Que, 1989, Whittaker and Lipscomb, 1984b). On the other hand an Fe^{2+} -semiquinone

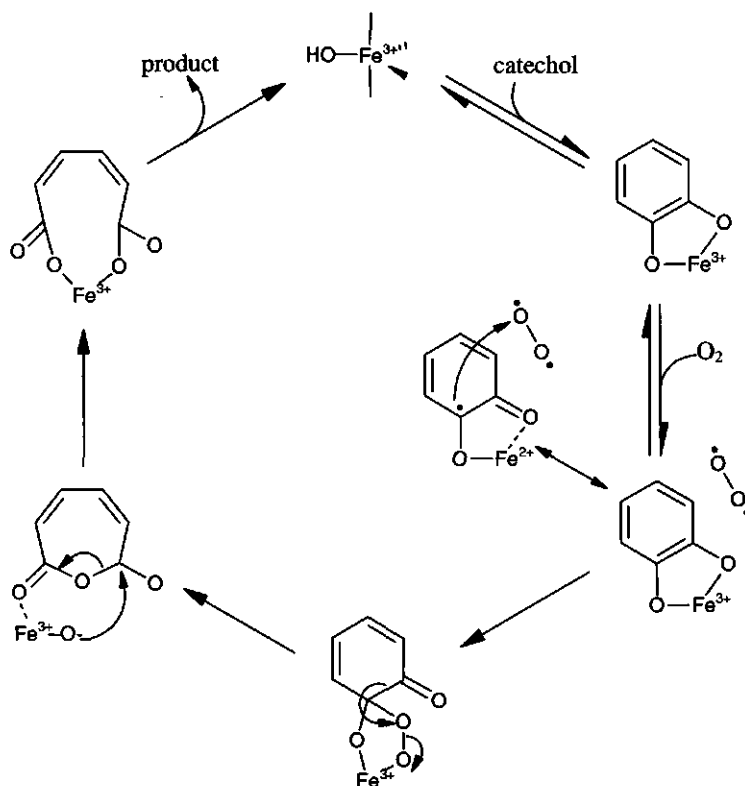


Figure 1.7 The reaction mechanism proposed for the intradiol dioxygenases.

radical resonance structure has been proposed to contribute to the activation of the substrate (Figure 1.7).

Substrate binding is proposed to cause structural changes, which result in an oxygen binding site (Figure 1.6). Oxygen enters the active site, without being coordinated to the iron. Due to its partial radical character, the substrate can react with this oxygen molecule via a nucleophilic attack. The subsequent reaction steps lead to cleavage of the aromatic ring between the two hydroxyl-substituted carbons (Figure 1.7).

The extradiol dioxygenases, which cleave the aromatic ring adjacent to the two hydroxyl-substituted carbons, have found a different solution to circumvent the spin-forbidden reaction with triplet oxygen. The substrate is, also bidentately, bound to ferrous iron (Fe^{2+}). In this case, however, the oxygen molecule does not directly react with the aromatic ring. It binds to iron first, to form a substrate-enzyme- Fe^{2+} - $\text{O}-\text{O}$ complex, thereby changing the electronic configuration of the dioxygen (Shu *et al.*, 1995). Subsequently, the distal oxygen reacts with the extradiol carbon centre.

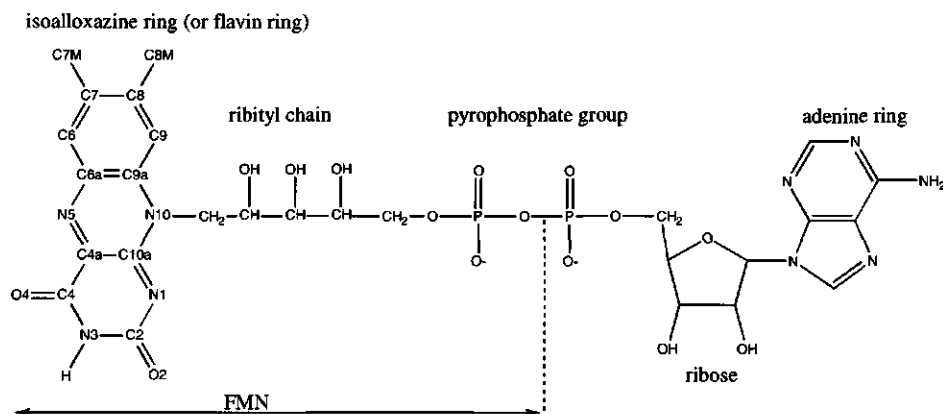


Figure 1.8 The chemical structures of FAD (flavin adenine dinucleotide) and FMN (riboflavin-5'-monophosphate)

The QSAR-study on catechol-1,2-dioxygenase, presented in chapter 4 of this thesis, focuses on the crucial reaction between the iron-bound substrate and molecular oxygen. In order to perform calculations relevant for the reaction within the enzyme complex, the extent of charge transfer from the dianic catechol to the iron was investigated by means of a calculation on the complete iron-ligand-substrate model. This model system was built on the basis of the crystal structure of the related protocatechuate dioxygenase enzyme, since the active sites of catechol-1,2-dioxygenase and protocatechuate dioxygenase are expected to be similar. Sequence alignment shows that the active site residues are conserved in both enzymes (Orville *et al.*, 1997).

1.6 The flavin dependent monooxygenases

Many hydroxylation reactions involved in the conversion of the various aromatic compounds to catechol derivatives (as a preparation for ring cleavage) are catalysed by flavin dependent monooxygenases (Dagley, 1987). Flavin is the general name for a group of compounds that contain an isoalloxazine ring (Figure 1.8). This isoalloxazine ring can adopt a whole series of electronic states (Zheng and Ornstein, 1996) and is capable of supporting a variety of reactions. Flavin is found as a cofactor in various groups of enzymes, for example as FMN (riboflavin-5'-monophosphate) or as FAD (flavin-adenine-dinucleotide) (Palfey and Massey, 1998).

The most widely studied flavin dependent monooxygenases are phenol hydroxylase (PH) and *p*-hydroxybenzoate hydroxylase (PHBH), which will be discussed in the remainder of this paragraph. PH and PHBH catalyse the hydroxylation of phenol and *p*-hydroxybenzoate, respectively, at the ortho position of the hydroxyl moiety. These

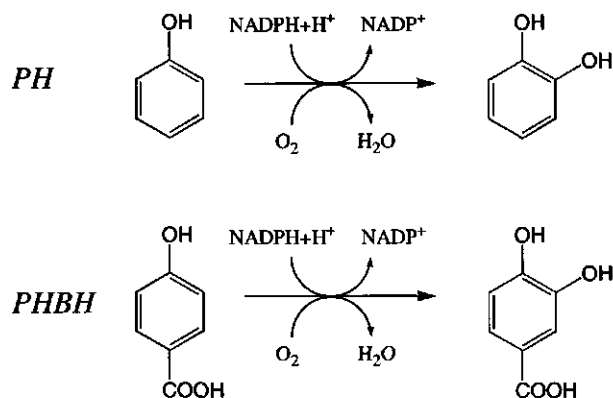


Figure 1.9 The reactions catalysed by phenol hydroxylase (PH) and *p*-hydroxybenzoate hydroxylase (PHBH)

related enzymes (catalysing homologous reactions) use NADPH and molecular oxygen, and produce catechol and protocatechuate respectively (Figure 1.9). The PHBH enzyme studied in this thesis is from the soil bacterium *Pseudomonas fluorescence* (Howell *et al.*, 1972). The enzyme is very homologous to PHBH from other *Pseudomonas* strains (Hosokawa and Stanier, 1966, Nakamura *et al.*, 1970), but, due to its high stability, PHBH from *Ps. fluorescence* has been most widely studied (Van Berkel and Müller, 1991). The most widely studied phenol hydroxylase, which has also been subject to investigation in this thesis, comes from the soil-yeast *Trichosporon cutaneum*.

Overall reaction mechanism

PH and PHBH are believed to operate via a similar overall reaction cycle (Figure 1.10). The first step in this common reaction mechanism is binding of the substrate. The binding of substrate induces the reduction of the FAD cofactor by NADPH, i.e. it stimulates the rate of reduction up to 10⁵ fold (Howell *et al.*, 1972). The reduced flavin then reacts with molecular oxygen to yield the C4a-peroxyflavin. For this reaction with triplet oxygen the ability of the flavin to form radicals is probably essential. The reaction is believed to proceed via a one-electron transfer from the flavin to oxygen (Harayama *et al.*, 1992, Kemal *et al.*, 1977). The resulting superoxide and flavin-semiquinone radicals quickly "collapse" to form the C4a-peroxyflavin (Palfey and Massey, 1998). The peroxide moiety is then protonated to generate the C4a-hydroperoxyflavin (Maeda-Yorita and Massey, 1993, Schreuder *et al.*, 1990).

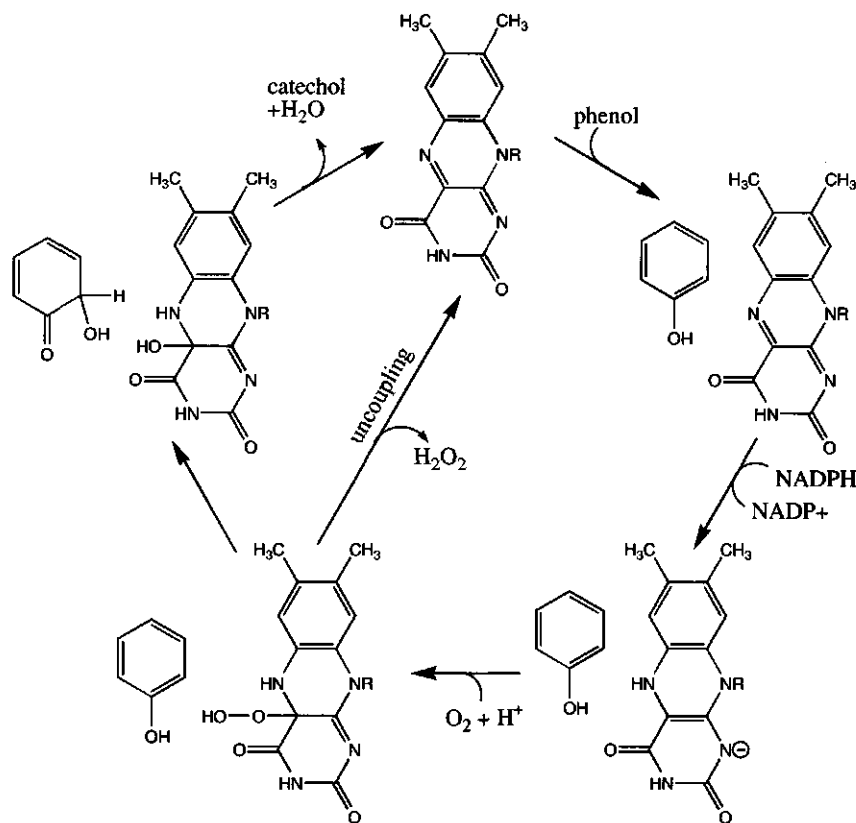


Figure 1.10 The proposed reaction cycle for the flavin dependent monooxygenases phenol hydroxylase and p-hydroxybenzoate hydroxylase

How the resulting C4a-hydroperoxyflavin subsequently reacts with the substrate has been subject to many experimental studies and to a lot of speculation. Among the proposed reaction mechanisms are dioxygen transfer to the substrate (Kemal and Bruce, 1979, Keum *et al.*, 1990), homolytic cleavage of the peroxide bond (Anderson *et al.*, 1987, Anderson *et al.*, 1991, Peräkylä and Pakkanen, 1993) and a mechanism which proceeds via an open ring form of the flavin (Entsch *et al.*, 1976, Husain *et al.*, 1980). Recent studies (Maeda-Yorita and Massey, 1993, Ridder *et al.*, 1999a), however, have lead to consensus in support of an electrophilic aromatic substitution proceeding by heterolytic cleavage of the peroxide bond and resulting in the formation of a deprotonated C4a-hydroxyflavin intermediate of the cofactor and a hydroxycyclohexadienone as an initial form of the product (Figure 1.10). This hydroxycyclohexadienone non-enzymatically converts to catechol via keto-enol tautomerisation

(Maeda-Yorita and Massey, 1993). The C4a-hydroxyflavin splits off water to yield oxidised flavin, which makes the enzyme ready to enter a next reaction cycle.

The simulations of PHBH and PH presented in this thesis concern the hydroxylation step. This reaction step has been proposed to be rate-limiting in the reaction cycles of PHBH and PH (Peelen *et al.*, 1995, Vervoort *et al.*, 1992), and is therefore most relevant for calculating the energy barriers to be compared with experimental rate constants. For the construction of the computational models, two aspects of the reaction mechanism by PHBH and PH are of importance: the position of the flavin to be modelled (in relation to flavin mobility) and the protonation state of the substrate.

Flavin mobility

Mobility of the flavin plays an important role in the reaction mechanism. Different conformations of the flavin cofactor have been observed in crystals of PHBH. In the crystal structures of the PHBH-substrate complex (Schreuder *et al.*, 1989, Wierenga *et al.*, 1979) and the PHBH-product complex (Schreuder *et al.*, 1988b) the flavin ring is buried in the interior of the protein and closes the active site. This conformation makes the active site inaccessible to solvent (Schreuder *et al.*, 1988a). It is referred to as the "in" conformation, to distinguish it from the "out" conformation, in which the flavin is moved away from the active site and exposed to solvent. This "out" conformation has been observed in crystal structures of PHBH in complex with 2,4-dihydroxybenzoate and 2-hydroxy-4-aminobenzoate (Schreuder *et al.*, 1994), of the Tyr222Ala (Schreuder *et al.*, 1994) and Tyr222Phe (Gatti *et al.*, 1994) mutants of PHBH, and of PHBH in complex with a stereochemical analog of the cofactor, arabinoflavin adenine dinucleotide (a-FAD) (Van Berkel *et al.*, 1994). The crystal structure of the free enzyme has recently been described and contains a "flexible" FAD, located at an intermediate position between the "in" and "out" conformations (Eppink *et al.*, 1999). Mobility of the flavin to the "out" conformation is believed to be important for accessibility of the active site for substrate (Gatti *et al.*, 1994, Schreuder *et al.*, 1994). Furthermore, it was found that the isomeric form of the enzyme-substrate complex with arabino-FAD, in which this isomeric cofactor preferentially adopts the "out" conformation, is reduced faster by NADPH than the native enzyme-substrate complex (Van Berkel *et al.*, 1994). This supports the proposal that the reduction of the flavin by NADPH occurs in the "out" conformation (Schreuder *et al.*, 1994). Recently, A model for the ternary enzyme-substrate-NADPH complex was proposed with the flavin in the "out" position (on the basis of kinetic effects of mutations on the interdomain surface) (Eppink *et al.*, 1998).

Once the FAD cofactor is reduced, it is believed to move back into the active site, as supported by a crystal structure of the reduced enzyme-substrate complex with the

flavin in the "in" conformation (Schreuder *et al.*, 1992). In this position the reduced flavin may react with molecular oxygen, and the resulting peroxide can be protonated (Schreuder *et al.*, 1990), while the solvent accessibility is minimised. The latter is important to prevent decomposition of the C4a-hydroperoxyflavin into oxidised flavin and hydrogen peroxide (referred to as "uncoupling", see Figure 1.10) (Kemal *et al.*, 1977, Van der Bolt *et al.*, 1996) and to allow efficient hydroxylation of the substrate (Entsch and Van Berkel, 1995, Schreuder *et al.*, 1992). The recently refined x-ray structure of PH, in which both the "in" and the "out" conformations are present in a single crystal, indicates that the flavin mobility is relevant as well in the reaction cycle of PH (Enroth *et al.*, 1998).

In line with the above, the models of the C4a-hydroperoxyflavin intermediates for PHBH and PH presented in chapters 5 to 8 have been based on the crystal structure with the flavin in the "in" conformation.

Substrate deprotonation

A second important factor in the catalytic cycle of PHBH is the deprotonation of the substrate hydroxyl moiety. The substrate *p*-hydroxybenzoate becomes deprotonated, directly upon its binding to the active site of PHBH. Deprotonation of the hydroxyl moiety at C4 of the aromatic substrate is achieved via a hydrogen bonding network formed by Tyr201, Tyr385 (Entsch *et al.*, 1991, Eschrich *et al.*, 1993), two solvent molecules and His72 (Gatti *et al.*, 1996, Schreuder *et al.*, 1994). This deprotonation appears to be essential for substrate conversion, based on the observations that compounds lacking this hydroxyl moiety are not converted (Entsch *et al.*, 1976) and that mutants in which the hydrogen bonding network had been disabled showed a considerable decrease in efficiency of hydroxylation (Entsch *et al.*, 1991, Eschrich *et al.*, 1993). Molecular orbital studies on the *p*-hydroxybenzoate substrate have demonstrated that deprotonation increases the nucleophilic reactivity of the C3 centre in the substrate, which could be essential for an electrophilic attack on this C3 centre by the C4a-hydroperoxyflavin cofactor (Vervoort *et al.*, 1992).

In addition to its requirement for the hydroxylation step, deprotonation has recently been proposed to facilitate movement of the flavin to the "out" conformation, possibly by inducing a conformational change in the active site loop around residue 294, thereby stimulating flavin reduction (Palfey *et al.*, 1999). This coupling of flavin reduction to substrate deprotonation may help to prevent hydrogen peroxide formation and, more importantly, would provide PHBH with a selection mechanism to distinguish between the substrate *p*-hydroxybenzoate and a related compound *p*-aminobenzoate. The latter can not be deprotonated and is therefore not converted by the enzyme (Palfey *et al.*, 1999). Such a conversion would yield a toxic aminophenol.

In PH, binding of the substrate also accelerates the reduction of FAD (Neujahr and Kjellén, 1978). However, the proton channel present in PHBH does not exist in PH (Enroth *et al.*, 1998). Furthermore, several studies indicate that the substrate is not deprotonated in the oxidised enzyme–substrate complex (Detmer and Massey, 1985, Neujahr and Kjellén, 1978, Peelen *et al.*, 1993). Nevertheless, ^{19}F -NMR experiments and molecular orbital calculations suggested that deprotonation of the substrate is relevant later in the reaction cycle, in order to activate the substrate for an electrophilic attack by the C4a-hydroperoxyflavin intermediate (Peelen *et al.*, 1993). The importance of substrate deprotonation for the hydroxylation reaction has been further investigated in the simulations described in chapters 6 and 8 of the present thesis.

In this thesis, the reaction mechanisms of the specific rate limiting steps in the catalytic cycles of the enzymes described above have been studied in more detail, by comparing experimental results with outcomes from computer calculations. The next chapter will provide some background information about the computational techniques applied in this research.

2 Introduction to the theoretical principles of molecular orbital theory and its combination with molecular mechanics

Due to the increasing power and availability of computers, the use of theoretical techniques to simulate and study enzyme reactions becomes more and more accessible to biochemists. Although the size of protein molecules prevents computer calculations of absolute accuracy, a theoretical simulation of protein reactions can already offer very useful information to complement experimental data.

In general, computer calculations of reactions, or of molecular properties determining reactivity, require a method based on quantum mechanics, such as molecular orbital calculations. The first part of this chapter will briefly introduce the theoretical principles that form the basis of molecular orbital theory. More practical aspects of the application of quantum mechanical calculations in biochemical research will be discussed in chapter 3.

The second part of the present chapter will focus on the theoretical aspects of the combination of quantum mechanics and molecular mechanics, which enables the simulation of reactions within the actual environment of the enzyme.

2.1 Quantum mechanical methods

The basis of all quantum mechanical calculations is the time-independent Schrödinger equation[#]: $H\Psi = E\Psi$

This equation describes how the energy of a molecular system depends on its structure in terms of quantum mechanical theory. The system is described as a wavefunction Ψ , which is a function of the coordinates of all electrons and nuclei in the system. The Hamiltonian H is a so-called energy operator, which by definition means that operating the Hamiltonian on the wavefunction Ψ results in the energy of the system multiplied by the wavefunction. This equation has multiple solutions: the eigenfunctions Ψ describe different states of the molecular system and the eigenvalues E present the corresponding energies of these states.

For molecules, containing at least three particles (for example H_2^+ has two nuclei and one electron) the Schrödinger equation is too complex to be solved analytically. The following paragraphs describe a number of approximations, which need to be made in order to solve this problem.

[#]The Schrödinger equation (Szabo and Ostlund, 1982)

The Schrödinger equation defines the Hamiltonian operator as the energy operator of a quantum mechanical particle:

$$H\Psi = E\Psi$$

For a molecule the Hamiltonian is given by:

$$H = -\sum_i \frac{1}{2} \nabla_i^2 - \sum_\alpha \frac{1}{2m_\alpha} \nabla_\alpha^2 + \sum_{ij} \frac{1}{r_{ij}} + \sum_{\alpha\beta} \frac{Z_\alpha Z_\beta}{r_{\alpha\beta}} - \sum_{i\alpha} \frac{Z_\alpha}{r_{i\alpha}}$$

In which the five summations represent the kinetic energy of all electrons i , the kinetic energy of all nuclei α , the repulsion between electrons i,j , the repulsion between nuclei α,β and the attraction between electrons i and nuclei α .

The energy is given by:

$$E_{QM} = \frac{\int \Psi^* H \Psi}{\int \Psi^* \Psi}$$

When the Born-Oppenheimer approximation is applied, the Schrödinger equation is solved for the electrons only, with the nuclei regarded as point-charges fixed in space. Consequently, two terms in the Hamiltonian, the kinetic energy of all nuclei α , and the repulsion between nuclei α,β vanish:

$$H_{elec} = -\sum_i \frac{1}{2} \nabla_i^2 + \sum_{ij} \frac{1}{r_{ij}} - \sum_{i\alpha} \frac{Z_\alpha}{r_{i\alpha}}$$

For each configuration of the nuclei the energy can now be evaluated as the sum of the electronic energy, obtained by solving the Schrödinger equation, and the nuclear repulsion:

$$E_{QM} = \frac{\int \Psi_{elec}^* H_{elec} \Psi_{elec}}{\int \Psi_{elec}^* \Psi_{elec}} + \sum_\alpha \sum_{\beta > \alpha} \frac{Z_\alpha Z_\beta}{R_{\alpha\beta}}$$

The Born-Oppenheimer approximation

A fundamental approximation made in quantum mechanical calculations on molecules is the Born-Oppenheimer approximation. Due to a difference in mass, electrons move much faster than nuclei and respond almost instantaneously to changes in the nuclear positions. Therefore, the electrons can be described as moving in a static field of fixed nuclei. This assumption leads to the Born-Oppenheimer approximation, stating that the Schrödinger equation can be solved for the electrons only, with the nuclei regarded as point-charges fixed in space. For each configuration of the nuclei (i.e. geometry of the molecule) the potential energy can be evaluated as the sum of the electronic energy, obtained by solving the Schrödinger equation, and the nuclear repulsion. This energy as a function of the nuclear coordinates is referred to as the Born-Oppenheimer potential energy surface.

How the electronic Schrödinger equation is tackled

The Schrödinger equation for the electrons only is still too complex to be solved for all but the simplest molecules (i.e. H_2^+). A further approximation, the so-called molecular orbital (MO) approximation is made by assuming that the electronic wavefunction of a specific electronic state can be described as a set of individual molecular orbitals, which are occupied by electrons according to that specific electronic state. In the ground state, for example, the electrons occupy the molecular orbitals with the lowest energies. A next step of approximation is achieved by the assumption that the molecular orbitals can be described as linear combinations of atomic orbitals (LCAO). These atomic orbital functions are derived from solving the Schrödinger equation for the hydrogen atom, and therefore are often referred to as hydrogen-like (Slater) atomic orbitals (Atkins, 1983). The set of atomic orbital wave-functions used to build the molecular orbitals is called a *basis-set*.[#]

By introducing these two approximations (the MO-LCAO approximation), the problem of solving the Schrödinger equation for the electronic system is transformed into the problem of finding the correct set of coefficients, which determine the contributions of the atomic orbitals to the various molecular orbitals in the system. This problem is mathematically feasible via an iterative process which results in a so-called self consistent field[#]. This level of approximation is referred to as the Hartree-Fock method. Hartree-Fock computations (using sufficient basis sets) often give reasonable descriptions of both the optimal geometry and the electronic structure of many closed-shell molecules. However, various adaptations/extensions have been made to the HF method to either speed up the calculations or to further improve the results.

Basis sets

Different basis sets exist, which vary in the number of atomic orbital wavefunctions included in the LCAOs, describing the molecular orbitals. A so-called minimal basis set contains only the atomic basis functions occupied by electrons. An example of a minimal basis set is the STO-3G basis set. (3G denotes that the Slater type orbitals are approximated by a combination of 3 gaussian wavefunctions, which speeds up the calculations). However, larger basis sets have been developed, imposing fewer restrictions on the spatial distribution of the electrons and therefore resulting in a better approximation to the exact molecular orbitals (the so-called Hartree-Fock limit). The first extensions to the minimal basis set are the so-called split valence basis sets, such as 3-21G, 6-31G and 6-311G, which include two or more basis functions, with different sizes, for each valence orbital. In the 3-21G basis set, for example, each inner shell orbital is described by one basis function (obtained by combining 3 gaussian functions), while the valence shell consists of two sets of basis functions (built from 2 and 1 gaussian functions respectively). These basis sets can be further extended by adding so-called polarisation functions, i.e. basis functions of higher angular momentum than required to describe the ground state of each atom (e.g. p-type functions on hydrogens and d-type functions on C, O and N, etc.) These polarisation functions allow for the possibility of small displacements of electronic charge away from the nuclear positions and are denoted by adding the type of the additional function in brackets, e.g. 6-31G(d) or 6-31G(d,p), or with an asterisk, e.g. 6-31G* or 6-31G** respectively. Finally, for the description of negatively charged species, so-called diffuse functions, generally designated via a '+'-sign, should be included in the basis set. These are 'oversized' orbital functions to allow for an expansion of the orbitals due to electron-electron repulsion in the anionic molecule (Foresman and Frisch, 1996).

The self-consistent field (SCF)

The shape of a molecular orbital in the Hartree-Fock approximation, determined by the contributions of the various atomic orbitals (the so-called atomic orbital coefficients), depends on the electrostatic field due to all nuclei and other electrons, and thus on the shape of the other molecular orbitals. In other words, the MOs can only be calculated directly if the solution is already known. This problem is solved by using an iterative process which starts from a guess of the molecular orbitals, on the basis of which new (and better) molecular orbitals are calculated. These then serve as the input for another calculation of the molecular orbitals, and so on. This process converges towards a point where the input orbitals are identical to the output orbitals. In practice, this so-called self-consistent field (SCF) is recognised as the point where the Hartree-Fock energy doesn't change anymore (to within a certain tolerance.)

Including electron correlation

The HF approximation of introducing molecular orbitals, describing individual electrons within the average field of the other electrons, neglects the fact that the movement of the electrons is correlated (due to interactions between the electrons). Due to this electron correlation, the actual field influencing the electron is not identical to the average field as assumed in the MO approximation. One way to improve on the MO approximation is by applying configuration interaction (CI). This involves mixing (i.e. combining) all (in case of full CI), or a number of, different configurations built up from the molecular orbitals, which leads to a better wavefunction. This

computationally expensive procedure is often applied in an approximate way (based on perturbation theory) via Møller-Plesset theory, denoted as MP2, MP4, etc. Inclusion of configuration interaction, together with an increase in the size of the basis set, leads to a systematic convergence of the *ab initio* calculations towards the exact solution of the Schrödinger equation.

A second approach to account for electron correlation is based on density functional theory (DFT). By using the principle that electron correlation is a function(al) of the electron density, the DFT methods provide very accurate results (especially the hybrid density functional methods like B3LYP) in a relatively inexpensive way (Foresman and Frisch, 1996).

Semiempirical methods

The *ab initio* and DFT calculations involve the evaluation of numerous integrals over the atomic orbital functions. Semiempirical methods introduce further approximations, to reduce the number of integrals to be calculated and to replace them by parameterised, semiempirical expressions. A first approximation, referred to as *neglect of differential overlap*, assumes that integrals involving the overlap terms between atomic orbitals of different atoms, can be neglected (Dewar, 1975). Furthermore, the various remaining integrals are replaced by simpler functions containing parameters which are optimised to fit the results with experimental data (Dewar and Thiel, 1977). In this way the semiempirical calculations are much faster than HF calculations. In many cases, they also yield better results than HF, due to a correction of shortcomings in the HF theory (i.e. lack of electron correlation) by the parameterisation on the basis of experimental data. Various schemes and parameterisations have been applied, among which MNDO (Modified Neglect of Differential Overlap, Dewar and Thiel, 1977), AM1 (Austin Model 1, Dewar *et al.*, 1985) and PM3 (Parametric Method 3, Stewart, 1989) were most successful.

2.2 The molecular mechanical approximation

The computation time required for quantum mechanical calculations increases with at least the square of the number of basis functions, which in practice limits the number of atoms that can be included in the quantum mechanical model. Quantum mechanical treatment of systems as large as proteins is therefore either impossible or impracticable for most purposes. Therefore, computational methods designed to model complete proteins often make a more rigorous approximation. A type of methods, which is widely used for protein modelling, is referred to as molecular mechanics, which assumes that molecules can be described by classical mechanics. Instead of (quantum mechanically) calculating the average electronic distribution, all the electronic

phenomena which govern the behaviour of a molecule, e.g. chemical bonds, spatial configurations of atoms and electrostatic interactions, are described by empirically derived (classical) potential energy functions. A set of such simple analytical expressions is called a forcefield.[#]

Due to the simplicity of the empirical potential energy terms used, molecular mechanical calculations on very large systems like proteins are efficiently performed. Molecular mechanics is therefore a useful basis for various biochemically relevant computational techniques, including homology modelling or structure prediction, molecular dynamics and docking. However, the molecular mechanical potential is not suitable for studying processes which involve changes in electronic structure, e.g. charge redistributions and chemical reactions involving the breaking and/or formation of bonds. To illustrate this, Figure 2.1 compares the potential energy curves for the stretching and dissociation of the HF molecule obtained from high level quantum chemical calculations and with the harmonic bond potential of the CHARMM forcefield. The minima are located at the same H-F distance and the shape around the minima of both energy profiles is similar, which illustrates that the molecular mechanical potential can be used to describe the optimal geometries and dynamics of groundstate molecules. However, molecular structures further away from the equilibrium geometry, for example an intermediate state towards the breaking of a bond, are not well described by the CHARMM forcefield. The harmonic potential for the covalent bond incorrectly predicts an infinite rise in energy upon breaking of this bond.

[#]Forcefield

In this thesis, molecular mechanical calculations are performed with the program CHARMM (Brooks *et al.*, 1983). The CHARMM forcefield consists of harmonic potential energy terms for all covalent bonds, bond angles and dihedral angles in the molecule. In addition, Van der Waals interactions and electrostatic (Coulomb) interactions, due to defined point charges centred at various atoms within the molecule, are included in the total potential energy. Thus:

$$\begin{aligned}
 E_{MM} = & \sum_b K_b (b - b_0)^2 && \text{Bonds} \\
 & + \sum_{\theta} K_{\theta} (\theta - \theta_0)^2 && \text{Bond angles} \\
 & + \sum_{\phi} K_{\phi} (1 + \cos(n\phi - \delta)) && \text{Dihedral angles} \\
 & + \sum_M \sum_{N>M} \frac{q_M q_N}{DR_{MN}} && \text{Coulomb interactions} \\
 & + \sum_M \sum_{N>M} 4\epsilon_{MN} \left[\left(\frac{\sigma_{MN}}{R_{MN}} \right)^{12} - \left(\frac{\sigma_{MN}}{R_{MN}} \right)^6 \right] && \text{Van der Waals interactions}
 \end{aligned}$$

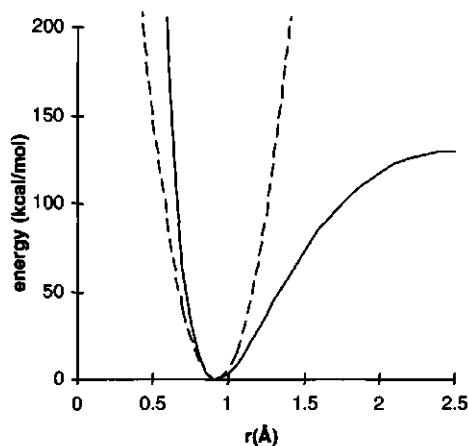


Figure 2.1 Comparison of high level quantum mechanical (QCISD(T)/6-311+G(2d,p) - solid line) and molecular mechanical (CHARMM 22 forcefield - dotted line) potential energy curves for the dissociation of the HF molecule in vacuum. These curves have been calculated using Gaussian98 (Frisch *et al.*, 1998) and CHARMM v24b1 (Brooks *et al.*, 1983) respectively.

2.3 Combined quantum mechanical and molecular mechanical methods

During the last decade, combined quantum mechanical and molecular mechanical (QM/MM) potentials have been developed[#], which enable the simulation of reactions within the environment of complete proteins. The QM/MM approach combines the strengths of the two theoretical methods, i.e. the accuracy of quantum mechanical methods required for modelling reactions and the efficiency of molecular mechanical methods required for large systems like proteins. The QM/MM method implies that a computationally expensive quantum mechanical description is applied only to the atoms involved in the reaction of interest, whereas a fast molecular mechanical method is used to describe the large number of surrounding enzyme and/or solvent atoms. The two parts of this QM/MM model system interact through electrostatic and Van der Waals interactions. The electrostatic QM/MM interactions are accounted for by including the point charges of the surrounding MM atoms in the Schrödinger equation for the QM atoms. Since no electron density is calculated for the MM atoms, a classical energy term for the Van der Waals interactions between the QM and MM atoms needs to be included. The remainder of this section describes some technical aspects of the QM/MM approach which require special attention for a proper application of this method to biochemical models. Some of these technical aspects are still subject to discussion and scientific development. The solutions implemented in CHARMM, and applied in the present thesis will be described.

#The QM/MM potential

The QM/MM potential is the sum of the QM and MM potentials plus a QM/MM interaction potential: $E_{QM/MM} = E_{QM} + E_{MM} + E_{QM/MM}$

The QM/MM interactions involve three terms.

- The interaction between the QM electrons and the MM (point) charges
- The interaction between the QM nuclei and the MM (point) charges
- Since no electron density distribution is calculated for the MM atoms, no Pauli repulsion and no London dispersion interactions are included in the terms above. Therefore, a classical Lennard-Jones (VDW) interaction energy is calculated between the QM and MM atoms.

Thus, the total QM/MM potential is given by:

$$E_{QM/MM} = \frac{\int \Psi^* H \Psi}{\int \Psi^* \Psi} + \sum_{\alpha} \sum_{\beta > \alpha} \frac{Z_{\alpha} Z_{\beta}}{R_{\alpha\beta}} + \sum_{\alpha} \sum_M \frac{Z_{\alpha} q_M}{R_{\alpha M}} \quad \text{b}$$

$$+ \sum_b k_b (b - b_0)^2 + \sum_{\theta} k_{\theta} (\theta - \theta_0)^2 + \sum_{\phi} k_{\phi} (1 + \cos(n\phi - \delta)) + \sum_M \sum_{N > M} \frac{q_M q_N}{DR_{MN}}$$

$$+ \sum_M \sum_{N > M} 4\epsilon_{MN} \left[\left(\frac{\sigma_{MN}}{R_{MN}} \right)^{12} - \left(\frac{\sigma_{MN}}{R_{MN}} \right)^6 \right] + \sum_{\alpha} \sum_{M > \alpha} 4\epsilon_{\alpha M} \left[\left(\frac{\sigma_{\alpha M}}{R_{\alpha M}} \right)^{12} - \left(\frac{\sigma_{\alpha M}}{R_{\alpha M}} \right)^6 \right] \quad \text{c}$$

In which the Hamiltonian is given by:

$$H = -\sum_i \frac{1}{2} \nabla_i^2 + \sum_{ij} \frac{1}{r_{ij}} - \sum_{i\alpha} \frac{Z_{\alpha}}{r_{i\alpha}} - \sum_{iM} \frac{q_M}{r_{iM}} \quad \text{a}$$

The labelled boxes mark the terms that account for the QM/MM interactions. The other terms are identical to those for pure QM or pure MM calculations as presented in the textboxes on page 20 and page 24.

Link atoms

One of the problems that arise when QM/MM methods are applied on proteins is that it is often necessary to use the QM and MM treatment for atoms within the same (protein) molecule. In such cases, covalent bonds need to be defined between QM and MM atoms. The QM atom requires this bond to be described in terms of orbitals in order to obtain a correct valence shell. The MM atom, however, for which electrons are not treated explicitly, does not allow the calculation of such a bonding orbital.

The approximate solution implemented in CHARMM, is to introduce dummy atoms, which are referred to as link-atoms (Field *et al.*, 1990). These link atoms are treated as hydrogen atoms in the QM system, whereas they are invisible to the MM atoms (Figure 2.2). In addition MM energy terms are used to calculate QM/MM bonded interactions. In this way realistic geometries can be obtained for the QM and MM atoms at both sides of the QM/MM boundary.

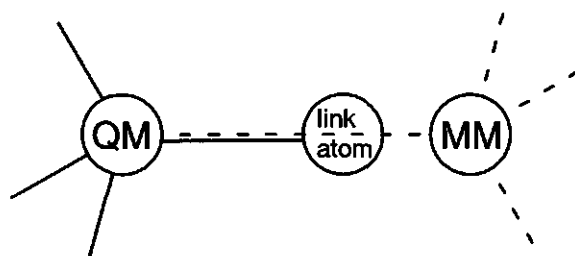


Figure 2.2 A link atom can be used to obtain realistic structures around covalent bonds crossing the QM/MM boundary. No interactions between the link atom and MM atoms are calculated. The QM/MM covalent bond is accounted for by an MM type harmonic energy term. Furthermore, MM type angle energies are calculated for all angles and dihedrals involving one or more MM atoms, to ensure a realistic geometry around the QM/MM boundary.

The dielectric constant in proteins

The strength of electrostatic interactions between charged particles depends on the dielectric properties of the medium in which the particles are present. In water, for example the dipolar water molecules interact with, and adjust themselves around, a charged moiety, thereby shielding it and diminishing its interaction with other charged particles. Consequently, electrostatic interactions in water are 80-fold weaker than in vacuum, which is accounted for by a dielectric constant of 80, in the classical Coulomb law. In proteins, the presence of water molecules as well as the polarisation of protein atoms and the reorientation of polar groups lead in a similar way to a decrease of electrostatic interactions relative to vacuum. The actual dielectric effect of a protein on charge-charge interactions, however, depends strongly on the specific local environment, such as its hydrophobicity and the presence of polar groups (Warshel, 1978). In many MM methods, which do not account for polarisation of atoms explicitly, a dielectric constant of up to 4 is considered a reasonable scaling of the electrostatic interactions in proteins (Grootenhuis and Van Galen, 1995). Some MM methods provide an option to use a "distance dependent dielectric constant". (For example $\epsilon = 1$ at a distance of 1\AA , $\epsilon = 10$ at a distance of 10\AA .) Such a treatment takes full account of short distance interactions, which yields more realistic results for hydrogen bond interactions, whereas the interactions are scaled down as the distance increases, to account for dielectric screening.

QM methods do not allow the scaling of the electrostatic type interactions by such a classical parameter as the dielectric constant. In QM/MM calculations, therefore, the use of a dielectric constant of 1 for the MM and QM/MM electrostatic interactions can be considered to be most consistent. This means, however, that the electrostatic

interactions are not corrected for dielectric screening effects and this may result in an overestimation of long range MM and QM/MM electrostatic interactions (Mulholland and Richards, 1997, Ridder *et al.*, 1999a). Nevertheless, in this thesis, some correction was applied by means of a cut-off of the non-bonded interactions described in the next paragraph (see also chapter 7).

Cut-off of the non-bonded interactions

In MM methods, often the VDW and electrostatic interactions are not calculated for all non-bonded atom-pairs. In principle, the number of non-bonded interactions would increase with the square of the number of atoms. In practice, since these interactions become weaker as the distance between the atoms increases, a cut-off distance is defined, beyond which the non-bonded interactions are neglected. To prevent that such a cut-off results in a discontinuous non-bonded potential (especially the electrostatic interactions may sometimes not be negligible at the cut-off distance) the interactions are usually smoothly scaled down, using various mathematical functions.

In the QM/MM approach, the use of such a non-bonded cut-off is not as trivial as in pure MM methods. The QM system should preferably interact with a well-defined set of MM atomic cores, i.e. all QM-atoms should 'feel' the same MM atomic cores. Thus, the non-bonded QM/MM interactions that are neglected on the basis of a cut-off distance, and the scaling of the remaining interactions, should be the same for all QM atoms independently of their individual positions in the system. In the QM/MM models described in the present thesis, this was done by applying the cut-off for the QM/MM interactions on the basis of the geometrical centre of all QM atoms together, instead of for the individual QM atoms. This is called a group-based non-bonded cut-off.

Studying approximate reaction coordinates

When studying reaction mechanisms it is essential to investigate how one reaction intermediate transforms into the next intermediate. A reaction pathway can be defined as the minimum energy path (on the multidimensional potential energy surface) connecting the reactants and products. The geometric changes along such a reaction pathway, are considered to be a function of the *reaction coordinate*, a parameter which indicates the progress in the transformation from reactant(s) to product(s). The most important point on the reaction pathway is the transition state, the intermediate structure with the maximum potential energy, since the energy difference between the transition state and the reactant state represents the activation energy of the reaction. The transition state can be identified as a saddle point on the potential energy surface. A saddle point corresponds to a structure that has a minimal energy with respect to all

degrees of freedom, except for the one corresponding to the reaction process (for which it has a maximal energy). Once a transition state is found, the intrinsic reaction pathway is found as the steepest descent down to the reactants and products (Fukui, 1981, Gonzalez and Schlegel, 1989). Often, the calculation of an intrinsic reaction coordinate is computationally too expensive for molecular systems of the size of proteins. A relatively simple and approximate way to study reaction pathways in large systems is to define a suitable, approximate reaction coordinate. This approximate reaction coordinate is used to deform the (optimised) molecular model in a series of small steps. At each step, energy minimisation is performed to allow the model to adapt to the local change in geometry. This procedure, which does not include entropic contributions, is referred to as adiabatic mapping (McCammon and Harvey, 1987). Within the CHARMM program, approximate reaction coordinates, defined in terms of atomic distances, can be applied using the RESD (REstrained Distances) method (Eurenius *et al.*, 1996). With this method, atomic distances (d_i), or linear combinations of them, can be harmonically restrained (with a force constant k) to a reaction coordinate parameter (r), by adding the following restraint term to the total energy:

$$E_{\text{RESD}} = k \left(\sum_i (a_i d_i) - r \right)^2$$

The weighting factors in the linear combination of interatomic distances (a_i) may be positive or negative, such that sums as well as differences of interatomic distances can be restrained. In addition, more than one restraint-term can be defined. This allows the independent variation of multiple reaction coordinates. Chapter 8 gives an example in which two reaction coordinates were systematically varied on a grid, resulting in a 2-dimensional (approximate) potential energy surface. Such multi-dimensional potential surfaces can be used to obtain information about the order in which the reaction processes occur (Cunningham *et al.*, 1997, Harrison *et al.*, 1997).

The application of the QM/MM method in biochemical research

The present chapter reveals that, in order to simulate reactions of systems as large and complex as protein molecules, approximations need to be made at various levels of theory. It is clear that, at the current state of the art, QM/MM calculations can not be expected to give absolute outcomes. Therefore, in order to successfully apply the QM/MM method in biochemical research, the following considerations are important.

In the first place, it can be important to test the QM/MM method for the specific enzyme system of interest. Test-calculations can be performed on essential fragments of the QM/MM model. Results obtained for these small fragments can be compared to higher-level *ab initio* calculations (or to relevant experimental data if these are available).

In the second place the results obtained from QM/MM calculations should be used in a comparative way. For example, in this thesis, the energy barriers obtained from QM/MM reaction pathway calculations are used as relative parameters, instead of as absolute activation energies. Other outcomes of the QM/MM calculations, such as Mulliken charge distributions on the QM atoms, and interaction energies between the QM system and surrounding MM residues (analysed to gain insight into the role of individual amino acids and solvent residues) are interpreted in a qualitative way.

Finally, it is important to validate as many as possible outcomes of the QM/MM model by comparison to experimental data. First of all, the optimised geometry of a QM/MM model can be carefully examined and compared to the experimental structure. For example such a comparison may help to identify protonation states of ionisable amino acids, different from the MM standard values (Mulholland and Richards, 1997) (see also chapter 8). In general, a close similarity between the QM/MM optimised geometry and the experimental structure, is a first indication that the QM/MM potential can be successfully applied to the system of interest. Furthermore, other outcomes may be compared to experimental results. In the present thesis, the energy barriers obtained for the QM/MM reaction pathways are correlated to the logarithm of experimental rate constants for series of substrates, to establish whether essential chemical features of the reaction mechanism are correctly accounted for in the QM/MM model.

Once the QM/MM model has been tested and validated it can provide detailed information on the reaction mechanism of the simulated reaction step and on the role of individual active site residues in the catalytic function of the protein. This kind of information is often difficult (or impossible) to obtain from experimental techniques. Provided the considerations mentioned above are taken into account, the QM/MM simulations can therefore yield valuable information to complement experimental biochemical research.

3

Computational methods in flavin research

Lars Ridder, Han Zuilhof, Jacques Vervoort and Iivonne M.C.M. Rietjens
Methods in Molecular Biology 131, 207-228 (1990)

3.1 Introduction

With the continuously increasing power of computers, quantum chemistry is becoming a valuable theoretical tool in enzyme research. Molecules as large as flavins can now be treated by computational methods of reasonable theoretical level. The present chapter focuses on the possibilities and restrictions of some quantum chemical methods with respect to research on the chemistry of flavin cofactors in enzyme catalysis.

Flavin cofactors are involved in a wide variety of enzymatic reactions like oxidation, reduction and monooxygenation. Consequently, flavin cofactors occur in many different forms including for example the oxidised, semiquinone and reduced states and the C4a-(hydro)peroxyflavin form (Figure 3.1). In flavoprotein research the geometrical and electronic properties of all these different flavin species are of interest. With the use of quantum chemical methods many properties, such as optimal geometry, redox potential, charge distribution, dipole moment and the localisation of the electrophilic and nucleophilic reactivity can be calculated for the different flavin molecules. Furthermore, reactions of flavins with other compounds can be studied and transition states and reaction intermediates can be characterised. These theoretical techniques will become increasingly important to complement experimental results, especially in the case of unstable reaction intermediates which are difficult to investigate by experimental techniques.

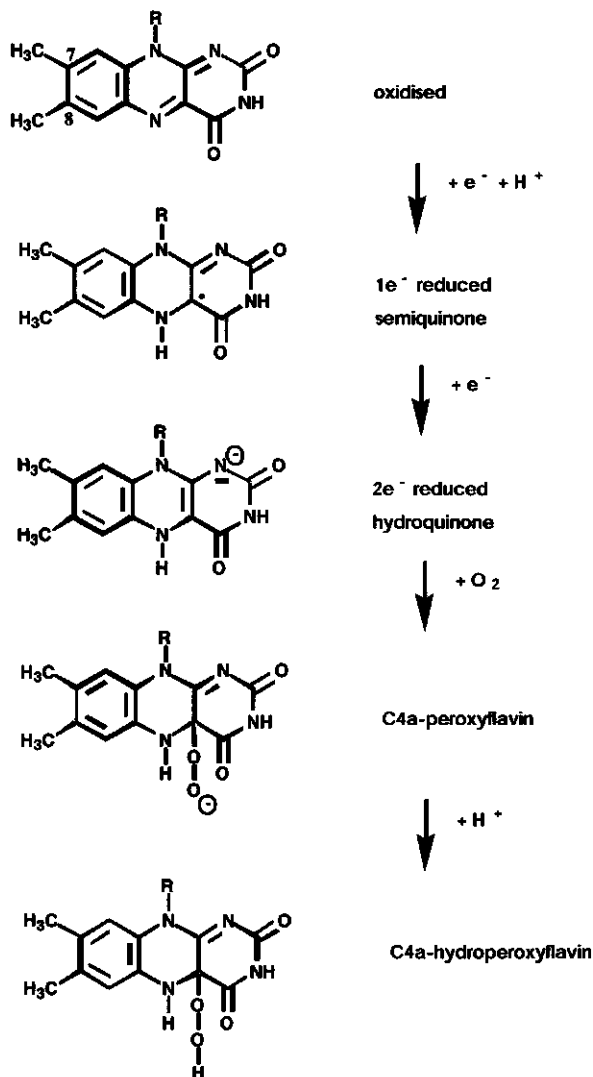


Figure 3.1 Various forms of the flavin cofactor relevant for flavin based biochemical reactions.

3.2 Materials

3.2.1 Software

Quantum chemical computations can be performed with a variety of (often commercially available) programs. Which program is most useful depends on the specific goal of the computation, which itself is linked to the quantum chemical method used (vide infra, section 3.3.2). Reasonably user-friendly programs, e.g.

Spartan (see for specific information: <http://www.wavefun.com>), Hyperchem (<http://www.hyper.com>) and ChemOffice (<http://www.camsoft.com>), allow one to draw the structure of the species of interest in a convenient way. Subsequently, the structures can be 'cleaned up' via the use of very fast molecular mechanics options (such as MM2 or MMX), which may be desirable for some purposes. With molecular mechanical methods, no electronic information can be derived, but the geometry can be estimated relatively efficiently. In the next step, these programs can be used for the quantum chemical calculations themselves, via simple pull-down menu-options. Within ChemOffice semiempirical methods such as AM1, PM3 and MNDO/d are available, as the full version of this program includes MOPAC97. The same options are also available within Spartan and Hyperchem, which are in many respects similar programs. The graphical interface of ChemOffice is, dependent on which property is studied, varying from good (pictures of the orbitals) to very good (renderings of the geometry), while Hyperchem and Spartan have overall a very good graphical display and output. ChemOffice and Spartan/Hyperchem differ significantly in other aspects: ChemOffice can take 2-dimensionally drawn chemical structures (such as given in Figure 3.1) and convert them into readable input structures for the computations, while the strength of Hyperchem and Spartan is that they include more advanced computational methods. Besides the semiempirical methods, they also include *ab initio* methods -such as Hartree-Fock (HF) or second order Møller-Plesset (MP2)- and density functional methods (DFT), including both relatively fast local DFT methods such as the Vosko-Wilk-Nusair method (VWN) and more accurate (and slower) gradient-corrected methods such as Becke-Perdew (BP) and its perturbational variation pBP (in Spartan).

The latter types of computations can also be performed with a wide variety of quantum chemical programs that are optimised for use with larger *ab initio* or DFT computations. The most well-known of these is Gaussian (<http://www.gaussian.com>), with which one can do almost all types of quantum chemical computations. This program does not provide a graphical user interface, but its output can be visualised easily with for example ChemOffice and especially Spartan. Other programs that can handle relatively large molecules such as flavins via *ab initio* or DFT methods include Jaguar (<http://www.schrodinger.com>) and Q-Chem (<http://www.q-chem.com>). The results of the latter can be visualised easily with especially Hyperchem.

3.2.2 Hardware

Molecular mechanics and semiempirical quantum chemical computations can conveniently be run on standard PC's (Pentium and later, with Windows 95/98/NT) or Macintoshes (PowerPC and later, with MacOS 7/8) using either ChemOffice,

Hyperchem (PC only) or PC Spartan/MacSpartan. Minimum system requirements for efficient computations do usually include 16 MB of available physical RAM (besides what is used by the operating system). For *ab initio* or DFT computations relatively fast processors are required. Gaussian and Hyperchem are practically useful for this purpose on PII PC's, G3 Macintoshes and later systems. This offers the advantage of the usage of the same operating system as used standardly on such machines (Windows 95/98/NT or MacOS 7/8). However, in case of computations on molecules as large as flavins, UNIX-machines are often preferable. Typical machines that can be used for this aim are Silicon Graphics computers such as the O2 or Octane, IBM/R6000 computers or DEC/Alpha. Such machines are the preferred type of hardware for use of Gaussian, Q-Chem and Jaguar, and also for Spartan and Hyperchem if *ab initio* or DFT computations of systems of the size of flavins are frequently used options. The latter two programs can also run on fast PC's with adequate amounts of RAM (typically 64 MB minimally required, but 128 MB or more advisable) and hard disk space (typically 2 GB minimally), given that Linux is used as operating system. This is a frequently significantly cheaper alternative to the use of work stations, for all but the largest class of computations. At present no really stable version of Linux is well-tested for the Macintosh. When Linux becomes available for Macintosh, all these programs can also be efficiently used on the G3 Power Macintosh machines, as their processor speed approaches that of standard UNIX-based work stations.

3.3 Method

The method for performing quantum mechanical calculations on flavins, but also on other molecules, include the following basic steps:

1. Define the starting geometry.
2. Select a calculation model.
3. Define and select the output parameters required.
4. Run the calculation job.
5. Validate and analyse the results.

3.3.1 Define the starting geometry.

The input geometries for calculations on flavin structures can be defined basically in two ways. First, the initial geometry can be designed by applying chemically reasonable atomic distances, angles and dihedral angles. In case of a text-oriented user interface (Gaussian) the distances, angles and dihedrals required to construct the molecule are entered in a so-called Z-matrix format. It is, however, often easier and faster to use a program with a graphical interface which provides a molecular editor,

enabling the user to build the molecule interactively and including chemically reasonable bonds and angles automatically. These editors are sometimes supplemented with libraries of molecular fragments to further speed up the molecular building process. The structure built in this way can subsequently be submitted in the program, or saved in a format suitable for submission in another program, for example Gaussian.

The second approach to define the input geometry is to use existing structural data. For a growing number of compounds including proteins, crystal structure coordinates are available from the Cambridge crystallographic data base (*dat* files) and the Brookhaven databank (*pdb* files), the latter containing the crystal structure coordinates of over 50 different flavin-containing proteins. Such structure files or parts of these files, can be used as input by most programs, either by importing them directly, or after conversion to a recognisable format. Often, the crystallographic data do not contain hydrogen coordinates. These must be added by using either one of the building procedures mentioned above, or by using an automatic hydrogen building routine provided by a number of molecular modelling packages (e.g. ChemOffice, Insight II or Cerius²).

3.3.2. Select a calculation model

Once the starting geometry has been defined one can use this structure for the quantum mechanical calculations. Selection of a calculation model involves considering the following questions.

1. Is the initial geometry relevant as such (for example in case of a crystal structure) or is (partial) geometry optimisation required ?
2. Is the system of interest charged? What is its multiplicity, i.e. how many unpaired electrons are present ?
3. Is relevant information obtained from a calculation in vacuum or should the effect of the environment be taken into account ?
4. Which quantum chemical method is best to use for the desired goal?

1. Is the initial geometry relevant as such (for example in case of a crystal structure) or is (partial) geometry optimisation required ? In many cases the input geometry may be relevant as such, for example when it is derived from an experiment (X-ray, NMR) or when it is optimised using another adequate computational method. A single point energy calculation is then sufficient to obtain all the electronic characteristics of the molecule.

Alternatively, when the initial structure is not based on available structural information, optimisation of all atomic coordinates is preferable, leading to the

energetically favourable geometry. Full geometry optimisation is essential when the vibrational modes of the molecule are studied, for example in relation to spectroscopic properties, or in case of identifying a reaction coordinate.

In some cases geometric constraints can be helpful to obtain a relevant geometry. For example when the calculation is performed on an isolated fragment (e.g. an active site or a flavin molecule) on which, in reality, geometric constraints are imposed by the surrounding structure (e.g. an enzyme). For these purposes most programs provide possibilities to freeze atomic coordinates and to constrain atomic distances and angles to specific values during the geometry optimisation. These constraint values can be derived on the basis of a relevant crystal structure or may be obtained from molecular mechanical calculations on the complete structure in which the isolated fragment is embedded.

2. Is the system of interest charged? What is its multiplicity, i.e. how many unpaired electrons are present? The number of electrons, which is an essential input parameter for the calculation, is calculated by most programs on the basis of the nature and number of atoms in the molecule in combination with the charge defined. Another input parameter of interest, which is related to the number of electrons, is the number of unpaired electrons present. This characteristic is indicated by the multiplicity of the system ($\text{multiplicity} = \text{number of unpaired electrons} + 1$). This implies that when calculations on an anion, cation, or radical flavin species are of interest, the input for the calculations should define this by setting the correct charge and multiplicity values. By default most programs assume zero charge and the absence of unpaired electrons ($\text{multiplicity} = 1$). Charge and multiplicity numbers given should be compatible. Furthermore, the presence of unpaired electrons poses some additional requirements on the method to be used. By default most programs use the restricted Hartree-Fock method for singlet molecules, while unrestricted Hartree-Fock is used for molecules with one or more unpaired electrons.

3. Is relevant information obtained from a calculation in vacuum or should the effect of the environment be taken into account? Quantum chemical calculations are generally performed in vacuo (or 'in gas-phase') i.e. assuming a dielectric constant of 1 and not including possible effects of the surroundings. Many properties of molecules can be derived from the in vacuo calculations, even if one is actually interested in their properties as solid, liquid or solute. Especially when a series of molecules is compared, in vacuo calculations often work well. However, in some cases it may be very important to include some effects of a specific environment. For example, polar solvents, with a high dielectric constant, reduce the electrostatic effects of local charges (dielectric screening) and drastically influence the energy of charged species. This effect is important when comparing differently charged molecules in solution or

when energy differences are studied between intermediates of a reaction involving a redistribution of charge. A common way to take the effect of polar solvents into account is to include a so-called *reaction field* of uniform dielectric constant, surrounding the molecule. The polarisation of this reaction field by the molecule results in an electric field, which in turn influences the electronic structure of the molecule (Foresman and Frisch, 1996).

When more specific interactions with the surroundings are important, for example in the case of hydrogen bonds, it is preferable to include these interacting molecules explicitly in the structure to be calculated. However, the additional number of atoms, included in the quantum chemical calculation to model a specific environment, is limited due to rapidly increasing computational demands (as will be discussed in section 3.2.4). In connection to this limitation it should be considered that when only a small part of, for example, an active site of an enzyme is included in the model, the geometric constraints imposed by the (rest of the) surrounding protein should be reproduced by well-chosen constraint parameters in the input (as discussed in 3.2.1.).

A simplified approach to include the electrostatic effects of surrounding atoms, as for example implemented in the Gaussian program, is to define an approximate electrostatic potential field by means of fixed point charges around the molecule (Foresman and Frisch, 1996). In more recently developed approaches, the environment surrounding the quantum chemical system, is fully modelled by a molecular mechanical treatment. In these so-called combined quantum mechanical and molecular mechanical (QM/MM) methods electrostatic as well as steric effects of the surrounding molecular model are included in the quantum chemical calculation (e.g. 2). Although promising, these QM/MM methods are not yet widely used and various implementations are still in the process of development.

4. Which quantum chemical method is best to use for the desired goal ? Quantum chemical methods vary in two aspects. First, different methods are present to handle the Hamiltonian operator in order to solve the Schrödinger equation. Second, the form and number of mathematical functions, which are used to describe the wave function and electron density of the molecule, are variable; these functions are organised in so-called basis sets

Semiempirical methods, such as AM1 (Dewar *et al.*, 1985) and PM3 (Stewart, 1989), introduce a number of approximations in solving the Schrödinger equation to speed up the calculations drastically. Consequently these methods can handle molecules containing up to 300 atoms relatively fast, but results should be used qualitatively or to study relative effects within series of related compounds.

Ab initio methods offer the advantage of systematic convergence towards the exact solution of the Schrödinger equation via increase of electron correlation effects and

basis set increase. However, this does not automatically happen in a straightforward way. Three methods can be commented on. Hartree-Fock computations are the most well-known of these, and give reasonable descriptions of both the geometry and electronic structure of many closed-shell molecules. For geometries, improvement over semiempirical methods for this type of species is usually small, so if geometrical information is the primary interest, use of semiempirical methods can save hours of CPU time. When chemical reactions, i.e. formation and breaking of bonds, are to be studied, the Hartree-Fock method is in principle not sufficient. Two improvements are possible in going from Hartree-Fock to methods with better accounts of electron correlation. Most classically this is done via perturbation theory, which is implemented in many programs via Møller-Plesset theory (specifically MP2). For neutral closed-shell species this gives a very good description of both geometry and electronic structures. The price is, however, a significant increase in computational demands, which renders this method usually completely impractical for systems of the size of flavins. A second, interesting approach consists of DFT methods. Especially the so-called hybrid-methods such as B3LYP are recommended for both geometrical and electronic structures, while also for quantitative agreement with experiment DFT methods often perform better than both semiempirical and *ab initio* HF methods.

Besides the applied method, the basis set used to describe the wave function is an important factor determining the quality of the calculation. The semiempirical methods all use a minimal basis set, containing the minimum number of basis functions needed for the valence electrons only. *Ab initio* methods can also be used with a minimal basis set, referred to as the STO-3G basis set, containing the minimum number of basis functions needed for all electrons. However, larger basis sets have been developed, imposing fewer restrictions on the spatial distribution of the electrons (and therefore resulting in a better approximation to the exact wave function). The first extensions to the minimal basis set are the so-called split valence basis sets, such as 3-21G, 6-31G and 6-311G, which include two or more basis functions, with different sizes, for each valence orbital. In the 3-21G basis set, for example, each inner shell orbital is described by one basis function (obtained by combining 3 gaussian functions), while the valence shell consists of two sets of basis functions (built from 2 and 1 gaussian functions respectively). These basis sets can be further extended by adding so-called polarisation functions, i.e. basis functions of higher angular momentum quantum number than required to describe the ground state of each atom (e.g. p-type functions on hydrogens and d-type functions on C, O and N, etc.) These polarisation functions allow for the possibility of small displacements of electronic charge away from the nuclear positions and are denoted by adding the type of the additional function in

Table 3.1 Time required for full geometry optimisation of the oxidised flavin by different quantum mechanical methods. The input geometry was created using the build option in Spartan. * = polarisation functions (see text)

method	basis set for		no of basis functions	CPU time	time relative to PM3
	H atoms	C,O,N atoms			
semiempirical					
RHF/AM1	1s	2s 2p _x 2p _y 2p _z	88	1 min 51 sec	1.24
RHF/PM3	1s	2s 2p _x 2p _y 2p _z	88	1 min 30 sec	1
ab initio					
RHF/STO-3G	1s	1s 2s 2p _x 2p _y 2p _z	107	6 h 6 min	244
RHF/3-21G*	1s 1s	1s 2s 2p _x 2p _y 2p _z	195	18 h 28 min	739
RHF/6-31G*	1s 1s	2s 2p _x 2p _y 2p _z 2s 2p _x 2p _y 2p _z 3d _{xx} 3d _{yy} 3d _{zz} 3d _{xy} 3d _{xz} 3d _{yz}	309	113 h 11 min	4527
density functional					
pBP86/DN	1s 1s	1s 2s 2p _x 2p _y 2p _z 2s 2p _x 2p _y 2p _z	195	23 h 47 min	951
pBP86/DN*	1s 1s	1s 2s 2p _x 2p _y 2p _z 2s 2p _x 2p _y 2p _z 3d _{zz} 3d _{zx} 3d _{yz} 3d _{xx} 3d _{xy}	290	40 h 44 min	1629
pBP86/DN**	1s 1s 2p _x 2p _y 2p _z	1s 2s 2p _x 2p _y 2p _z 2s 2p _x 2p _y 2p _z 3d _{zz} 3d _{zx} 3d _{yz} 3d _{xx} 3d _{xy}	326	47 h 47 min	1911
BP86/DN*	as for pBP86/DN*		290	54 h 20 min	2173

brackets, e.g. 6-31G(d) or 6-31G(d,p), or with an asterisk, e.g. 6-31G* or 6-31G**, respectively. (A brief overview is given by Foresman and Frisch (1996).)

Table 3.1 illustrates the performance of different methods and basis sets when applied on the flavin cofactor in its oxidised form with a CH₃ replacing the ribityl side chain. Full geometry optimisation was performed, using a Silicon Graphics O² R10K workstation. Table 3.1 also presents the basis functions, used to describe the H, C, O or N atoms in the various basis sets.

From the data listed in Table 3.1 it is clear that, with an increase in basis set, the time required for the calculation increases rapidly, thereby limiting the size of the molecules that can be handled by a specific method in the available amount of time. Clearly, one should choose a method by finding a balance between quality and costs. In the case of flavin molecules it is advisable, from a practical point of view, to start with the semiempirical models. They give a first, qualitative, indication whether it is

Table 3.2 Quantum mechanical parameters that can be calculated, and their relation to characteristics of flavin based reaction chemistry

calculated parameter	flavin based reaction characteristic
optimised geometry	geometry of cofactor in its various reactive forms: flat, butterfly or bend flavin ring, orientation of substituents geometry of substrates
energy of highest occupied molecular orbital = E_{HOMO}	ionisation potential of reduced or of semiquinone flavin nucleophilic reactivity of the cofactor or of substrates
energy of lowest unoccupied molecular orbital = E_{LUMO}	reduction potential of semiquinone or of oxidized flavin electrophilic reactivity of the cofactor or of substrates
heat of formation	differences in heat of formation between starting and transition state geometry or between starting and end geometry can be correlated to activation or reaction enthalpies, to the rate of a reaction and/or to pK_a values
HOMO/HOMO-1 distribution	indicates sites of nucleophilic reactivity within the cofactor and/or substrate
LUMO/LUMO+1 distribution	indicates sites of electrophilic reactivity within the cofactor and/or substrate
charge distribution	relates to attractive/repulsive coulomb interactions
spin distribution	relative stability of semiquinone flavin form, as spin delocalisation stabilises the radical relates to electrophilic/nucleophilic reactivity of the sites within the radical
bond orders	indicates to what extent bonds are formed or broken, for example in reaction pathway geometries

worthwhile to perform more time-consuming *ab initio* or density functional methods with more extended basis sets.

3.3.3 Define and select the output parameters required.

Quantum chemical calculations may produce a large amount of output data. Therefore it is important to decide which output is relevant for the biochemical mechanism underlying the activity of the flavin cofactor studied. Many programs offer the possibility to select the data to be listed in the output and the analysis tools to be run at the end of the calculations. The selection proceeds either interactively or using keywords. It is advisable to choose not only the unknown characteristics of interest but also those electronic properties that allow comparison to existing (experimental) data. Table 3.2 lists various of the parameters that can be calculated by quantum mechanical calculations, and links them to characteristics relevant for flavin-based reaction chemistry. In addition, some programs can generate graphical displays of the

optimised geometries as well as of the calculated molecular orbital, spin and charge distributions. Although the numerical output is frequently most relevant, the visualisation of abstract quantities is often helpful. In the Notes section several examples are presented, to illustrate the use of the various quantum mechanically calculated molecular orbital parameters for flavin based research.

3.3.4 Run the calculation job

After the starting geometry, method and all output to be generated have been defined the calculation can be started. Depending on the size of the molecule and the method chosen the calculation will take seconds, hours, days or even weeks. While running the calculation, the output generated reports the status of the calculation and can be monitored. When a geometry optimisation is performed the output reports the results of each optimisation cycle, for example in terms of the energy gradient which indicates to what extent the geometry is converging to an energy minimum. Since the geometry optimisation is the most time-consuming part of many quantum mechanical calculations this also indicates when the calculation is about to be completed. When the convergence values fall beneath the optimisation criteria defined in the program, the geometry optimisation stops.

3.3.5 Validate and analyse the results

Validation and analysis of the results is an important step in the application of quantum mechanical calculations to biochemical problems. Since the programs started to be used by biochemists during the past decade, validation of the data obtained for biochemical systems is still an essential part of research in this field. Most of the effort is still devoted at comparing geometries, reaction pathways, energy barriers and rates to data obtained from biochemical experiments. More and more calculations may then start to be of use not only as a tool for interpretation of experimental data, but also yielding additional information, for example when experimental data is unavailable due to the short life-time of reaction pathway intermediates. The next section includes some examples of validation of computational results in studies on flavin dependent systems, showing the potential of this technique in flavin research.

3.4 Notes

3.4.1 Introduction

Although some people tend to say that a computer experiment is always successful, this is not the case. First of all a calculation may fail due to inconsistencies in the input. For example, the multiplicity defined could be impossible for a given atomic

structure and charge, the required parametrisation or basis set may not be available for all the atoms of the system, or the method chosen may not be compatible with the size of the system, in relation to the computer memory and power available. Second, a calculation may fail when the geometry does not converge to an energy minimum. This may be related to the complexity of the molecule and the optimisation routine implemented in the program. Some programs provide options or keywords to control some of the optimisation parameters, such as for example the step-size and the maximum number of optimisation steps required. When optimisations fail one might consider to apply less precise optimisation criteria, or use a different input geometry. Finally, the molecular orbital wave function may not converge to a "self-consistent" solution of the Schrödinger equation. This problem may be solved by defining a different initial guess wave function, obtained for example from a lower level calculation. In addition some programs provide a "damping" option which can be used to prevent oscillations in the self-consistent field procedure.

The computational failures discussed above are generally detected by the program and can often be solved. More serious failures, however, which require chemical knowledge to detect, relate to the fact that the generation of large output files does not necessarily imply the correctness of the data. Thus, a calculation can be "successful" but this does not guarantee that its results are "reasonable" and have a physical meaning. Several reasons implicit to the calculations themselves can be the cause of these failures. The following paragraphs discuss some problems or faults to be aware of, and strategies to solve and/or overcome these problems.

3.4.2 Geometry optimisation

Geometry optimisation is a critical aspect of the application of quantum chemistry on large systems. First, each optimisation step may demand significant computational effort. Second, a large number of optimisation steps may be required to find an energy minimum. Third, the optimised geometry could be a so-called local energy minimum, instead of to a global energy minimum.

A common approach to limit the computation time for large molecules is to perform geometry optimisation at a lower level of theory, followed by a single point calculation at a higher level of theory with a more extended basis set. Although this approach is often satisfactory, it should be noted that it does not always give good results. One should especially be aware of potential errors of the semiempirical methods in equilibrium geometries. Geometries representing an energy minimum on the semiempirical potential surface are possibly far off the corresponding minimum on the *ab initio* potential surface. This is illustrated in Figure 3.2. Thus, results obtained from a single point calculation at the Hartree-Fock level using the geometry optimised with

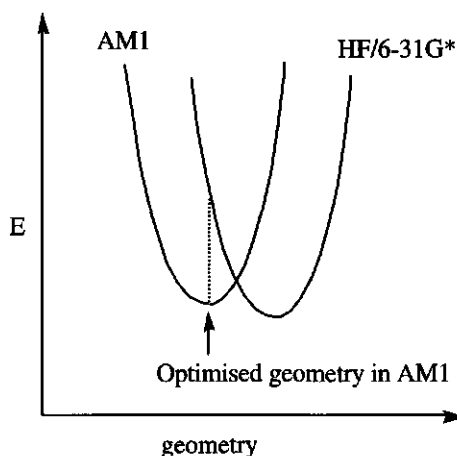


Figure 3.2 The effect of different potential energy surfaces: an optimised geometry using a semiempirical method (e.g. AM1), may represent a distorted structure using the DFT or *ab initio* methods (e.g. HF/6-31G*).

a semiempirical method may represent a distorted system and may not be relevant for the equilibrium situation. On the other hand the optimum geometries obtained with *ab initio* methods, going from Hartree-Fock 3-21G* to higher level calculations, show a somewhat more systematic and often smaller variation. Therefore, higher level (e.g. MP2) energy calculations on Hartree-Fock 6-31G* geometries are generally believed to give reasonable results.

Another approach to limit the number of optimisation steps is to perform partial optimisation by fixing part of the system to its initial coordinates or by applying geometric constraints. However, on the basis of the arguments given above, the quality of the results depends largely on the initial geometry. Constraints on bond lengths and angles should in general be avoided in equilibrium geometry calculations, where CPU-capacity allows this.

With increasing complexity of a molecule the chance to obtain a local minimum instead of a global energy minimum increases. Among the available methods to obtain a global energy minimum are Monte Carlo techniques and genetic algorithms which investigate and compare many possible conformations. In practice, all quantum mechanical methods, including even the semiempirical molecular orbital methods are likely to be too time-consuming for these jobs and molecular mechanics methods should be applied. Then, the best geometries obtained by these molecular mechanics methods can be used as input for quantum mechanical energy minimisations. Alternatively, one may choose to use a relevant crystal structure geometry, since this geometry is (close to) a real life energy minimum. However, this cannot be done for

transition state geometries for which crystal structural data are unavailable. Transition state geometries need to be verified by means of a vibrational analysis, to check whether the vibration corresponding to the reaction coordinate (with an imaginary/negative eigenvalue) corresponds to the expected reaction. Even better, but often very time-consuming, is to calculate an intrinsic reaction coordinate starting from the transition state, which should lead to correct reactant and product geometries.

3.4.3 Absolute versus relative outcomes

Computations on complex flavin systems can at present not be expected to yield reliable results for all calculated properties in an absolute way. Some calculated parameters, like bond lengths and angles, are frequently close to their physical values, but many electronic parameters are only useful on a relative basis. This implies that most calculations are to be used to study properties within a series of related structures, to correctly predict relative differences. Experience obtained over the past decade has demonstrated that this relative approach, using series of related structures, is effective and can be used to study complex systems in a reliable way, even by the simple and fast semiempirical methods (Cnubben *et al.*, 1994, Peelen *et al.*, 1995, Soffers *et al.*, 1996, Van Haandel *et al.*, 1996). To illustrate this, the next paragraphs present some examples of quantum mechanical computer calculations in flavin biochemistry.

Example a: Redox potentials of a series of C7/C8-substituted flavins. The redox potential, i.e. the energy involved to donate or accept electrons, is an essential property of flavins since they are often involved in electron transfer reactions. The reduction potentials for a large number of flavin molecules have been reported in literature (Eckstein *et al.*, 1993, Hasford and Rizzo, 1998, Ortiz Maldonado *et al.*, 1997). The energy required to take up an electron is to a first approximation related to the energy of the lowest unoccupied molecular orbital (LUMO). Table 3.3 presents the experimental reduction potentials for the two-electron reduction of a series of C7/C8-substituted flavins as reported in the literature, as well as the LUMO energies calculated for these structures using various computational methods. Although the E_{LUMO} values obtained with the various methods differ significantly in an absolute way, they compare very well in a relative way (AM1 vs. HF/3-21G* : $r=0.96$; HF/3-21G* vs. pBP/DN* : $r=0.97$). Furthermore, with all three methods a clear correlation is observed between the calculated E_{LUMO} values and the experimentally determined reduction potentials of the oxidised flavin derivatives (Figure 3.3). The results illustrate that when series of related compounds are studied, the absolute outcomes of the calculated parameters may deviate, whereas good quantitative structure activity relationships can still be obtained. The data also illustrate that the results from fast semiempirical methods can be very useful in these relative studies.

Table 3.3 Redox potential for the two- electron reduction of a series of oxidised substituted flavin derivatives, obtained from the literature (Eckstein *et al.*, 1993, Hasford and Rizzo, 1998, Ortiz Maldonado *et al.*, 1997), and the E_{LUMO} values calculated by several quantum mechanical calculations. For the calculations the N10-ethyl substituted derivatives were used. Reduction potentials reported by Hasford and Rizzo (1998) were converted by +240 mV to be comparable to those reported by Eckstein *et al.* (1993) and Ortiz-Maldonado *et al.* (1997).

substituent at		reduction potential	E_{LUMO} (eV) calculated by		
C7	C8	for two electron reduction (mV)	AM1	HF/3-21G*	pBP86/DN*
CH ₃	NH ₂	-310/ -330	-1.423	0.9951	-3.780
CH ₃	OCH ₃	-260	-1.700	0.8019	-3.938
CH ₃	CH ₃	-208/ -207	-1.619	0.6993	-4.085
CH ₃	H (nor)	-180	-1.647	0.6109	-4.199
CH ₃	Cl	-152	-1.820	0.3720	-4.324
CH ₃	F	-167	-1.821	0.4857	-4.264
CH ₃	SCH ₃	-204	-1.663	0.4234	-4.251
CH ₃	SOCH ₃	-161	-1.903	0.2416	-4.471
CH ₃	SO ₂ CH ₃	- 50	-2.209	0.0607	-4.580
CH ₃	H (iso)	-200	-1.620	0.6466	-4.131
CH ₃	OCH ₂ CH ₃	-246	-1.549	0.8245	-3.908
CH ₃	Br	-148	-1.867	0.3959	-4.338
CH ₃	N(CH ₃) ₂	-254	-1.569	0.7687	-3.810
H	H	-167	-1.683	0.5608	-4.286
H	CH ₃	-199	-1.651	0.6607	-4.161
Cl	H	-131	-1.832	0.3094	-4.474
H	Cl	-145	-1.860	0.3167	-4.416
Cl	Cl	- 98	-1.985	0.1097	-4.550
Cl	CH ₃	-162	-1.793	0.4174	-4.332
F	H	-126	-1.878	0.2879	-4.474
H	F	-159	-1.861	0.4199	-4.376
F	CH ₃	-158	-1.837	0.4076	-4.324
F	Cl	-104	-2.045	0.0680	-4.563
OCH ₃	H	-156	-1.709	0.5442	-4.215
H	N(CH ₃) ₂	-340	-1.396	0.9791	-3.733
H	CN	- 43	-2.118	-0.0797	-4.844
CN	H	- 81	-1.960	0.1497	-4.748
CH ₃	O ⁻	-334	2.225	4.6750	-0.199
CH ₃	S ⁻	-290	1.596	3.7887	-0.542

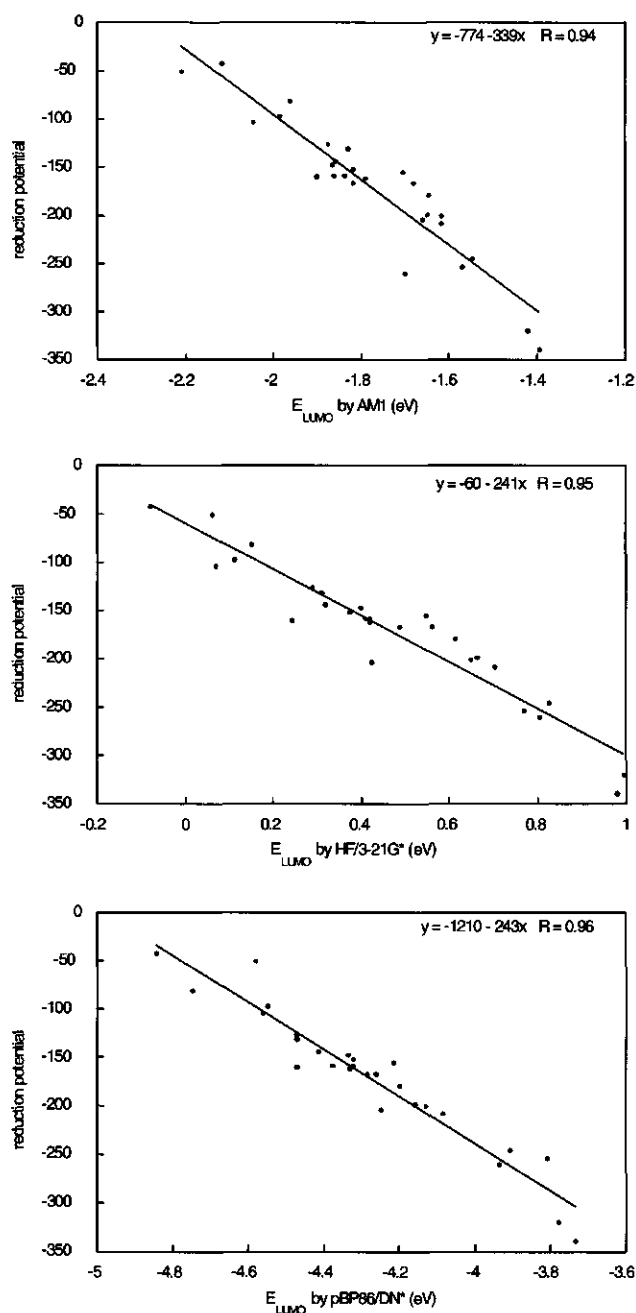


Figure 3.3 Quantitative structure activity relationships for the correlation between the experimentally determined reduction potential for the two electron reduction of a series of substituted flavins (Eckstein *et al.*, 1993, Hasford and Rizzo, 1998, Ortiz Maldonado *et al.*, 1997) and the E_{LUMO} as calculated by a) the AM1 semiempirical method, b) the HF/3-21G* method and c) the pBP86/DN* method.

Example b: Calculated nucleophilic reactivity and activation barriers for a flavin dependent monooxygenase reaction. According to frontier orbital theory, the energies of LUMO and HOMO (highest occupied molecular orbital) are not only related to reduction and oxidation potentials, respectively, but also to electrophilic and nucleophilic reactivity, i.e. the rate by which electrophilic and nucleophilic reactions proceed. As an example, the electrophilic attack by a C4a-(hydro)peroxyflavin cofactor on the C3 of the substrate parahydroxybenzoate, as catalysed by the enzyme parahydroxybenzoate hydroxylase (PHBH), can be studied.

First, the HOMO *density* on the individual atoms within the *p*-hydroxybenzoate-flavin complex is related to their nucleophilic reactivity, relative to the other atoms. Comparison of the HOMO distributions within the substrate in two protonation states (Figure 3.4) illustrates that deprotonation of the hydroxyl moiety is required to obtain substantial nucleophilic reactivity at the C3 position of *p*-hydroxybenzoate required for the electrophilic attack by the cofactor.

Furthermore, the calculated HOMO *energies* of *p*-hydroxybenzoate and four of its fluorinated analogues correlate ($r=0.99$) with the natural logarithm of their experimental rate of conversion (Husain *et al.*, 1980, Vervoort *et al.*, 1992). Since the reactivity of the flavin cofactor is a constant factor in the enzymatic conversion of the series of substrates, the reactivity of the substrates systematically influences the rate of the reaction. Thus, calculations on the series of isolated substrates yield insight in the chemistry of the reaction process.

Calculations are not restricted to structures and characteristics of stable molecules but can also be performed on reactive reaction pathway intermediates for which no experimental data for validation can be obtained. The lack of experimental data on transition state geometries complicates direct validation of transition state and reaction pathway calculations. As an example of reaction pathway calculations in flavin based reaction chemistry we refer to recent QM/MM reaction pathway calculations for the monooxygenation of parahydroxybenzoate by the C4a-hydroperoxyflavin intermediate in the active site of para-hydroxybenzoate hydroxylase (Ridder *et al.*, 1998). Such reaction pathway calculations do not yield absolute activation energies, since entropy is not included. However, the calculated energy barriers for the transfer of the OH moiety from the C4a-hydroperoxyflavin intermediate to the C3 of *p*-hydroxybenzoate and four of its fluorinated analogues, correlate well ($r=0.96$) with the logarithm of the experimentally determined k_{cat} for conversion (Husain *et al.*, 1980), indicating they are useful in a relative approach. Since the fluorine substituent is relatively small, it has (almost) no steric effect, while it influences the electronic characteristics of a reactant to a large extent. Thus, such correlations with kinetic data allow validation of transition state and reaction pathway calculations.

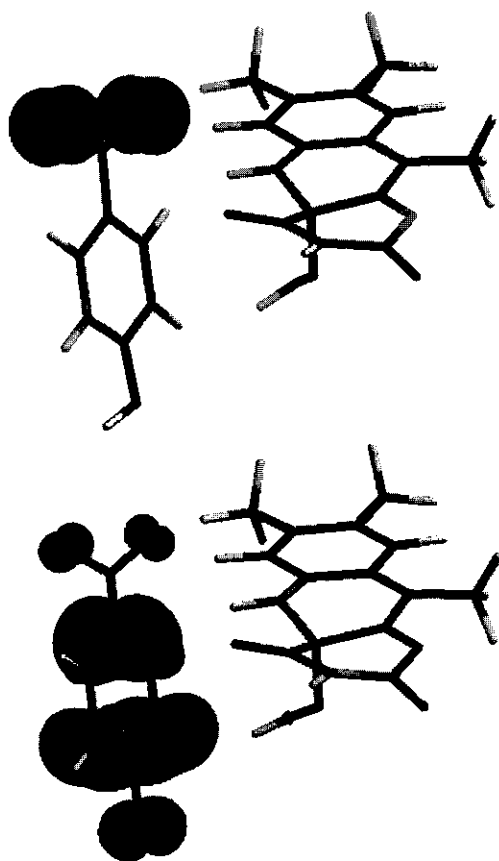


Figure 3.4 Calculated (AM1) HOMO distribution in *p*-hydroxybenzoate flavin complex, with a) a protonated OH moiety, and b) a deprotonated O⁻ moiety. Note that upon deprotonation the HOMO density, representing nucleophilic reactivity, on C3 of the substrate parahydroxybenzoate increases from almost zero to 22 %. The geometry is obtained from a combined QM/MM calculation (12).

It is important to note that the above-mentioned correlations are based on the assumption that the reaction of the C4a-hydroperoxyflavin cofactor with the substrate is the rate limiting step in the overall reaction. If not, another parameter than the calculated nucleophilicity of the substrate would determine the reaction rate.

3.4.4 Influence of charge

Comparison of properties in species with different overall charge requires caution. A striking example can be found in the comparison of the reduction potential data for two C8 substituted flavins, i.e. with R = N(CH₃)₂ (Table 3.3, entry 13) and R = S⁻ (last

entry). The E_{LUMO} of the $R = S^-$ species is significantly higher than that of the species with $R = N(\text{CH}_3)_2$, yet the reduction potentials differ by only 50 mV.

As explained in section 3.4.3, the LUMO energy dominates the capacity of a compound to take up electrons. However, in case of a difference in overall charge two other terms come in for a correct comparison of the systems in solution. First, a difference in charge leads to a difference in solvation energy. The increase in solvation energy is usually different in going from an uncharged species to a negatively charged one, than for addition of a second electron to an already negatively charged species. Second, the addition of electrons has to overcome an increasing electron-electron repulsion, which is larger in going from a negatively charged species to a doubly negatively charged one, than in going from a neutral species to a charged one. These two factors usually cause the reduction of neutral species to proceed more facile than that of analogous negatively charged species, and hamper the use of one straightforward correlation for species with overall different charges.

3.4.5 Meaning of calculated charge distributions

The distribution of charge is not an unambiguously determinable molecular property. A variety of models are available, each with their specific advantages and disadvantages. The differences between several of these models can be very significant, so that differences in charge distributions between different species can only be meaningfully compared when using the same method. For the determination of how the charge of a species will affect its interaction with the surrounding medium (e.g. solvent or protein), so-called electro-static potential (ESP) charges are most significant, e.g. ChelpG or Merz-Kollmann (MK). For determination of the electronic structure of a species natural-orbital derived (NBO or NPA) charges are most reliable. Most programs do supply so-called Mulliken charges by default. These have as their only advantage that they are the most frequently used, but display notorious dependencies of theoretical methods and basis sets. Comparison of Mulliken charges obtained via different methods -even if only slightly different- renders them frequently completely useless. If ESP or NBO-charges are available, these should be preferred for the specific aim chosen.

3.4.6 Influence of method used

It is common to assume that an increase of required computation time more or less automatically leads to an improved description of the system. This is, however, often not true in the comparison of the faster semiempirical methods to low-level *ab initio* calculations. In view of the size of flavin molecules, the applicability of these fast methods on the problem of interest is to be considered first.

The results of semiempirical methods are usually good in a qualitative sense, and very useful in the comparison of data obtained from structurally related compounds. However, in an absolute sense these methods tend to be less reliable for electronic properties, and also energetic differences between different conformations of a molecule should not be taken for granted using these methods. (In fact, if conformational analysis is of interest, molecular mechanics is usually the method of primary choice.) When choosing a specific semiempirical method one should take into account that the quality of the parametrisation (AM1, PM3, etc.) varies for different elements. The elements C, H, N and O -as occurring in unfunctionalised flavins- are usually reasonably well described with the different semiempirical techniques. In contrast, for halogens, in which lone pair interactions can be substantial, the quality is generally improved in the series $AM1 < PM3 < MNDO/d$. The latter treats C, H, N and O, however, via the less reliable MNDO method. Semiempirical methods are generally used within the framework of restricted Hartree-Fock calculations, in which orbitals are filled, starting from those with the lowest energy, with two paired electrons. For neutral molecules with closed electron shells (2 paired electrons in all orbitals) this is generally without problems, but for open-shell species (radicals, radical anions, radical cations or triplet states) this approach does have its limitations. In many programs the default treatment of open-shell species is via the unrestricted Hartree-Fock method, in which the functions describing α or β electrons are not necessarily identical anymore (although the resulting α and β function are usually still virtually identical). This approach works frequently reasonably well as long as the so-called spin contamination, i.e. S^2 , is close to the theoretical value of 0.75 for a doublet spin system (i.e. a molecule with one unpaired electron), or 2.0 for a triplet system. Nearly all of the above mentioned programs will print values for S^2 by default in UHF computations. If the value is > 0.85 for a doublet system, the result should be seriously distrusted.

Semiempirical computations use so-called minimal basis sets. These have as a serious disadvantage that intermolecular interactions (e.g. π - π stacking interactions) are generally poorly described. Semiempirical computations compensate this limitation via element-specific parametrisations. In this way, PM3 approximates hydrogen-bonds to a frequently quite acceptable degree (Jurema and Shields, 1993, Zuilhof *et al.*, 1997). For all species with elements containing lone pairs or (partially filled) d-orbitals, the results of semiempirical methods should be treated with caution, although some success has been claimed for methods including d-orbitals (MNDO/d) and for metals (PM3[tm]).

As indicated in section 3.2.4, *ab initio* methods systematically converge towards the exact solution of the Schrödinger equation upon inclusion of electron correlation and

increase of the size of the basis set. However, for large systems like flavin molecules, the size of the basis set should not be larger than necessary in relation to the problem studied. The use of *ab initio* and DFT computations with minimal basis sets (e.g. STO-3G) is strongly discouraged. The extra effort of *ab initio* or DFT calculations, compared to semiempirical calculations, is only paying off with basis sets such as 3-21G(d), 6-31G(d) and larger. If hydrogen bonds play an important role, then polarisation p-functions on hydrogen atoms are important, and increase to 6-31G(d,p) is recommended. Finally, for the description of negatively charged species, so-called diffuse functions -generally designated via the '+'-sign- should be included in the basis set. Convergence of the wave function with such basis sets is usually slowed down significantly, but electronic features are described more accurately. For anionic systems of the size of flavins the 3-21+G(d) basis would be recommended; if ample computational resources are available, then the 6-31+G(d,p) basis set is the next step. If one chooses to use diffuse functions in the description of anionic systems, direct comparison to neutral species requires computation of the neutrals with diffuse functions as well (Koch, 1998). For neutral species a single point calculation with diffuse functions at the optimised geometry obtained without diffuse functions is often adequate, as the inclusion of diffuse functions does not have major effects on the geometry of neutral species, but is important for electronic structure data and relative electronic energies of neutrals and anions. Further enlargement of the basis set via addition of more polarisation functions is not paying off in most cases.

The exception here is formed by *ab initio* or DFT computations including (transition) metal atoms, which do demand large basis sets for accurate descriptions. Finally, the use of numerical basis sets is in some programs recommended for DFT work (e.g. DN in Spartan). Such basis sets might speed up the computation, but systematic validation of the quality of this approach, specifically for open-shell species, is currently not yet available.

3.4.7 Conclusions

The examples given in section 4.3 indicate that quantum chemical calculations become a useful tool in the field of flavin research. When possible, computational studies on flavins should be defined in a relative way, by investigation of trends in relevant electronic properties within a series of flavin molecules or with a series of substrates converted by the same flavin cofactor. In such problems, semiempirical methods are often very effective. Comparison of differently charged molecules is often problematic, especially in the solute state.

When computational results are to be reliable in an absolute sense, *ab initio* and density functional methods are preferable, provided that the basis set used and the

electron correlation included give a sufficient description of the computational system. Furthermore, for absolute outcomes, one should consider the effect of the relevant environment on the properties of the molecule one is actually interested in. In biochemistry, the relevant environment is often the active site of a protein. A full description of all effects of a protein environment that surrounds a molecule of interest is beyond the current state of computational chemistry. However, the combined QM/MM methods as developed in the last decade are a first approach towards computational chemistry of biomolecules and, provided the results are used on a comparative basis, become an interesting tool in biochemistry.

4

Quantitative structure activity relationship for the rate of conversion of C4-substituted catechols by catechol-1,2-dioxygenase from *Pseudomonas putida (arvilla)* C1

Lars Ridder, Fabrizio Briganti, Marelle G. Boersma, Sjef Boeren, Eric H. Vis,
Andrea Scozzafava, Cees Veeger and Ivonne M.C.M. Rietjens
European Journal of Biochemistry 257, 92-100 (1998)

The influence of various C4/C5 substituents in catechol (1,2-dihydroxybenzene) derivatives on the overall rate of conversion by catechol-1,2-dioxygenase from *Pseudomonas putida (arvilla)* C1 was investigated. Using catechol, 4-methylcatechol, 4-fluorocatechol, 4-chlorocatechol, 4-bromocatechol, 4,5-difluorocatechol and 4-chloro-5-fluorocatechol, it could be demonstrated that substituents at the C4 and/or C5 position decrease the rate of conversion, from 62% (4-methylcatechol) down to 0.7 % (4-chloro-5-fluorocatechol) of the activity with non-substituted catechol. The inhibition was reversible upon addition of excess catechol for all substrates tested. This indicates that the lower activities are neither due to irreversible inactivation of the enzyme nor to product inhibition. Based on the reaction mechanism proposed in the literature (Que, L., Ho, R.Y.N. (1996) *Chem. Rev.* 96, 2606 - 2624), the nucleophilic reactivity of the catecholate was expected to be an essential characteristic for its conversion by catechol-1,2-dioxygenase. Therefore, the rates of conversion were compared with calculated energies of the highest occupied molecular orbital (E_{HOMO}) of the substrates. A clear quantitative relationship ($R > 0.97$) between the $\ln k_{\text{cat}}$ and the calculated electronic parameter E_{HOMO} was obtained. This indicates that the rate-limiting step of the reaction cycle is dependent on the nucleophilic reactivity of the substrate and not sterically hindered by the relatively large bromine or methyl substituents used in the present study. Possible steps in the reaction mechanism determining the overall rate at 20°C are discussed.

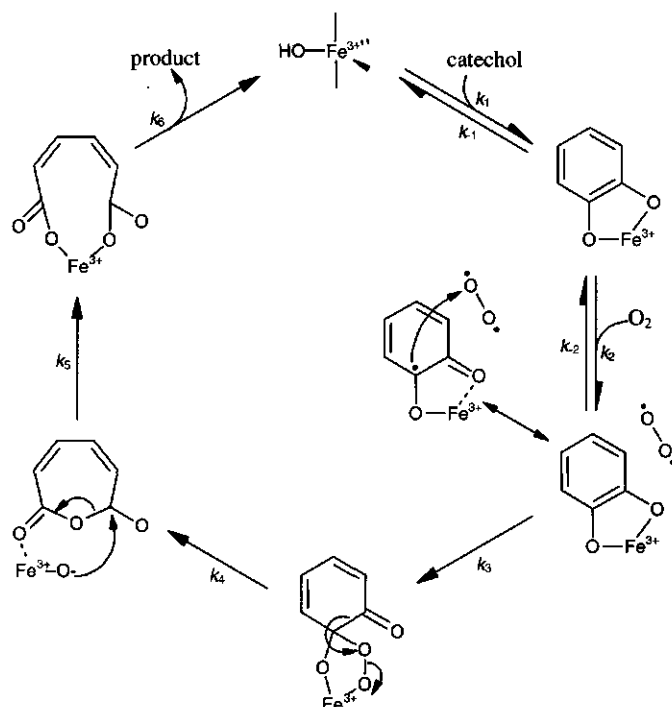


Figure 4.1 The general reaction scheme for intradiol catechol dioxygenases (Lipscomb and Orville, 1992, Que and Ho, 1996, Walsh *et al.*, 1983). The five ligands of the iron(III) centre in the resting enzyme are a solvent molecule, two tyrosine residues and two histidine residues. The solvent molecule and one of the tyrosine ligands are replaced by the hydroxyl groups of the catechol substrate upon bidentate binding.

4.1 Introduction

Catechol dioxygenases are essential in microbial degradation of aromatic compounds, since they catalyse the oxidative cleavage of *ortho*-dihydroxy substituted aromatic rings while incorporating both atoms of molecular oxygen. The catechol-1,2-dioxygenases convert catechols to *cis,cis*-muconates which can be further degraded to central carbon metabolites. Substrate specificity is an interesting characteristic of the catechol dioxygenases. Aromatic xenobiotics encountered in the environment often carry substituents that may not be removed during degradation steps prior to ring cleavage, consequently influencing the essential dioxygenase activity. Besides, the influence of substituents on the rate of conversion of the substrate could provide information on the catalytic mechanism of the enzyme.

A multistep reaction cycle (Figure 4.1), proceeding via a nucleophilic attack of the iron(III)-bound deprotonated substrate on molecular oxygen, has been proposed for the intradiol cleavage of *ortho*-dihydroxybenzenes (Lipscomb and Orville, 1992, Que and

Ho, 1996) by the most often studied *ortho*-cleaving enzymes, i.e. protocatechuate-3,4-dioxygenase from *Pseudomonas aeruginosa* (Fujisawa and Hayaishi, 1968, Que *et al.*, 1976) and catechol-1,2-dioxygenase from *Pseudomonas putida* (arvilla) C1 (Kojima *et al.*, 1967). Walsh *et al.* (Walsh *et al.*, 1983) have performed transient kinetic studies on catechol-1,2-dioxygenase identifying an enzyme-substrate complex and two intermediates during the subsequent reaction with oxygen. The catalytic constants obtained indicate that the last step of the reaction cycle, possibly product release, is rate limiting at 4°C. No such kinetic data are available for the reaction at higher temperatures.

In the present study the effect of varying C4/C5 substituents in catechol derivatives on their overall rates of conversion by a catechol-1,2-dioxygenase was investigated at 20°C. Comparison of the rates obtained with the outcomes of molecular orbital calculations on the substrates may reveal whether differences observed must be ascribed to sterical hindrance (preventing efficient conversion of substrates with relatively large C4/C5 substituents) or to electronic reactivity. Previous studies for other enzyme systems have demonstrated that it is indeed possible to relate experimental rates of conversion quantitatively to calculated frontier orbital energies of substrates characterising their reactivity. For example, such quantitative structure activity relationships (QSAR) were reported for *p*-hydroxybenzoate hydroxylase (Vervoort *et al.*, 1992), cytochromes *P*-450 (Cnubben *et al.*, 1994), phenol hydroxylase (Peelen *et al.*, 1995), horseradish peroxidase (Sakurada *et al.*, 1990, Van Haandel *et al.*, 1996) and glutathione-*S*-transferases (Rietjens *et al.*, 1995, Soffers *et al.*, 1996). As demonstrated in these previous studies, the existence of a frontier-orbital-based QSAR can provide additional insight into the reaction mechanism and the possible nature of the rate-limiting step.

4.2 Materials and methods

Chemicals

Catechol was obtained from Sigma, 4-methylcatechol and 4-fluorophenol were obtained from Aldrich and 4-chlorophenol, 4-bromophenol, 3,4-difluorophenol and 4-chloro-3-fluorophenol were obtained from Janssen Chimica. 4-Fluorocatechol, 4-chlorocatechol, 4-bromocatechol, 4,5-difluorocatechol and 4-chloro-5-fluorocatechol were prepared by conversion of the corresponding substituted phenols by phenol hydroxylase from *Trichosporon cutaneum*, purified as described below. The incubation mixtures contained 5 mM of phenol derivative, 50 mM potassium phosphate, pH 7.6, 12 mM ascorbic acid, to prevent auto-oxidation of the catechol formed, 0.5 mM FAD and 6 mM NADPH. The reaction was started, at room

temperature, with varying amounts (depending on the reactivity of the phenol) of phenol hydroxylase. The mixture was stirred and the activity was monitored by measuring the rate of oxygen consumption using a Clark electrode. The incubations were stopped after conversion of at least 60% of the phenol, as estimated from the measured rate of conversion and the period of incubation. The catechols thus obtained were concentrated by freeze/drying and purified by HPLC using a Microsphere C18 column. An eluent gradient was applied from 100% H₂O to 20% H₂O / 80% methanol in 14 min. Detection was at 280 nm. Fractions containing the catechols were collected from the HPLC, freeze/dried and stored at -20°C.

Enzyme purification.

Phenol hydroxylase was purified from the yeast *Trichosporon cutaneum* essentially as described by Neujahr and Gaal (Neujahr and Gaal, 1973). The final preparation had a specific activity of 5.0 U/mg protein (over 90% pure) and was free of catechol-1,2-dioxygenase activity. Catechol-1,2-dioxygenase from *Pseudomonas putida* (arvilla) C1 was purified essentially as described by Nakai (1988). The specific activity was 38 U/mg at room temperature and air-saturated oxygen concentration (260 µM), which is comparable to values obtained by others (38.4 U/mg) (Fujiwara *et al.*, 1975).

Activity measurements

The activity of catechol-1,2-dioxygenase with the different catechol substrates was measured by following the oxygen consumption using a Clark electrode. The assays were carried out at 20°C in 2 ml 100 mM potassium phosphate, pH 7.6. Buffer and substrate were mixed in the reaction vessel, which was then closed and stirred until a constant oxygen level was reached. The reaction was started by injection of enzyme and the enzyme activity was derived from the initial rate of oxygen decrease detected. The K_m values for catechol are in the order of micromoles and too low to be measured. Instead, all measurements were carried out under saturating catechol concentrations. This was confirmed by verifying that the activity was not affected by a twofold reduction of the catechol concentration. The concentration of the catechol substrate in the assays varied between 10 µM and 300 µM (Table 4.1). The actual concentrations of the synthesised catechols were determined from assays in which the catechol concentration is smaller than the oxygen concentration by measuring the total oxygen consumption upon complete conversion of the catechol. The K_m values for oxygen, using the various substrates, were determined in a series of activity measurements at different concentration of O₂. The oxygen concentration was lowered by bubbling N₂ through the reaction mixture during different time intervals. The actual oxygen concentration was measured using the Clark electrode. The data were analysed by

fitting the values to the Michaelis-Menten equation using Kaleidagraph version 2.1.3. The enzyme activity k_{cat} was calculated on the basis of an average molecular mass of catechol-1,2-dioxygenase of 62 kDa (Nakai *et al.*, 1990).

¹⁹F-NMR measurements

¹⁹F-NMR measurements were performed on a Bruker AMX 300 spectrometer as described previously (Vervoort, 1992). The sample volume was 1.6 ml and for calibration and locking the magnetic field a coaxial insert was used which contained a known amount of *p*-fluorobenzoic acid as internal standard in D₂O.

Molecular orbital calculations

A nucleophilic attack of the iron(III)-bound catecholate on molecular oxygen is an essential step in the catechol-1,2-dioxygenase reaction (Que and Ho, 1996). According to the frontier orbital theory the reactivity of soft nucleophiles depends on the energy of the highest occupied molecular orbital (E_{HOMO}), i.e. the energy of the most reactive π electrons in the aromatic system (Fleming, 1976). Therefore, the HOMO energies of the catechol derivatives are calculated and compared with experimental k_{cat} values. For calculations on the catechol derivatives, the semiempirical AM1 method and the closed-shell Hartree-Fock *ab initio* method with the 3-21G* and 6-31+G** basis sets were applied. To study the effect of bidentate coordination of the substrate to iron in the active site a calculation was performed on a complete iron coordination site using a pBP-86 density functional method. All molecular orbital calculations were performed with Spartan 5.0 or, in case of the HF/6-31+G** calculations, with Gaussian94 (Frisch *et al.*, 1995).

Solvation effects and other influences due to binding of the catechols in the active site of catechol-1,2-dioxygenase are assumed to have no significant influence on the relative differences between electronic parameters of the homologous catechols. The outcomes of the in gas-phase calculations can thus be used to study relative differences within a series of related compounds when bound to the active site of the enzyme (Cnubben *et al.*, 1994, Peelen *et al.*, 1995, Rietjens *et al.* 1995, Sakurada *et al.*, 1990, Soffers *et al.*, 1996, Van Haandel *et al.*, 1996, Vervoort *et al.*, 1992).

4.3 Results

Biosynthesis of the substituted catechols

Phenol hydroxylase from *Trichosporon cutaneum* catalyses a regioselective hydroxylation at the *ortho* position with respect to the hydroxyl moiety of the phenol (Neujahr and Gaal, 1973, Neujahr and Kjellén, 1978). Therefore, all 4-substituted

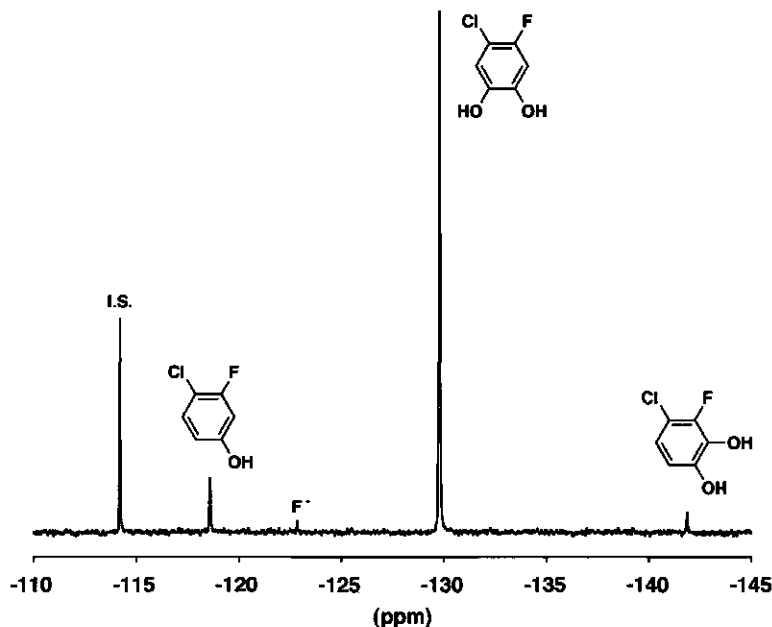


Figure 4.2 ^{19}F -NMR spectrum of the incubation of 4-chloro-3-fluorophenol with phenol hydroxylase from *Trichosporon cutaneum*, showing formation of 4-chloro-5-fluorocatechol as the major (94%) catechol metabolite. Metabolites were identified according to Peelen *et al.* (1995). The resonance marked I.S. is from the internal standard *p*-fluorobenzoic acid.

phenols are converted selectively to the corresponding 4-substituted catechols. 3,4-Difluorophenol and 4-chloro-3-fluorophenol can, in principle, be hydroxylated at both ortho positions yielding different products. However, as previously described (Peelen *et al.*, 1995), the 4,5-difluorocatechol and 4-chloro-5-fluorocatechol are formed preferentially (93% and 94% respectively). This is illustrated by the ^{19}F -NMR spectrum of the incubation mixture of 4-chloro-3-fluorophenol (Figure 4.2). The non-substituted catechol was also prepared from phenol, using phenol hydroxylase. The activity obtained with this biosynthetically produced parent catechol was equal to the activity with the commercially obtained substrate (data not shown), indicating that the biosynthetic catechol preparation does not contain inhibiting compounds. Characterisation of the catechols by HPLC analysis with diode-array detection showed that the ultraviolet-visible absorbance spectra of the substituted catechols are, apart from small shifts in λ_{max} (< 10 nm), similar to the absorbance spectrum of the non-substituted catechol. The differences in retention times (t_r) on the hydrophobic column used, were in accordance with the expected differences in hydrophobicity ($t_{r,H} < t_{r,F} < t_{r,Cl} < t_{r,Br}$). The purity of the catechols prepared by the biosynthetic phenol

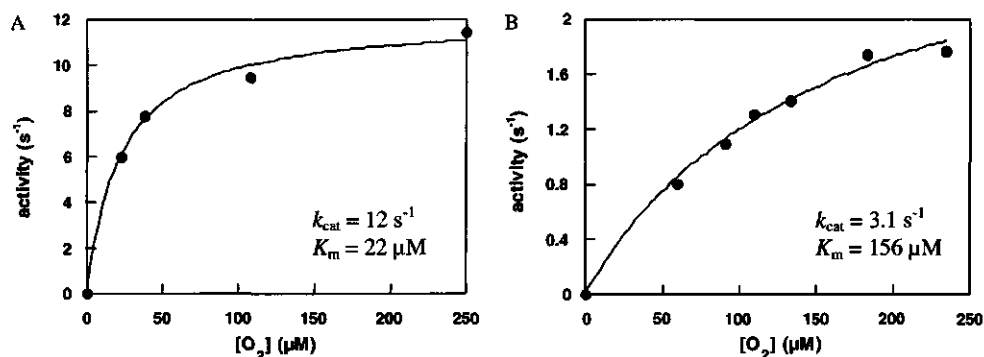


Figure 4.3 The activities obtained with a) 4-fluorocatechol and b) 4-chlorocatechol, plotted against the oxygen concentration. The data were fitted using the Michaelis-Menten equation.

hydroxylase method was at least 96%, based on HPLC analysis, and, for the fluorine-containing catechols, based on ¹⁹F-NMR analysis (data not shown).

Conversion of the catechol derivatives by catechol-1,2-dioxygenase from Pseudomonas putida (arvilla) C1

K_m values for oxygen and k_{cat} values for conversion of the catechol derivatives at saturating catechol and oxygen concentrations were determined (Table 4.1). This was done by measuring the rate of conversion at different oxygen concentrations and saturating catechol concentrations and applying Michaelis-Menten analysis. As an example, Figure 4.3 presents the Michaelis-Menten curves obtained for 4-fluorocatechol and 4-chlorocatechol. The k_{cat} values obtained decreased in the order catechol > 4-methylcatechol > 4-fluorocatechol > 4-chlorocatechol > 4-bromocatechol > 4,5-difluorocatechol > 4-chloro-5-fluorocatechol.

After the complete reaction with a substituted catechol was measured, a high concentration of non-substituted catechol (at least five times the concentration of the converted substituted catechol) was added and the resulting activity was measured. For all assays containing substituted catechols, the addition of the parent catechol resulted in an activity that equalled the activity obtained with catechol as the only substrate (data not shown). This indicates that differences in the rate of conversion are not due to irreversible inhibition, i.e. to the formation of an abortive enzyme-substrate complex or to product inhibition.

Rate constant k_2 for oxygen binding derived on the basis of k_{cat} and K_m

The ratio $k_{cat}/K_m(\text{O}_2)$ gives information about the process of oxygen binding. Figure 4.1 presents a general reaction scheme for the catalytic cycle of an intradiol cleaving

Table 4.1 k_{cat} values at saturating catechol and O_2 concentrations (mean of two experiments with independent enzyme preparations) for the conversion of catechol derivatives by catechol-1,2-dioxygenase from *Pseudomonas putida* (arvilla). The k_{cat} at saturating O_2 concentration with 4-chloro-5-fluorocatechol is given as a range, indicating possible values, based on the measured activity at 260 μM O_2 and the highest and the lowest K_m values obtained with other substrates. The K_m for 4-chloro-5-fluorocatechol could not be determined because the activity is too low.

substrate	catechol concentration (μM)	K_m value for oxygen (mM)	k_{cat} at saturating catechol and O_2 concentration (s^{-1})	k_{cat} relative to reaction with parent catechol (%)
catechol	300	0.11 ± 0.01	57 ± 3	100 ± 5
4-methylcatechol	300	0.09 ± 0.01	36 ± 2	62 ± 4
4-fluorocatechol	10	0.02 ± 0.003	12 ± 1	21 ± 1
4-chlorocatechol	50	0.17 ± 0.03	3.4 ± 0.3	6.0 ± 0.6
4-bromocatechol	10	0.30 ± 0.09	1.3 ± 0.2	2.2 ± 0.4
4,5-difluorocatechol	10	0.02 ± 0.01	0.63 ± 0.06	1.1 ± 0.1
4-chloro-5-fluorocatechol	300		0.40 ± 0.17	0.7 ± 0.3

dioxygenase (Lipscomb and Orville, 1992, Que and Ho, 1996, Walsh *et al.*, 1983). It can be derived by using the King-Altman method (King and Altman, 1956) that under substrate saturating conditions the following relation exists.

$$v = \frac{\frac{k_3 k_4 k_5 k_6}{k_3 k_4 k_5 + k_3 k_4 k_6 + k_3 k_5 k_6 + k_4 k_5 k_6} [E]}{1 + \frac{(k_{-2} + k_3) k_4 k_5 k_6}{(k_3 k_4 k_5 + k_3 k_4 k_6 + k_3 k_5 k_6 + k_4 k_5 k_6) k_2} \frac{1}{[O_2]}} \quad (1)$$

Assuming that $k_3 \gg k_{-2}$, which implies that the reaction with O_2 is much faster than the rate of O_2 dissociation, equation (1) changes to:

$$v = \frac{\frac{k_3 k_4 k_5 k_6}{k_3 k_4 k_5 + k_3 k_4 k_6 + k_3 k_5 k_6 + k_4 k_5 k_6} [E]}{1 + \frac{k_3 k_4 k_5 k_6}{(k_3 k_4 k_5 + k_3 k_4 k_6 + k_3 k_5 k_6 + k_4 k_5 k_6) k_2} \frac{1}{[O_2]}} \quad (2)$$

Equation (2) can be further simplified by introducing the catalytic constant $k_{cat} = k_3 k_4 k_5 k_6 / (k_3 k_4 k_5 + k_3 k_4 k_6 + k_3 k_5 k_6 + k_4 k_5 k_6)$, which results in

$$v = \frac{\frac{k_{cat} [E]}{1 + \frac{k_{cat}}{k_2} \frac{1}{[O_2]}}}{1 + \frac{K_m(O_2)}{[O_2]}} \quad (3)$$

From this it follows that the ratio $k_{cat}/K_m(O_2)$ equals k_2 . Table 4.2 presents the values for k_2 derived in this way on the basis of the k_{cat} and $K_m(O_2)$ values of the present study (Table 4.1). The differences in the values for k_2 indicate that the substituents on the aromatic ring of the substrate affect the binding of molecular oxygen in the active site.

Table 4.2 Rate constants k_2 representing binding of O_2 , calculated from Table 4.1 as $k_{cat} / K_m(O_2)$, for the various catechol derivatives in the present study, compared to values derived for the catechol-1,2-dioxygenases I and II of *Pseudomonas* sp. B13, on the basis of data taken from Dorn and Knackmuss (1978). Van der Waals radii of the substituents, are given as a possible parameter influencing the rate of oxygen binding. It should be noted that k_{cat} , as defined in equation (3), is independent of k_2 presented here.

substrate	$k_2 \times 10^5$ ($s^{-1} M^{-1}$)	relative values (%)	relative values (%) for k_2 for catechol-1,2- dioxygenase I and II		Van der Waals radius of C4 substituent (Å) (Chang, 1981)
catechol	5.2 ± 0.7	100 ± 14	100	100	1.1
4-methylcatechol	4.0 ± 0.7	77 ± 13	96	652	2.00
4-fluorocatechol	6.0 ± 1.4	115 ± 27	-	223	1.35
4-chlorocatechol	0.20 ± 0.05	3.8 ± 1.0	4.2	123	1.80
4-bromocatechol	0.04 ± 0.02	0.8 ± 0.4	-	-	1.95
4,5-difluorocatechol	0.3 ± 0.2	6.2 ± 3.7	-	-	1.35

Relative k_{cat} and $K_m(O_2)$ values have been reported for two different catechol-1,2-dioxygenases, I and II, from *Pseudomonas* sp. B13 by Dorn and Knackmuss (Dorn and Knackmuss, 1978) with the substrates catechol, 4-fluorocatechol (catechol-1,2-dioxygenase II only), 4-chlorocatechol and 4-methylcatechol. Based on these literature data, relative values for k_2 were calculated as described above (Table 4.2). The relative k_2 values for catechol-1,2-dioxygenase I are similar to those obtained for catechol-1,2-dioxygenase of *Pseudomonas putida* (arvilla), whereas the k_2 values thus obtained for catechol-1,2-dioxygenase II differ significantly. This suggests that catechol-1,2-dioxygenase of *Pseudomonas putida* (arvilla) and catechol-1,2-dioxygenase I of *Pseudomonas* sp. B13 differ from catechol-1,2-dioxygenase II of *Pseudomonas* sp. B13 with respect to the way molecular oxygen binds to the active site.

Charge delocalization upon coordination of substrate to iron in the active site

Bidentate coordination of the substrate to the iron in the active site of catechol-1,2-dioxygenase influences the electronic characteristics of the catechol. The coordination could result in the partial delocalization of negative charge from the oxygen atoms of the catecholates to the iron atom. Since electronic charge significantly influences the reactivity of the π electrons (i.e. the E_{HOMO}) it was important to define the charge on the substrate that should be taken into account when performing calculations on the isolated substrates. Therefore, an initial quantum mechanical molecular orbital calculation was performed on a model for the complete iron coordination site which is presented in Figure 4.4. This model was built on the basis of the crystal structure of a homologous enzyme, protocatechuate dioxygenase in complex with the substrate protocatechuate (Orville *et al.*, 1997) (Figure 4.4). The pBP-86 density functional

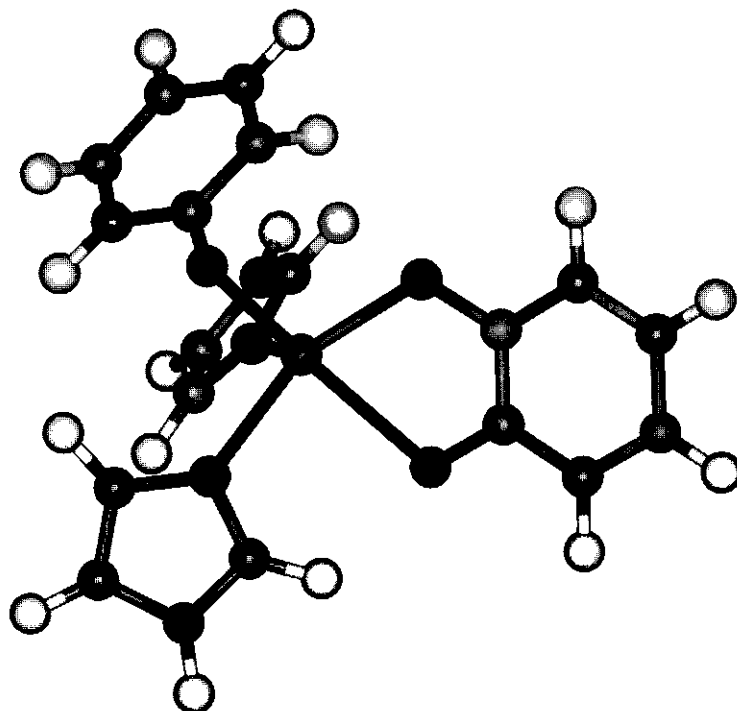


Figure 4.4 A model for the iron-containing active site in catechol 1,2-dioxygenase built on the basis of the crystal structure of the homologous enzyme protocatechuate dioxygenase in complex with the substrate protocatechuate (3PCA) (Orville *et al.*, 1997). The model consists of an iron atom, coordinated by a catecholate, a phenolate and two imidazole rings. The non-hydrogen atoms in this model were fixed to the crystal coordinates of the protocatechuate, Tyr408 and His460 and His462, respectively. Hydrogen atoms were added and optimised using the Merck molecular mechanical geometry optimiser in the builder routine in Spartan 5.0.

method as implemented in Spartan 5.0 was applied using the DN* basis set to obtain a self-consistent field. Mulliken analysis at the end of this calculation indicated a total charge of $-0.8 e$ on the catecholate atoms. On the basis of this result, subsequent molecular orbital calculations were performed on the series of catechols in their mono-anionic form in which one of the two hydroxyl moieties is deprotonated.

It should be noted that the iron-ligand complex model of the present study was designed only to give a reasonable estimate of the total charge on the catecholate. The geometry of the complex was derived from a crystal structure of a homologous enzyme. Small geometrical differences in active site coordination between the model and our intradiol dioxygenase may influence the orbital distribution significantly. Therefore we did not attempt to use the complete iron-ligand complex model for calculations to compare with experimental data. Instead, we assumed that coordination

Table 4.3 Energies of the highest occupied molecular orbitals (eV) of the various substrates, obtained with the AM1 semiempirical method and the *ab initio* method using the 3-21G* and 6-31+G** basis sets.

substrate	E _{HOMO} (eV) obtained with		
	AM1	3-21G*	6-31+G**
catechol	-2.7460	-1.8877	-2.4186
4-methylcatechol	-2.7615	-1.8900	-2.3879
4-fluorocatechol	-3.0015	-2.2178	-2.6820
4-chlorocatechol	-3.0990	-2.4047	-2.8449
4-bromocatechol	-3.1760	-2.3917	-2.9144
4,5-difluorocatechol	-3.2440	-2.5070	-2.9171
4-chloro-5-fluorocatechol	-3.3190	-2.6702	-3.0306
correlation with $\ln(k_{\text{cat}})$ (<i>R</i>)	0.987	0.971	0.972

of the catecholates to the active site of catechol-1,2-dioxygenase influences their nucleophilicity in a systematic way, thereby not changing the order in reactivity of the catecholates used.

Calculation of chemical reactivity of the mono-anionic catechol derivatives

Based on the reaction mechanism proposed in the literature (Lipscomb and Orville, 1992, Que and Ho, 1996), the nucleophilic reactivity of the catecholate was expected to be an essential characteristic for its conversion by catechol-1,2-dioxygenase. Therefore, the HOMO energies of the substrates were calculated for the various catecholate mono-anions using the semiempirical AM1 method and the *ab initio* Hartree-Fock method with the 3-21G(*) and 6-31+G** basis sets. The results are presented in Table 4.3. For the asymmetric catechol derivatives, for which two isomeric forms of the mono-anions exist, the average of the HOMO energies of both isomers were calculated. The differences in the values obtained for the various catechol derivatives indicate that the substituents influence the nucleophilic reactivity of the substrate.

Molecular-orbital-based QSAR for the conversion of C4/C5-substituted catechols by catechol-1,2-dioxygenase from Pseudomonas putida (arvilla) C1

In Figure 4.5 the natural logarithm of the k_{cat} values for the conversion of the catechol derivatives by catechol-1,2-dioxygenase from *Pseudomonas putida* (arvilla) C1 are plotted against the energies of the reactive HOMO electrons in the aromatic ring of the catecholate mono-anions as calculated using the semiempirical AM1 method and the *ab initio* method with 3-21G and 6-31+G** basis sets. Linear correlations are obtained with $R = 0.99$, 0.97 and 0.97 respectively (Figure 4.5).

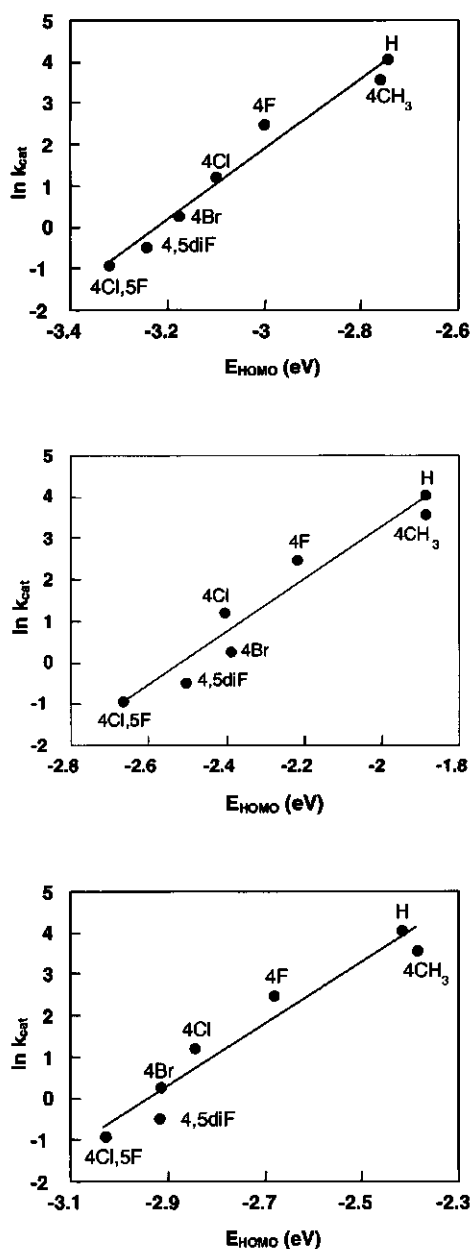


Figure 4.5 Plot of the natural logarithm of the k_{cat} for the conversion of a series of catechol derivatives by catechol-1,2-dioxygenase from *Pseudomonas putida* (arvilla) C1 plotted against the energy of the reactive HOMO electrons in the aromatic ring of the catechols as calculated using a) the semiempirical AM1 method and the *ab initio* Hartree-Fock method with the b) 3-21G(*) and c) 6-31+G** basis sets.

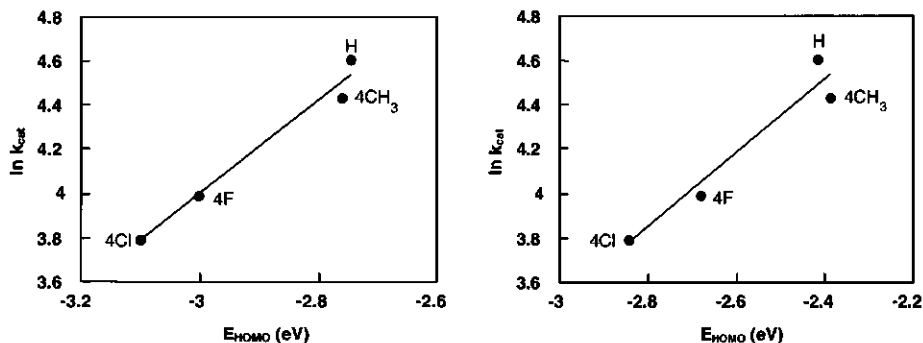


Figure 4.6 Plot of the natural logarithm of the relative k_{cat} values reported by Dorn and Knackmuss (1978) obtained with different catechol derivatives for catechol-1,2-dioxygenase II from *Pseudomonas* sp. B13 against the energy of the reactive HOMO electrons in the aromatic ring of the catechols as calculated using a) the semiempirical AM1 method and b) the *ab initio* Hartree-Fock method with a 6-31+G** basis set.

In addition, the k_{cat} values for the catechol-1,2-dioxygenase II, from *Pseudomonas* sp. B13 with four C4-substituted substrates (Dorn and Knackmuss, 1978) were correlated in the same way with the E_{HOMO} values obtained for the respective catecholate monoanions on the basis of the semiempirical AM1 and *ab initio* 6-31+G** calculations. The correlations obtained (Figure 4.6) suggests a similar mechanism for the type II catechol-1,2-dioxygenase with respect to the rate-limiting step.

From the definition of k_{cat} introduced in equation (3) it can be derived that if one of the rate constants k_3 , k_4 , k_5 or k_6 ($=k_x$) is much smaller than the other rate constants, this rate constant k_x rate will determine the overall rate (i.e. $k_{cat} = k_x$). To be able to discuss which step in the proposed reaction cycle could be the rate-determining step underlying the QSAR obtained, additional calculations were performed on the intermediate structures proposed in the multistep reaction cycle (Figure 4.1). The difference in heat of formation (ΔH_f), going from one intermediate to the next, can be assumed to correlate with the activation barrier of this reaction step, as long as a series of identical reactions, i.e. the conversion of homologous substrates, is compared. Differences in heat of formation, corresponding to step 3, 4 and 5 of the proposed reaction cycle (Figure 4.7) were calculated for the various substrates, using the AM1 method. The values obtained, relative to the value obtained with the parent catechol and expressed as $\Delta\Delta H_f$, were used as parameter for comparison to the natural logarithm of the k_{cat} values obtained in the present study (Figure 4.7). Reasonable correlations were thus obtained for step 3 and 4 of the reaction cycle (Figure 4.1), with $R = 0.99$ and 0.97 respectively, while the $\Delta\Delta H_f$ -values obtained for step 5 do not correlate with the natural logarithm of the rate of conversion ($R = 0.68$).

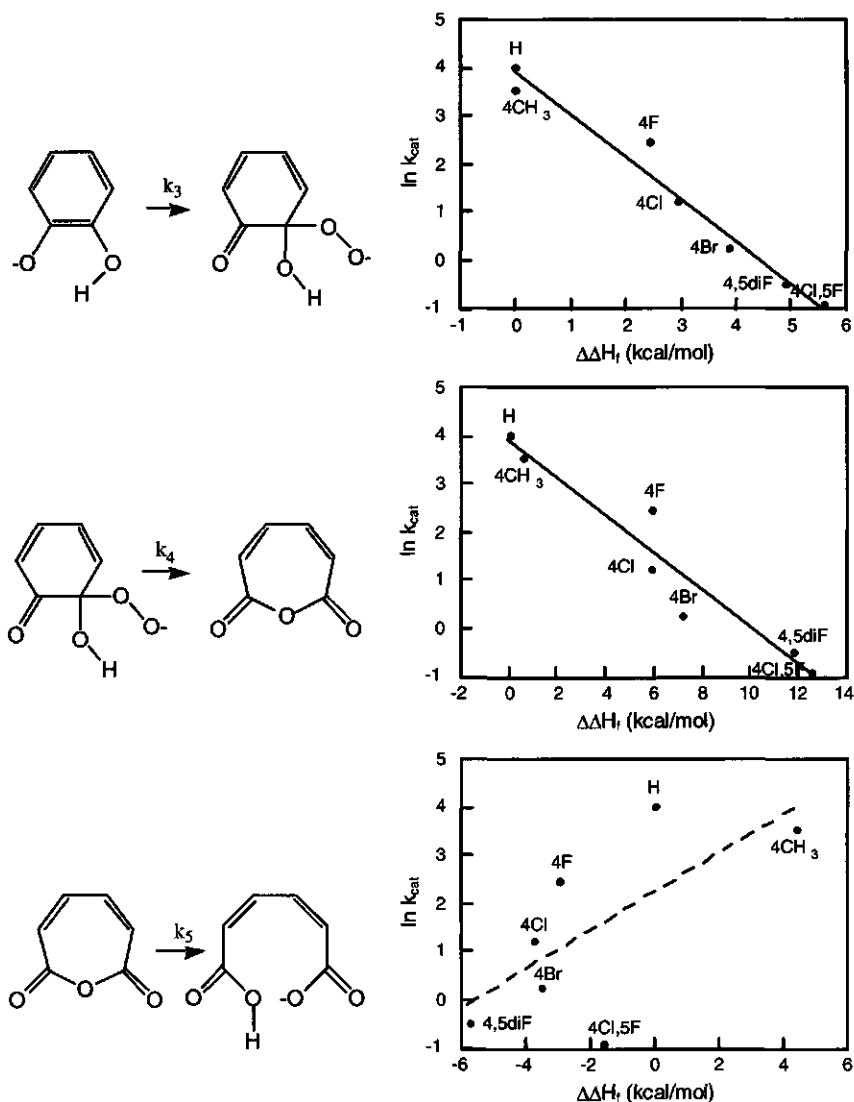


Figure 4.7 Plot of the natural logarithm of the k_{cat} values obtained with different catechol derivatives against $\Delta\Delta H_f$ values for step 3, 4 and 5 of the proposed reaction cycle (Figure 4.3), obtained from semiempirical molecular orbital calculations (AM1 method).

4.4 Discussion

In the present study the effect of substituents on the conversion of catechol derivatives by catechol-1,2-dioxygenase was investigated. This reaction is important in the degradation of aromatic pollutants. The results indicate the importance of substrate nucleophilicity for the intradiol dioxygenase activity.

Only C4-substituted and C5-substituted derivatives were used. C3-substituted substrates were excluded because catechol-1,2-dioxygenase is known to convert some of these substrates, i.e. 3-methylcatechol and 3-methoxycatechol, via extradiol cleavage in addition to the normal intradiol cleavage (Fujiwara *et al.*, 1975). Furthermore, C3-substituted catechols have the complicating possibility of the formation of abortive enzyme-substrate complexes. Walsh and Ballou (1983) proposed the formation of such abortive enzyme-substrate complexes, in the case of protocatechuate-3,4-dioxygenase when the protocatechuate derivatives are halogenated at a position *ortho* to the hydroxyl moieties. According to their kinetic scheme these complexes are inactive or converted very slowly to the normal type of enzyme-substrate complex. The conversion of the inactive type into the normal type of complex is rate limiting in this case and therefore the overall activity cannot be expected to correlate with the calculated nucleophilic reactivity of the catechol substrates, which was an objective of the present study. Dorn and Knackmuss (1978) also suggested the formation of an abortive complex, in case of catechol-1,2-dioxygenase with 3-chlorocatechol as the substrate, indicating that the phenomenon observed for protocatechuate-3,4-dioxygenase occurs with catechol-1,2-dioxygenases as well.

Using catechol, 4-methylcatechol, 4-fluorocatechol, 4-chlorocatechol, 4-bromocatechol, 4,5-difluorocatechol and 4-chloro-5-fluorocatechol as model substrates, k_{cat} values at saturating catechol and O_2 concentration as well as K_{m} values for O_2 were obtained. The rate constant k_2 , derived from k_{cat} and $K_{\text{m}}(\text{O}_2)$, represents the rate constant for oxygen binding to the enzyme-substrate complex (Figure 4.3). To explain the differences in the values obtained for k_2 , the size of the substituents (Table 4.2) may be a factor of influence. Larger halogen substituents result in lower k_2 values (compare 4-fluorocatechol, 4-chlorocatechol and 4-bromocatechol), which could be interpreted as sterical hindrance by the substituents, lowering the accessibility of the active site for oxygen. However, the differences observed in the k_2 values cannot be explained by sterical hindrance alone. For example, 4-methylcatechol and 4-bromocatechol, with substituents that are similar in size, have completely different k_2 values. Similarly the k_2 values obtained with 4-fluorocatechol and 4,5-difluorocatechol differ significantly from each other, which cannot be explained by sterical hindrance, because the Van der Waals radius of a fluorine is only slightly larger than the radius of a hydrogen substituent. Apparently another, as yet unidentified, factor influences the rate of oxygen binding as well. For four substrates, catechol, 4-methylcatechol, 4-fluorocatechol and 4-chlorocatechol, the relative k_2 values could be compared with relative k_2 values obtained with catechol-1,2-dioxygenases I and II from *Pseudomonas sp.* B13. The values obtained for catechol-

1,2-dioxygenase I are similar to the values obtained in the present study for catechol-1,2-dioxygenase from *Pseudomonas putida* (arvilla), while the results for catechol-1,2-dioxygenase II differ significantly. This suggests that the accessibility of the active site for oxygen, or the way oxygen binds in the active site differs for the catechol-1,2-dioxygenase I and II. In this respect, the catechol-1,2-dioxygenase from *Pseudomonas putida* (arvilla), presently studied, is more similar to the catechol-1,2-dioxygenase I, than to the catechol-1,2-dioxygenase II from *Pseudomonas* sp. B13. It is important to notice that the assumption that $k_2 \ll k_3$, used to simplify the kinetic formula, may not be totally correct. It cannot be excluded that in case of substrates that are converted relatively slow k_3 is too small to justify the neglect of k_2 in equation (1). Nevertheless, the ratio k_{cat}/K_m can be used as a parameter to detect the influence of the substrate on oxygen binding and to compare different enzymes in this respect.

To explain the variations in the overall k_{cat} values for conversion of different catechol derivatives by catechol-1,2-dioxygenase, the size of the substituents and the nucleophilic reactivity of the substrates are considered as parameters of influence. The differences in Van der Waals radii of the substituents, which may influence the overall rate of conversion through sterical effects, cannot explain the differences in k_{cat} values. For example, 4-methylcatechol is converted much faster than 4-bromocatechol, while the methyl substituent is similar in size compared to the bromo substituent. Similarly, the k_{cat} values for catechol, 4-fluorocatechol and 4,5-difluorocatechol differ significantly, in spite of the similarity in size of the substituents.

Instead, results of the present study show that the k_{cat} values correlate quantitatively to a calculated parameter for nucleophilic reactivity. The correlation is based on the frontier orbital theory, which explains that, during the reaction between a soft nucleophile and a soft electrophile, the interaction between the frontier orbitals, the highest occupied molecular orbital (HOMO) of the nucleophile and the lowest unoccupied molecular orbital (LUMO) of the electrophile, is essential and lowers the activation barrier (Fleming, 1976). The energy difference between these frontier orbitals is the most important factor leading to the differences in the activation energies of the reactions of a series of related compounds. Thus, in case of a series of related nucleophiles reacting with a given electrophile, the HOMO energy of the nucleophiles is the essential parameter influencing the activation energies. Based on the finding that the nucleophilic attack of the aromatic π electrons of the iron(III)-bound catecholate on the oxygen molecule is an important step in the proposed reaction mechanism (Lipscomb and Orville, 1992, Que and Ho, 1996), the HOMO of the catecholate was assumed to be the relevant frontier orbital in the present study. Thus, the natural logarithm of the k_{cat} , which is proportional to the activation energy of the rate-limiting step based on the Arrhenius equation, is correlated to the E_{HOMO} .

The correlations obtained in the present study, with r varying from 0.971 to 0.986 depending on the calculation method used (Table 4.3), indicate that the nucleophilic reactivity of the substrate is the important characteristic, contributing to the origin of the differences in rates of conversion of different catechol derivatives by catechol-1,2-dioxygenase. It suggests that the mechanism of the rate-limiting step of the reaction cycle is dominated by the nucleophilic reactivity of the substrate. Similar correlations observed for the type II catechol-1,2-dioxygenase of *Pseudomonas sp. B13*, indicate that, although the two types of catechol-1,2-dioxygenase differ significantly with respect to oxygen binding and substrate specificity (Dorn and Knackmuss, 1978), the rate-limiting step for both enzymes depends on the nucleophilic reactivity of the substrate. Dorn and Knackmuss concluded on the basis of correlations with substituent constants that in the case of catechol-1,2-dioxygenase II the rate-limiting step is dependent on the nucleophilic reactivity of the substrate, in agreement with the conclusion of the present study, while in case of catechol-1,2-dioxygenase I steric influences of the substituent play a role in the rate limiting step as well. This latter conclusion is not confirmed by the present study as far as 4-substituted substrate derivatives are concerned.

The present results give no definite answer to the question which step in the proposed multistep reaction cycle is rate determining and responsible for the effects of the substituents on the overall rate of conversion. From equations (2) and (3) it follows that four steps in the reaction cycle, step 3, 4, 5 and 6, are to be considered (Figure 4.3). Obviously, the attack of the substrate on molecular oxygen, step 3 of the reaction cycle, is dependent on the nucleophilicity of the substrate and may possibly determine the overall rate of conversion. This suggestion is supported by the observed correlation of the overall rate constant with the calculated values for the $\Delta\Delta H_f$ of this reaction step (Figure 4.7a, $R = 0.99$). However, it can not be excluded beforehand that the subsequent reaction steps are also influenced by the nucleophilic properties of the substrate. The correlation obtained with the parameter $\Delta\Delta H_f$ of reaction step 4 (Figure 4.7b, $R = 0.97$) suggests that this step might as well be the rate-limiting step. For step 5 such a correlation is not observed (Figure 4.7c, $R = 0.68$), indicating that this step is not likely to be rate limiting. Product release, step 6, is not expected to be influenced positively by an increasing nucleophilicity of the reacting compound. Instead, an increased ability to accept electrons, related to electrophilicity rather than nucleophilicity, would be expected to enhance product release. Thus, the present results indicate that step 3 or 4 are rate limiting at 20°C. From comparison with Walsh *et al.* (1983) it appears that an increase in temperature from 4°C to 20°C changes the rate-limiting step from product release to a catalytic step in the reaction cycle.

Acknowledgements

This work was supported by an EU Copernicus grant no. ERBIC15CT960103 and by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organisation for Scientific Research (NWO). The contract N 96.01270.PF37 of the *Gruppo per la Difesa dai Rischi Industriali* of CNR is also gratefully acknowledged.

5

Correlation of calculated activation energies with experimental rate constants for an enzyme catalysed aromatic hydroxylation

Lars Ridder, Adrian J. Mulholland, Jacques Vervoort and Ivonne M.C.M. Rietjens
The Journal of the American Chemical Society 120, 7641-7642 (1998)

The reaction pathways for aromatic 3-hydroxylation of *p*-hydroxybenzoate and four fluorinated analogs by the reactive C4a-hydroperoxyflavin cofactor intermediate in *p*-hydroxybenzoate hydroxylase (PHBH) have been investigated by a combined quantum mechanical and molecular mechanical (QM/MM) method. A structural model for the C4a-hydroperoxyflavin intermediate state in the PHBH reaction cycle was built on the basis of the crystal structure coordinates of the enzyme-substrate complex.

A reaction pathway for the subsequent hydroxylation step, i.e. the electrophilic attack of the distal oxygen of the cofactor on the substrate, was calculated by imposing a reaction coordinate which involves cleavage of the peroxide oxygen-oxygen bond and formation of the carbon-oxygen bond between the substrate C3 and the distal oxygen of the peroxide moiety of the cofactor. The geometrical changes and the Mulliken charge distributions along the calculated reaction pathway are in line with an electrophilic aromatic substitution type of mechanism.

The energy profiles obtained with the various substrates reveal an influence of the fluorine substituents on the calculated energy barrier. Comparison of the results to experimental data yields a clear correlation ($R = 0.96$) between the natural logarithm of the k_{cat} for the enzymatic conversion of the various substrates and the calculated energy barriers. This correlation supports the electrophilic attack as rate limiting in the PHBH reaction cycle and validates the relative activation barriers obtained from the QM/MM method.

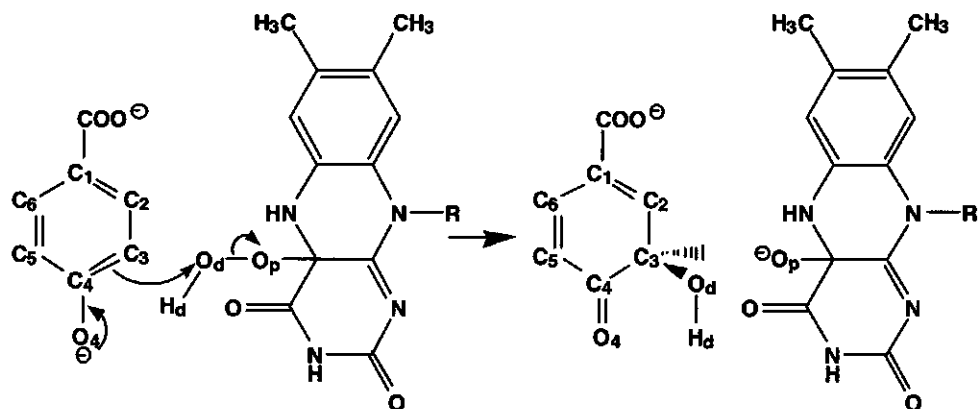


Figure 5.1 The electrophilic attack of the distal oxygen of the flavin cofactor on the substrate according to an electrophilic aromatic substitution mechanism. The cyclohexadienone formed converts into the final product 3,4-dihydroxybenzoate via keto-enol tautomerization.

5.1 Introduction and method

Para-hydroxybenzoate hydroxylase (PHBH, EC 1.14.13.2) is involved in the microbial degradation of aromatic compounds. It has become the model enzyme for the family of external flavoprotein monooxygenases (Entsch and Van Berkel, 1995). In the present study, a combined quantum mechanical (QM) and molecular mechanical (MM) method is applied to calculate energy barriers for the hydroxylation of the parent substrate and four of its fluorinated analogs within the active site of PHBH. The energy barriers obtained show a clear correlation ($R = 0.96$) with experimentally determined rate constants for their enzymatic conversion, demonstrating the biochemical relevance of the present approach.

The aromatic hydroxylation catalyzed by PHBH has been proposed to proceed via the electrophilic attack of an intermediate hydroperoxide form of the flavin cofactor (C4a-hydroperoxyflavin) on the substrate C3 (Figure 5.1) (Entsch *et al.*, 1976). This electrophilic attack is proposed to be the rate limiting step at physiological conditions (25 °C, pH 8) (Vervoort *et al.*, 1992). The crystal structure coordinates of PHBH in complex with the oxidized flavin cofactor and the substrate *p*-hydroxybenzoate (Schreuder *et al.*, 1989, Wierenga *et al.*, 1979) were used to build a three dimensional model of the C4a-hydroperoxyflavin intermediate of PHBH. The hydroxylation reaction in this large system was studied by using a combined QM/MM potential function (Field *et al.*, 1990) which treated the flavin ring, modified to the reactive C4a-hydroperoxyflavin intermediate (Schreuder *et al.*, 1990), and the dianionic form of the substrate (Entsch *et al.*, 1976, Vervoort *et al.*, 1992) quantum mechanically on the

basis of the closed shell semiempirical AM1 Hamiltonian (Dewar *et al.*, 1985). The steric and electronic influence of the surrounding enzyme was included by treating all other atoms molecular mechanically (Brooks *et al.*, 1983) using the CHARMM 22 forcefield (Quanta, 1993), describing polar hydrogens explicitly. For a correct description of the C1-C2 bond of the ribityl chain of the FAD cofactor, which crosses the QM/MM boundary, a 'link atom' (Field *et al.*, 1990) was introduced. The coordinates of 330 water molecules in the crystal structure were used to build explicit water molecules using the HBUILD facility in CHARMM. Energy minimization was performed with the Adopted Basis Newton-Raphson method for all atoms within a 10 Å sphere around the distal oxygen of the cofactor, including the substrate, the C4a-hydroperoxyflavin, a number of amino acid residues and 15 water molecules. All other atoms were fixed. A gradient tolerance of $0.01 \text{ kcal mol}^{-1} \text{ Å}^{-1}$ and nonbonded cutoff of 11 Å were applied. The optimized model for the enzyme-C4a-hydroperoxyflavin-substrate complex intermediate was used as the starting geometry in a reaction pathway calculation for the attack of the C4a-hydroperoxyflavin on the substrate.

A reaction coordinate r , describing the breaking of the peroxide oxygen-oxygen bond (Op-Od) and the formation of the bond between the distal peroxide oxygen and the substrate C3 (Figure 5.1), was applied by harmonically restraining ($k = 5000 \text{ kcal mol}^{-1} \text{ Å}^{-2}$) the difference in the lengths of these bonds. For each successive point along this reaction coordinate, energy minimization was performed as described above. The water molecules in the optimized region were harmonically restrained to their initial optimized positions ($k = 5.0 \text{ kcal mol}^{-1} \text{ Å}^{-2}$).

5.2 Results

Figure 5.2 presents the energies of the resulting pathway (solid line). The intermediate geometry at $r = -0.54$ is highest in energy and represents the approximate transition state structure of the hydroxylation step (Figure 5.3). Gas-phase normal mode analysis of this transition state geometry resulted in one significant negative eigenvalue corresponding to the transfer of the hydroxy group from the cofactor to the substrate, which validates this transition state geometry. The product geometry ($r = 1.55$) is a sigma adduct of the *p*-hydroxybenzoate in which the C3 is changed from a planar (sp^2) to a tetrahedral (sp^3) conformation. Mulliken charges on the QM atoms in the reactant, transition state and product geometries are given in Table 5.1. The total charges on the substrate ring and cofactor reveal a net charge transfer of one electron from the substrate to the C4a-hydroxyflavin (compare with Figure 5.1). Analysis of the geometrical changes in the substrate (results not shown) show that the C1-C2, C5-C6 and C4-O4 bond distances decrease whereas the other atomic distances in the carbon ring increase. Together with the observed decrease in charge on O4 these results show

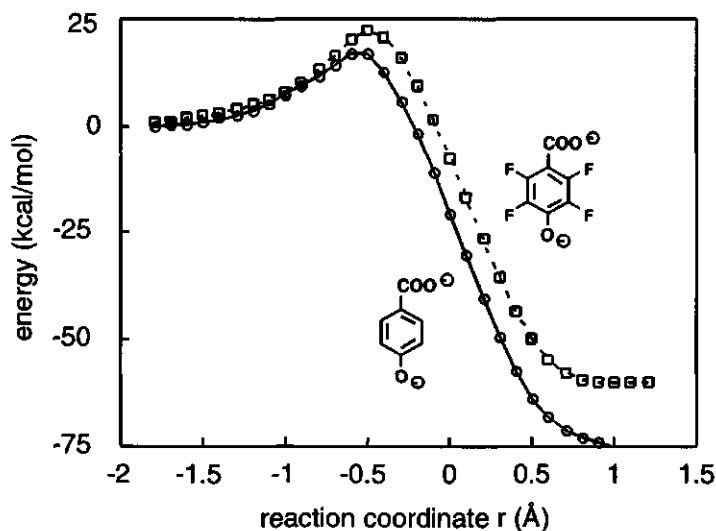


Figure 5.2 Energy profiles obtained for the electrophilic attack of the C4a-hydroperoxyflavin cofactor on *p*-hydroxybenzoate and 2,3,5,6-tetrafluoro-*p*-hydroxybenzoate. Energies relative to the reactants.

the formation of a cyclohexadienone, indicating that the present model is in agreement with an electrophilic aromatic substitution type of mechanism as proposed on the basis of transient kinetic studies for a homologous flavin dependent aromatic hydroxylase (Maeda-Yorita and Massey, 1993).

The reaction pathways for four fluorinated substrate derivatives were calculated following the same procedure. In the initial structure the substrate was placed such that the carbon atom to be hydroxylated is in the same position as C3 of the native substrate. For all fluorinated substrates each intermediate geometry obtained along the reaction coordinate was similar to the corresponding geometry of the reaction pathway

Table 5.1 Mulliken atomic charge distribution in the reactant, transition state and product geometries. Atom labeling as in Figure 5.1. The thin lines separate atoms which are included in the total charges for the cofactor and the substrate respectively.

atoms	reactants	transition state	products
flavin ring	0.10	0.00	-0.33
Op	-0.19	-0.37	-0.66
Od	-0.23	-0.13	-0.33
Hd	0.29	0.24	0.26
benzoate	-1.41	-1.21	-0.61
O4	-0.56	-0.53	-0.32
total cofactor	-0.03		-1.00
total substrate	-1.97		-1.00

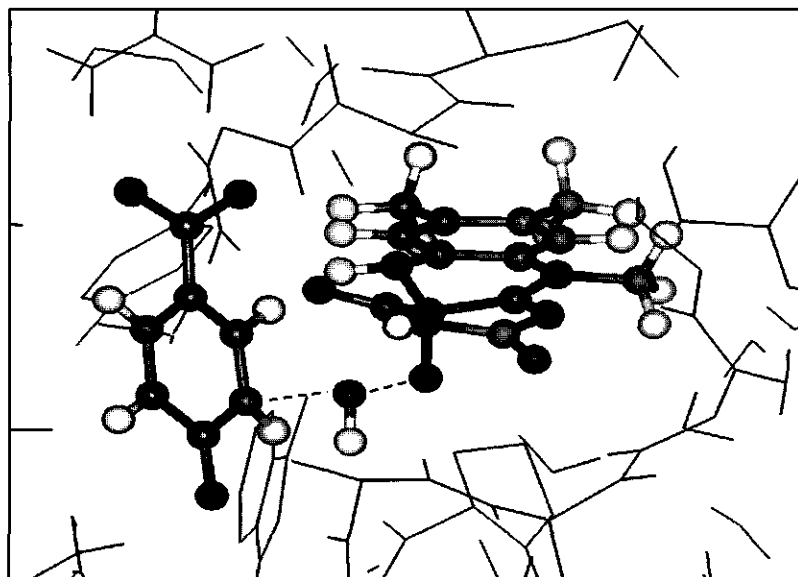


Figure 5.3 Approximate transition state structure of the present reaction pathway model. The dotted lines indicate the forming Od-C3 and breaking Op-Od bonds (Figure 5.1). The quantum mechanical atoms are represented in ball and stick.

with the native substrate, especially for the first part of the reaction coordinate towards the transition state (RMSD < 0.10 Å).

However, the energy profiles obtained for the different substrates differ with respect to the height of the barrier (Figure 5.2). The total energy differences between the initial minimized geometry and the transition state geometry, vary from 17.6 to 22.5 kcal/mol for the various *p*-hydroxybenzoates, increasing with increasing number of fluorine substituents (Table 5.2). These calculated energy barriers show a linear correlation (R

Table 5.2 Calculated energy barriers and the contributions of the QM/MM Van Der Waals interaction (QMVDw term) and the quantum mechanical energy, including the influence of the surrounding MM charges (QMElec term). Energies in kcal/mol. The positions of the transition states on the reaction coordinate are given (*r*). Labels of fluorine substituents are chosen such that hydroxylation is at C3 (Figure 5.2)

substrate	ΔE_{tot}	ΔQMVDw	ΔQMElec	<i>r</i> (Å)
pOHB	17.62	-0.24	19.53	-0.54
5-F-pOHB	18.30	-0.02	20.21	-0.52
2-F-pOHB	18.42	-0.26	20.11	-0.54
3,5-diF-pOHB	20.92	-0.62	22.99	-0.50
2,3,5,6-tetraF-pOHB	22.49	-0.67	24.08	-0.50

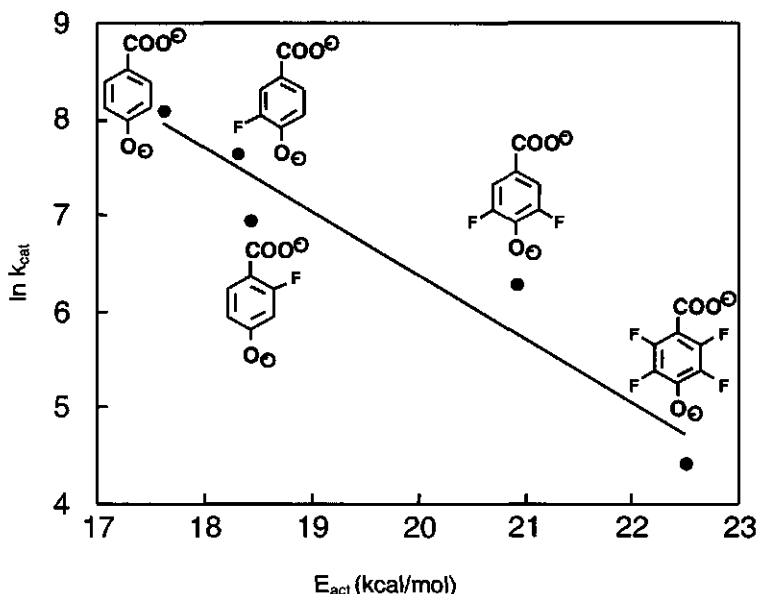


Figure 5.4 Linear correlation ($R = 0.96$) of experimental k_{cat} values for the overall hydroxylation of five *p*-hydroxybenzoate derivatives by PHBH at 25°C, pH 8 (Husain *et al.*, 1980) with calculated energy barriers for the present reaction pathway model.

= 0.96, Figure 5.4) with the natural logarithm of the experimental k_{cat} values for conversion of the different *p*-hydroxybenzoate derivatives by PHBH (Husain *et al.*, 1980). This correlation supports the proposal (Vervoort *et al.*, 1992) that the electrophilic attack is rate limiting under physiological conditions.

5.3 Discussion

Hydroxylation of some substrates by PHBH can be accompanied by a side-reaction leading to hydrogen peroxide formation. It has been suggested that this uncoupling reaction could be related to the tendency of the flavin to adopt the so-called "out" conformation (Gatti *et al.*, 1994). For the fluorinated substrates, however, this uncoupling reaction does not occur (Husain *et al.*, 1980, Van der Bolt *et al.*, 1997, Van der Bolt *et al.*, 1996).

The activation energy for *p*-hydroxybenzoate has been experimentally determined to be around 12 kcal/mol (Van Berkel and Müller, 1989). This indicates that the calculated energy barriers are roughly a factor 1.5 too high in comparison to experimental data. Since the energy barriers are dominated by the energy difference in the quantum mechanical energy term (not shown), the absolute overestimation of the energy barriers probably lies in the semiempirical AM1 method which is known to

overestimate absolute activation energies in many cases (Dewar *et al.*, 1985). Relative activation energies, however, appear to be more reliable (Hehre, 1997). The QM/MM method has been applied in several enzyme studies (Bash *et al.*, 1991, Cunningham *et al.*, 1997, Harrison *et al.*, 1997, Hartsough and Merz, 1995, Mulholland and Richards, 1997). To our knowledge the present study is the first demonstration of a correlation between calculated energy barriers and experimental rate constants for the enzymatic conversion of a series of substrates, which indicates that the QM/MM energy barriers are useful in a relative way. Furthermore, it indicates that studying a series of related fluorinated substrates provides a method to validate the QM/MM method. By using only fluorinated substrates, being similar in size compared to the non-substituted substrate, the present study was focused on the electronic effects of the substituents. Since steric effects are accounted for in the QM/MM method, the present approach may be extended to larger substituents. This indicates the potential of quantitative relationships between QM/MM energy barriers and experimental data to become a valuable tool in predicting rates of enzymatic conversions of new substrates.

Acknowledgements

This work was supported by the Netherlands Organization for Scientific Research (NWO). A. J. M. is a Wellcome Trust International Prize Travelling Research Fellow (Grant ref. no. 041229). We would like to thank Prof. Martin Karplus for useful discussions.

hydroperoxyflavin intermediate with the substrate. Analysis of the effect of individual active site residues on the reaction reveals a specific transition state stabilisation by the backbone carbonyl moiety of Pro293. The crystal water 717 appears to drive the hydroxylation step through a stabilising hydrogen bond interaction to the proximal oxygen of the C4a-hydroperoxyflavin intermediate which increases in strength as the hydroperoxyflavin cofactor converts to the anionic (deprotonated) hydroxyflavin.

6.1 Introduction

The flavoprotein monooxygenase *p*-hydroxybenzoate hydroxylase (PHBH, E.C. 1.14.13.2), catalyses the 3-hydroxylation of *p*-hydroxybenzoate, which is an important step in the microbial degradation pathways of a wide variety of aromatic compounds. The first steps in the reaction cycle of PHBH involve binding of the substrate followed by a two-electron reduction of the flavin cofactor by NADPH and the incorporation of molecular oxygen to form the C4a-peroxyflavin intermediate. Upon protonation of the distal oxygen of the C4a-peroxyflavin, the so-called C4a-hydroperoxyflavin is formed which reacts with the substrate leading to the formation of the hydroxylated product and the hydroxyflavin form of the cofactor. Finally, release of water from the hydroxyflavin leads to the formation of the original oxidised state of the flavin cofactor. The attack of the C4a-hydroperoxyflavin on the substrate has previously been suggested to be rate limiting at pH 8, 25°C (Vervoort *et al.*, 1992).

In the last decade, two questions concerning the mechanism of hydroxylation have been the subject of several studies. First, the required protonation states of the substrate and of the activated cofactor for the hydroxylation reaction are of interest. Deprotonation of the hydroxyl moiety at C4 of the aromatic substrate is suggested to be essential for its conversion. This conclusion was originally based on the observation that compounds lacking this hydroxyl moiety are not converted (Entsch *et al.*, 1976). Furthermore, studies on the Tyr201Phe mutant enzyme indicates that a hydrogen bond network formed by Tyr201 and Tyr385 has a crucial role in deprotonating the substrate in the active site. The considerable decrease in efficiency of hydroxylation in the mutant, as a result of a loss of substrate deprotonation, further supports the proposal that the deprotonation of the substrate hydroxyl moiety in wild type enzyme is needed to activate the substrate for hydroxylation (Entsch *et al.*, 1991, Eschrich *et al.*, 1993). Molecular orbital studies on the *p*-hydroxybenzoate substrate have demonstrated that deprotonation increases the nucleophilic reactivity of the C3' centre in the substrate which could be essential for an electrophilic attack on this C3' centre by the C4a-(hydro)peroxyflavin cofactor (Vervoort *et al.*, 1992). With respect to the protonation state of the C4a-peroxyflavin cofactor, some information comes from transient kinetic studies on a related flavin dependent aromatic hydroxylase, phenol

hydroxylase from *Trichosporon cutaneum* (Maeda-Yorita and Massey, 1993). In these experiments several intermediates were observed, two of which were ascribed to respectively the unprotonated and protonated form of the flavin intermediate, i.e. to the C4a-peroxyflavin and to the C4a-hydroperoxyflavin. This experimental observation supports the proposal that the peroxide moiety becomes protonated to activate the peroxyflavin for the transfer of oxygen to the substrate C3'.

A second major question with respect to the catalytic mechanism of PHBH concerns the splitting of the peroxide dioxygen bond in the hydroxylation step. Mechanisms involving heterolytic cleavage (Husain *et al.*, 1980, Keum *et al.*, 1990) as well as mechanisms involving homolytic cleavage (Anderson *et al.*, 1987, Keum *et al.*, 1990, Anderson *et al.*, 1991, Peräkylä and Pakkanen, 1993) of the peroxide oxygen-oxygen bond have been proposed. At present, one of the earliest suggestions of electrophilic aromatic substitution involving heterolytic cleavage of the peroxide bond is supported by experimental data (Maeda-Yorita and Massey, 1993). However, definite evidence with respect to the exact reaction mechanism seems to be lacking, which is mainly due to the short lifetime of the reaction intermediates which are formed upon hydroxylation. This makes it difficult to obtain detailed experimental information on the nature of this hydroxylation step and the reaction intermediates formed.

In the present study, a combined quantum mechanical and molecular mechanical (QM/MM) reaction pathway analysis of the hydroxylation step in the PHBH reaction cycle has been performed. The combination of quantum mechanics and molecular mechanics is a promising approach to describe enzymatic reactions within the environment of the protein (Bash *et al.*, 1991, Cunningham *et al.*, 1997, Harrison *et al.*, 1997, Hartsough and Merz, 1995, Mulholland and Richards, 1997). Quantum chemical (QM) treatment is preferred for chemical reactions, which involve breaking and forming of bonds and electronic redistribution. On the other hand, molecular mechanical (MM) methods are more practical for large protein structures. In QM/MM methods, these a molecular system is separated into a small quantum mechanical region, describing the reacting part of the system (substrate, cofactor and/or catalytic amino acids), and a molecular mechanical region describing the surrounding protein and solvent molecules (Mulholland and Karplus, 1996).

The validity of the present reaction pathway model for the hydroxylation in the PHBH reaction cycle, obtained with the QM/MM method described by Field *et al.* (Field *et al.*, 1990), has been demonstrated previously through the correlation of the calculated energy barriers for the hydroxylation of a series of fluorinated substrate derivatives with experimental rate constants for the conversion of these substrates by PHBH (Ridder *et al.*, 1998). In the present study this QM/MM reaction pathway model is

applied to yield additional insight into the mechanism of the hydroxylation reaction catalysed by PHBH and the electronic characteristics of the intermediates.

6.2 Methods

Preparation of the protein model

A three dimensional structure of the C4a-hydroperoxyflavin-substrate intermediate of PHBH was built on the basis of the crystal structure of oxidised PHBH in complex with the substrate *p*-hydroxybenzoate (Schreuder *et al.*, 1989, Wierenga *et al.*, 1979) obtained from the Protein Databank (entry code 1PBE.PDB). Two other crystal structures, of the reduced PHBH-substrate complex and of the PHBH-product complex, are very similar to the structure of the oxidised PHBH-substrate complex (Schreuder *et al.*, 1989, Schreuder *et al.*, 1991). On the basis of this similarity, no dramatic changes in the protein conformation around the active site are expected for the C4a-hydroperoxyflavin intermediate in the reaction cycle and during the subsequent reaction step.

The flavin cofactor present in the crystal structure in its resting state was modified to the active C4a-hydroperoxyflavin intermediate using the coordinates reported by Schreuder *et al.* (Schreuder *et al.*, 1990) on the basis of molecular modelling studies and experimental NMR data of the C4a-hydroperoxyflavin intermediate (Vervoort *et al.*, 1986). Hydrogen atoms were added to the protein according to an extended atom topology (i.e. only polar group hydrogens were modelled explicitly) using CHARMM (Brooks *et al.*, 1983). An all-hydrogen topology was used for the substrate and the isoalloxazine ring of the cofactor, because this part of the structure was treated quantum mechanically. The hydroxyl and carboxyl moieties of the substrate were modelled in the anionic state. The coordinates of 330 water molecules in the crystal structure were used to build explicit water molecules in the model using the HBUILD routine of the CHARMM program. Since in this structure the substrate and the peroxide moiety of the cofactor are not exposed to solvent it was not considered necessary to add more water molecules to the model.

QM/MM calculations

On the basis of the starting geometry described above, QM/MM calculations were performed using CHARMM version 24b1 (Brooks *et al.*, 1983). The C4a-hydroperoxyflavin and the dianionic substrate were treated quantum mechanically on the basis of a closed shell semiempirical AM1 Hamiltonian (Dewar *et al.*, 1985). The covalent C-C bond between the C1* and the C2* of the ribityl chain of the FAD cofactor, which crosses the boundary between the QM and MM regions, was replaced

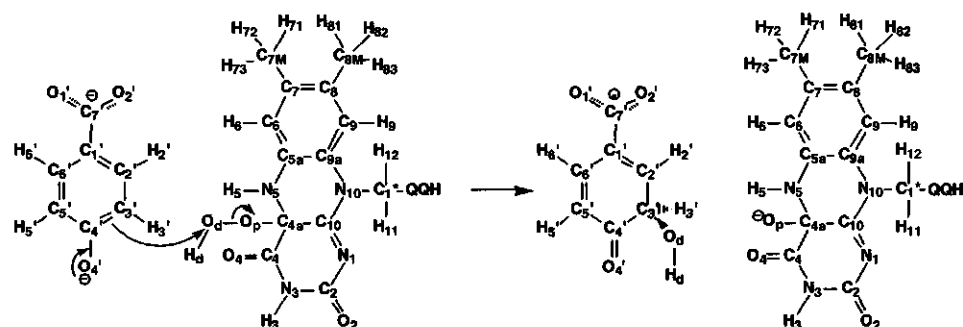


Figure 6.1 The hydroxylation of *p*-hydroxybenzoate according to an electrophilic aromatic substitution mechanism and numbering of the atoms as used throughout the present paper.

by a C-QQH (link atom) bond in the QM system (Field *et al.*, 1990). All other atoms were treated at the molecular mechanical level (MM), using the CHARMM 22 forcefield (Quanta, 1993). The complete simulation system consists of 50 QM atoms, and 4840 MM atoms. The QM region has a total charge of $-2e$. All energies were evaluated with a nonbonded cutoff distance of 11 Å, using a group based switching function to smooth the QM/MM interactions to zero. The coordinates of 295 atoms within a 10 Å sphere around the distal oxygen of the cofactor were optimised to a gradient tolerance of $0.01 \text{ kcal mol}^{-1} \text{ Å}^{-1}$ ($1 \text{ cal} = 4.184 \text{ J}$), requiring 780 steps of Adopted Basis Newton-Raphson (ABNR) minimization. This geometry optimised region included the substrate, the flavin ring, (parts of) 37 amino acid residues in the active site and 15 water molecules. All other atoms were fixed during the calculations.

Validation of AM1 for the QM system

In the model described above the substrate and the flavin ring of the cofactor were treated quantum mechanically (QM) by the semiempirical AM1 method. This QM system has a total charge of $-2e$. In the case of *ab initio* methods, calculations on anions require basis sets with diffuse functions. Semiempirical methods, such as AM1, use minimal basis sets which do not include diffuse functions. Nevertheless, AM1 often performs well on anionic systems and appears to be able to correct for the increased electron repulsion through the parameterization (Dewar *et al.*, 1985). The competence of AM1 for even a dianionic system is illustrated by a recent comparison of AM1 with *ab initio* methods for dianionic oxaloacetate, for which the AM1 results were similar to those from *ab initio* calculations including diffuse functions and performed even better than *ab initio* methods without diffuse functions (Mulholland and Richards, 1998a).

To test AM1 for the dianionic QM system in the present PHBH model, gas-phase calculations were performed for the separate reactants, *p*-hydroxybenzoate and C4a-hydroperoxyflavin, and the expected products, hydroxycyclohexadienone and ionized hydroxyflavin and the results were compared to ab initio calculations which include diffuse functions. These gas-phase calculations were performed with Spartan 5.1 and Gaussian98 using the AM1, HF/6-31+G(d) and B3LYP/6-31+G(d) methods.

Definition of the reaction pathway

The QM/MM model for the C4a-hydroperoxyflavin-substrate intermediate was used as a starting geometry for a reaction pathway calculation describing the reaction of the C4a-hydroperoxyflavin with the substrate. Figure 6.1 schematically presents this hydroxylation reaction according to an electrophilic aromatic substitution mechanism, as proposed for the aromatic hydroxylation of phenol by the homologous enzyme phenol hydroxylase (Maeda-Yorita and Massey, 1993). The calculation of the hydroxylation step was performed by applying the reaction coordinate parameter $r = d(\text{Op-Od}) - d(\text{Od-C3'})$ as an approximation to the intrinsic reaction coordinate. This reaction coordinate was harmonically restrained ($k = 5000 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$) using the CHARMM command RESDistance. By changing r in steps of 0.1 \AA , from -1.8 \AA , corresponding to the geometry of the reactants, to $+1.2 \text{ \AA}$, corresponding to the geometry of the products, the distal oxygen of the flavin cofactor was forced to move stepwise from the C4a-hydroperoxyflavin towards the substrate, corresponding to the process of hydroxylation. For each value of r , the active site region (i.e. the 10 \AA sphere) was optimised to the same gradient tolerance used for the optimisation of the initial geometry, requiring 60 to 250 steps of ABNR minimization. To prevent the water molecules in the optimised region from moving over large distances, which may cause undesired discontinuities along the reaction pathway, they were harmonically restrained to their initial (optimised) coordinates, using a small force constant of $5.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$.

Other reaction coordinates could have been chosen for the simulation. The reaction coordinate used here, (i.e. $r = d(\text{Op-Od}) - d(\text{Od-C3'})$) is justified by the smooth and continuous form of the resulting path and the corresponding energy profile (see Results section). Furthermore, the reaction coordinate was tested by performing the simulation in reverse direction, using the same method but starting from the optimised product geometry. This yielded identical results.

The top of the energy profile represents the approximate transition state of the hydroxylation step in the reaction cycle. To determine this maximum more precisely, the energy of some additional points on the reaction coordinate were calculated, varying r in steps of 0.01 \AA around the top in the energy profile. To analyse the

QM/MM transition state geometry obtained, vibrational analysis of the QM part of the transition state was performed *in vacuo* using Spartan 5.0. Charges on the quantum mechanical atoms within the electrostatic potential of the molecular mechanical atoms were obtained from the Mulliken analysis routine available in CHARMM.

It is useful to note that the closed shell AM1 method, which is used for the quantum mechanical region in the present model, cannot describe radical type mechanisms. Therefore, the present study is not an attempt to answer definitely the discussion with respect to heterolytic cleavage versus homolytic cleavage of the peroxide oxygen-oxygen bond. However, the results are relevant for heterolytic bond dissociation reactions, such as the electrophilic aromatic substitution mechanism which is most recently supported by experimental evidence for the flavin dependent monooxygenases (Maeda-Yorita and Massey, 1993). Other limitations of the closed shell AM1 method are that AM1 has not been parameterized for unstable structures, such as transition states, and that possible interactions of the groundstate configuration with other electronic configurations are not taken into account. However, in spite of these limitations, the usefulness of AM1 in the present model is supported by a correlation of the resulting energy barriers with the logarithm of experimental rate constants for a series of fluorinated *p*-hydroxybenzoates (Ridder *et al.*, 1998).

Frontier orbital analysis

Electronic details of the reaction mechanism described by the QM/MM model were obtained by analysing the energies and density distributions of the four highest occupied molecular orbitals (HOMOs) and the four lowest unoccupied molecular orbitals (LUMOs) on the QM atoms for all reaction pathway geometries. The orbital density on a specific QM atom is defined as the sum of the squared molecular orbital coefficients for the contributions of the relevant atomic orbitals. The HOMO and LUMO density distributions represent nucleophilic and electrophilic reactivity respectively, according to frontier orbital theory (Fleming, 1976). The molecular orbital energies and density distributions, were obtained from *in vacuo* single point calculations on the geometries of the reaction pathway intermediates, using the AM1 method implemented in Spartan 5.0.

Amino acid decomposition analysis

In order to obtain qualitative information on the contribution of specific amino acids to the QM/MM energy barrier and to the energy difference between the reactants and products of the hydroxylation step, energy decomposition analyses were performed with a procedure similar to those used previously (Bash *et al.*, 1991, Cunningham *et al.*, 1997, Mulholland and Richards, 1997). The analyses start from the energy

difference between the reactants and the transition state, or the reactants and the products, without the MM protein environment. Then, one by one the amino acid residues are included in the QM/MM energy difference calculation in order of increasing distance between their centre of mass (COM) and the distal oxygen of the hydroperoxyflavin. The effect of an (MM) amino acid on the QM/MM energy difference gives an approximate and qualitative indication of its influence on the hydroxylation step in the active site of PHBH. The results obtained from this analysis do not include dielectric screening and other solvation effects and therefore overestimate the effect of distant residues (Mulholland and Richards, 1997). It should be clear that the analysis relates specifically to the hydroxylation step in the reaction cycle. The results, therefore, are not necessarily directly comparable to experimental studies on mutant enzymes in which effects on other steps as well as changes in protein structure can play a role.

6.3 Results

Comparison of AM1, HF and B3LYP results

Table 6.1 presents the results of the gas-phase calculations performed to validate the AM1 method for the relevant reactant and product molecules in the PHBH model. Although the peroxide O-O bond length in the C4a-hydroperoxyflavin appears to be somewhat underestimated by AM1, the other geometrical variables involving the hydroperoxide moiety are in good agreement with the HF/6-31+G(d) and B3LYP/6-31+G(d) results. The differences in the C4a-Op bond length and the C10-C4a-Op angle between the C4a-hydroperoxyflavin and C4a-hydroxyflavin, as predicted by the B3LYP/6-31+G(d) method, are well reproduced by AM1. Furthermore, it is interesting to note that for most of the C-O bonds in the different compounds, the AM1 values are closer to the results from B3LYP/6-31+G(d) than to those from HF/6-31+G(d). The reasonable agreement between AM1 and B3LYP/6-31+G(d), which is the best level of theory used and includes electron correlation to a certain extent, supports the accuracy of AM1 for the QM system in the present QM/MM model.

QM/MM Reaction pathway calculation for the hydroxylation reaction in p-hydroxybenzoate hydroxylase

Figure 6.2a presents the geometry of the active site region obtained by energy minimization of the initial model derived from the crystal structure, using the constraints and cutoffs as defined in the Methods section. The non-hydrogen atoms deviated from the initial coordinates, with a root-mean-squared difference (RMSD) of

Table 6.1 A representative selection of bond lengths (Å), angles (in degrees), dihedral angles (in degrees) and energy difference (in kcal/mol) of AM1, HF and B3LYP optimised structures (in vacuum) of C4a-hydroperoxyflavin, *p*-hydroxybenzoate, C4a-peroxyflavin and hydroxycyclohexadienone. Labels as in Figure 6.1.

molecule	parameter	AM1	HF/6-31+G(d)	B3LYP/6-31+G(d)
C4a-hydroperoxyflavin	C4a-Op	1.484	1.425	1.486
	Op-Od	1.280	1.385	1.444
	C4a-Op-Od	113.3	110.3	108.9
	C10-C4a-Op	103.2	102.7	102.2
	C4a-Op-Od-Hd	-82.9	-92.8	-84.6
C4a-hydroxyflavin	C4a-Op	1.297	1.309	1.323
	C10-C4a-Op	109.9	108.5	108.5
<i>p</i> -hydroxybenzoate	C1'-C2'	1.402	1.399	1.410
	C2'-C3'	1.382	1.385	1.396
	C3'-C4'	1.438	1.434	1.443
	C4'-O4'	1.279	1.266	1.291
	C1'-C7'	1.497	1.522	1.531
	C7'-O1'	1.276	1.247	1.273
hydroxycyclohexadienone	C1'-C2'	1.345	1.329	1.352
	C2'-C3'	1.492	1.500	1.497
	C3'-C4'	1.524	1.523	1.533
	C4'-O4'	1.241	1.208	1.240
	C1'-C7'	1.525	1.545	1.553
	C7'-O1'	1.264	1.235	1.261
	C3'-Od	1.419	1.396	1.417
	O4'-C4'-C3'-Od	-26.9	-27.3	-23.4
$\Delta E_{\text{products-reactants}}^*$		-139.97	-130.94	-121.56

* Note: An unusual large zero-point energy (ZPE) correction (+80 kcal/mol) for the energy difference was found at the HF/6-31+G(d) level, mostly due to a large ZPE difference between the C4a-hydroperoxyflavin and the C4a-hydroxyflavin. Although higher level frequency calculations would be required to test the validity of this large ZPE correction, it indicates that the enthalpy values may be significantly different from the energy values below. This also suggests that the AM1 results should be interpreted as energies rather than enthalpies.

Table 6.2 Hydrogen bond distances between active site amino acids and the substrate in the reactants (*p*-hydroxybenzoate - C4a-hydroperoxyflavin), the transition state ($r=-0.54$) and the products (cyclohexadienone - C4a-hydroxyflavin). Substrate atom labels as in Figure 6.1.

hydrogen bond (Å)	reactants	transition state	products
amino acid - substrate			
Ser212-H - O1'	2.06	2.03	1.97
Arg214-H - O1'	1.82	1.81	1.80
Arg214-H - O2'	1.94	1.89	1.83
Tyr222-H - O2'	1.96	1.93	1.92
Tyr201-H - O4'	1.89	1.93	2.10
Pro293=O - Hd	2.35	1.98	2.42
Wat717-H - Op	2.20	2.11	2.01
Wat717-H - Od	2.16	3.14	4.01

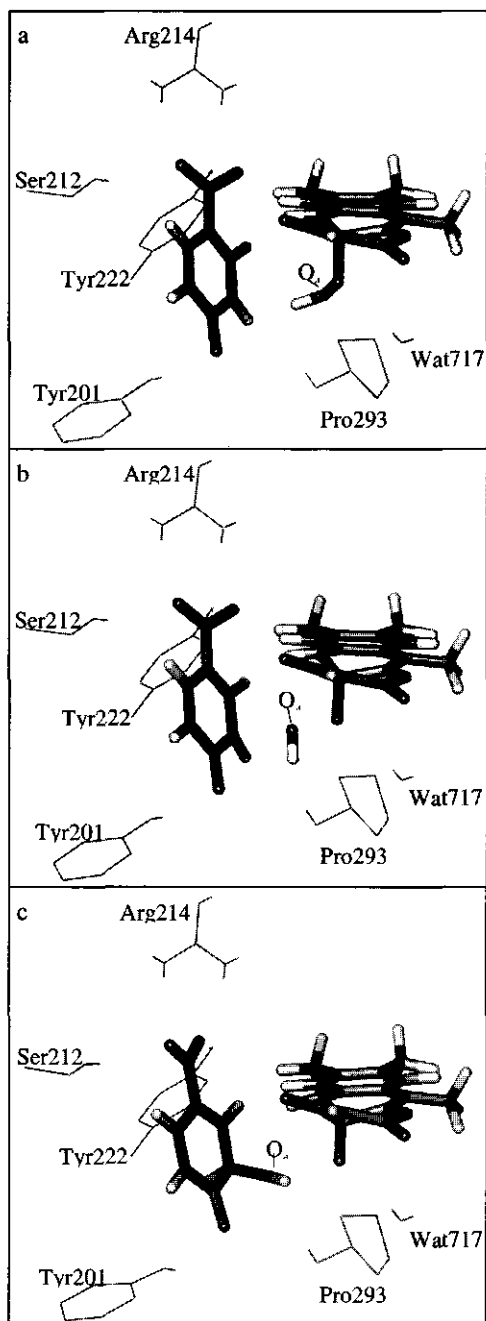


Figure 6.2 The active site region in a) the substrate - C4a-hydroperoxyflavin reactant state, b) the transition state and c) the cyclohexadienone - C4a-hydroxyflavin product state. The MM atoms are shown in thin lines and QM atoms are shown in thicker bonds.

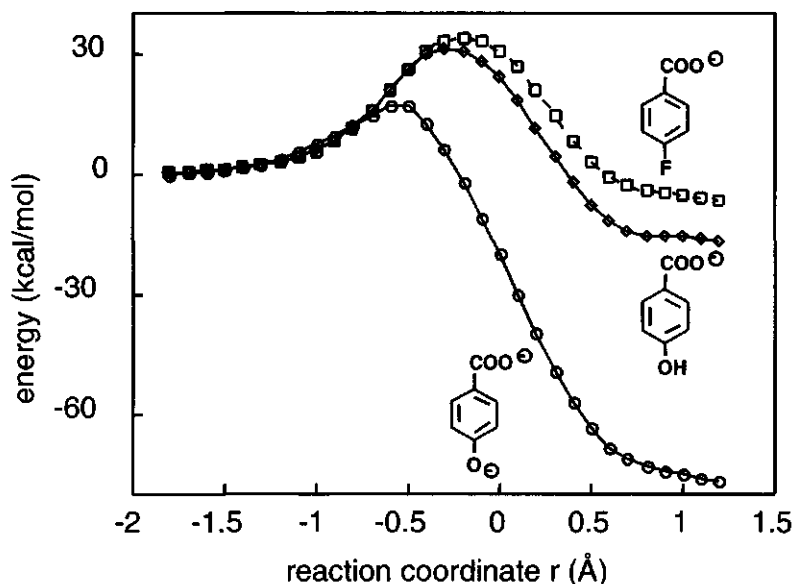


Figure 6.3 QM/MM reaction pathway energy profiles of the hydroxylation of *p*-hydroxybenzoate in its dianionic and monoanionic form and of the effector *p*-fluorobenzoate, as function of the reaction coordinate $r = d(\text{Op-Od}) - d(\text{Od-C3'})$. Energies are given relative to the reactants.

0.44 Å. Table 6.2 shows that notable hydrogen bond interactions between the substrate and active site amino acids, which have been reported to be present in the crystal structures of PHBH in complex with *p*-hydroxybenzoate and 3,4-dihydroxybenzoate (Schreuder *et al.*, 1989), are preserved in the optimised model. An additional hydrogen bond is observed between the distal oxygen of C4a-hydroperoxyflavin and a water molecule (Wat717). The optimised structure represents the enzyme–C4a-hydroperoxyflavin–substrate complex in the PHBH reaction cycle and was used as the starting geometry for the reaction pathway calculation for the aromatic hydroxylation of the substrate. Figure 6.3 presents the energies of the calculated reaction pathway intermediates as a function of the reaction coordinate variable r , relative to the energy of the initial optimised geometry. The energy maximum is at $r = -0.54$. This point represents the transition state structure of the hydroxylation step (Figure 6.2b). Gas-phase normal mode analysis validates this transition state geometry by showing only one large negative eigenvalue corresponding to the transfer of the hydroxy group from the cofactor to the substrate, which is absent in the reactant and product geometries (results not shown). The energy barrier to the hydroxylation step is given by the difference between the energy of this transition state and the energy of the initial structure. This QM/MM calculated energy barrier of 17.6 kcal/mol is approximately 1.5 times higher than the experimental activation energy (Van Berkel and Müller,

1989). The last point on the reaction pathway ($r = 1.2 \text{ \AA}$) was further optimised, without the reaction coordinate distance restraint term. The resulting optimised product geometry (Figure 6.2c) remained almost identical to the last pathway geometry (RMSD = 0.03 \AA) and contains deprotonated C4a-hydroxyflavin and a sigma adduct of *p*-hydroxybenzoate. In this product of the calculated reaction pathway, the planar conformation of the C3' of the substrate is changed to a tetrahedral conformation as expected for the cyclohexadienone product shown in Figure 6.1.

Table 6.3 Atomic charge distribution in the reactants, transition state and products according to Mulliken analysis of the QM/MM model. Atom labeling as in Figure 6.1.

atom	reactants	TS	products	atom	reactants	TS	products
N1	-0.326	-0.343	-0.359	Op	-0.190	-0.375	-0.665
C2	0.384	0.386	0.394	Od	-0.228	-0.132	-0.331
O2	-0.323	-0.342	-0.374	Hd	0.291	0.237	0.260
N3	-0.377	-0.384	-0.386				
H3	0.317	0.307	0.286	C1'	-0.319	-0.282	-0.129
C4	0.324	0.312	0.273	C7'	0.388	0.389	0.382
O4	-0.336	-0.355	-0.430	O1'	-0.699	-0.691	-0.612
C4a	0.132	0.197	0.307	O2'	-0.716	-0.704	-0.691
N5	-0.280	-0.282	-0.310	C2'	-0.020	-0.019	-0.030
H5	0.305	0.281	0.222	H2'	0.112	0.123	0.167
C5a	0.064	0.066	0.090	C6'	-0.047	-0.007	-0.136
C6	-0.134	-0.146	-0.183	H6'	0.077	0.091	0.135
H6	0.207	0.195	0.146	C3'	-0.312	-0.288	-0.245
C7	-0.068	-0.070	-0.069	H3'	0.132	0.140	0.173
C7M	-0.174	-0.173	-0.169	C5'	-0.398	-0.356	0.005
H71	0.084	0.081	0.082	H5'	0.105	0.115	0.122
H72	0.103	0.099	0.086	C4'	0.285	0.291	0.250
H73	0.054	0.053	0.049	O4'	-0.556	-0.526	-0.323
C8	-0.090	-0.099	-0.116				
C8M	-0.176	-0.174	-0.170	total			
H81	0.081	0.079	0.072	flavin-Op	-0.094	-0.380	-0.997
H82	0.088	0.087	0.086				
H83	0.086	0.083	0.079	total			
C9	-0.149	-0.149	-0.151	Od-Hd	0.063	0.105	-0.071
H9	0.120	0.118	0.113				
C9a	-0.007	-0.006	0.009	total			
N10	-0.185	-0.191	-0.198	substrate	-1.969	-1.725	-0.933
C10	0.180	0.179	0.131				
C1*	-0.076	-0.072	-0.077				
H11	0.060	0.054	0.045				
H12	0.113	0.112	0.108				
QQH	0.096	0.093	0.085				

Table 6.4 Bond lengths (Å) and bond orders (Spartan 5.1) in the reactants (*p*-hydroxybenzoate - C4a-hydroperoxyflavin), the transition state ($r=-0.54$) and the products (cyclohexadienone - C4a-hydroxyflavin). Atom labels as in Figure 6.1.

Bond	Bond length (Å)			Bond order		
	reactants	transition state	products	reactants	transition state	products
Op-Od	1.28	1.52	2.96	1.03	0.71	0.00
Od-C3'	3.09	2.06	1.41	0.01	0.21	1.01
C1'-C2'	1.40	1.38	1.35	1.39	1.51	1.82
C2'-C3'	1.38	1.41	1.49	1.44	1.27	0.97
C3'-C4'	1.44	1.45	1.52	1.14	1.10	0.90
C4'-C5'	1.44	1.45	1.46	1.12	1.08	1.00
C5'-C6'	1.37	1.36	1.35	1.54	1.62	1.78
C6'-C1'	1.41	1.42	1.45	1.30	1.22	1.06
C4'-O4'	1.27	1.27	1.24	1.53	1.57	1.84

Analysis of changes in charge distribution and bond distances along the reaction pathway

The atomic charges on the QM atoms in the reactant, transition state and product geometries, including the effect of the protein, were obtained by performing a Mulliken analysis in CHARMM (Table 6.3). The changes in atomic charge are indicative of an electrophilic aromatic substitution type of mechanism (Figure 6.1). This especially follows from the total charges on the substrate ring and cofactor which reveal -1e charge transfer from the substrate to the C4a-hydroxyflavin (Table 6.3). Table 6.4 presents the changes in bond distances and bond orders in the carbon ring of the substrate. Along the calculated reaction pathway, the bond orders of the C1'-C2' bond and the C5'-C6' bond increase, whereas the other bonds in the carbon ring become single bonds, demonstrating that the aromatic carbon ring has changed to a cyclohexadienone. Furthermore, the decrease in charge on O4' (Table 6.3) and the increase in C4'-O4' bond order (Table 6.4) are in agreement with the formation of a carbonyl group at the *para* position.

Changes in the orientation of active site amino acids

The RMSD for non-hydrogen atoms between the optimised coordinates of subsequent pathway geometries stayed within 0.05 Å, indicating that only small changes in the conformation of the optimised atoms in the active site occur along the reaction coordinate. The small structural changes occur especially at three sites in the optimised region. First, Ser212, Arg214 and Tyr222, which form the binding site for the carboxylate moiety of the substrate, adapt their orientation in response to a small rotation of the substrate. As a result, the 4 hydrogen bonds between the carboxylate moiety of

the substrate and these amino acid residues are well preserved along the reaction coordinate (Table 6.2). Second, Tyr201 and Tyr385, which form a hydrogen bonding network to stabilise the deprotonated hydroxyl moiety of the substrate, also adapt to the movement of the substrate. Notably, the hydrogen bond distance between the hydroxyl-proton of Tyr201 and O4' of the substrate increases along the reaction coordinate from 1.89 Å to 2.10 Å (Table 6.2). The third movement is observed around the peptide bond between Pro293 and Thr294. Towards the transition state, this part of the backbone shows a small rotation. Due to this rotation, the carbonyl group of Pro293 points towards Hd of the hydroperoxyflavin and interacts with the Od-Hd moiety in the process of being transferred to the substrate (Figure 6.2b, Table 6.2). At the end of the reaction pathway this carbonyl group rotates back to its original position.

In the crystal structure, Wat717 is hydrogen bonded to the backbone nitrogen of Gly298. Upon minimization of the present model in which oxidized flavin is replaced by C4a-hydroperoxyflavin, this water molecule forms a hydrogen bond with the peroxide moiety of the cofactor. As the reaction proceeds, this interaction between Wat717 and the proximal oxygen of the peroxide becomes stronger (Table 6.2).

Activation of the substrate

In the model described above, the hydroxyl moiety at the C4' position of the substrate is deprotonated. This deprotonation is proposed to activate the substrate for the nucleophilic attack on the cofactor (Entsch *et al.*, 1991). To investigate the importance of deprotonation, similar reaction pathway calculations were performed for the aromatic hydroxylation of the protonated substrate and of the effector *p*-fluorobenzoate in which the hydroxyl moiety is replaced by a fluorine substituent. Figure 6.3 shows that the energy barriers obtained for these two substrate analogues are a factor of 2 higher than for the native substrate. On the basis of a correlation between the calculated energy barriers and the logarithm of experimental rate constants of conversion a series of fluorinated substrates (Ridder *et al.*, 1998) these barriers would correspond to a rate constant of about 0.01 s^{-1} . Thus, these results support the proposal that deprotonation of the *p*-hydroxybenzoate substrate is essential for the hydroxylation reaction. They are also in agreement with the experimental observation that with *p*-fluorobenzoate hydroxylation does not proceed, although *p*-fluorobenzoate binds to the active site of PHBH and initiates reduction of the flavin and formation of the C4a-hydroperoxyflavin.

To obtain more detailed insight into the reaction process the changes in energy and distribution of the molecular orbitals involved were analyzed. Since the analysis of the AM1 orbitals does not include the effects of the protein it is only expected to give a qualitative picture of the changes along the reaction coordinate. Figure 6.4a presents

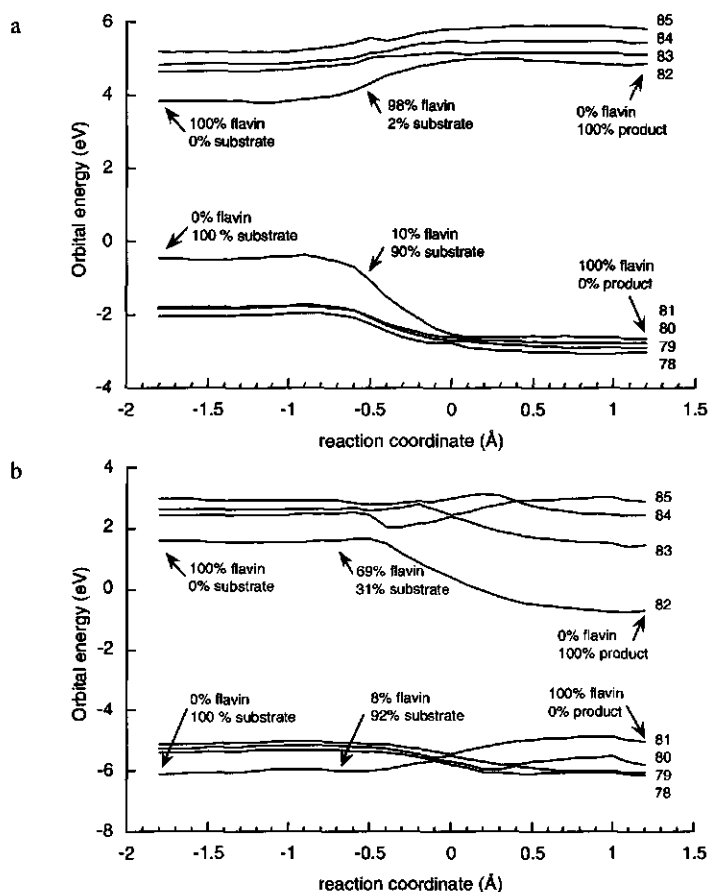


Figure 6.4 Walsh diagram, showing the energies of orbitals 78–85, i.e. from HOMO–3 to LUMO+3 as a function of the reaction coordinate for a) hydroxylation of deprotonated substrate and b) protonated substrate. The density distributions of the interacting HOMO and LUMO orbitals are represented as percentages on substrate and cofactor.

the energies of the four highest occupied molecular orbitals (HOMOs) and the four lowest unoccupied molecular orbitals (LUMOs), as functions of the reaction coordinate, for the model with deprotonated substrate. This so-called Walsh diagram can be interpreted in terms of frontier orbital theory. In the initial cofactor substrate complex the HOMO and LUMO frontier orbitals are energetically separated from the other occupied and unoccupied orbitals. Furthermore, the HOMO is fully localized on the *p*-hydroxybenzoate (graphically presented in Figure 6.5a) which is in accordance with its role as nucleophile in the hydroxylation reaction, whereas the LUMO is localized on the C4a-hydroperoxyflavin in line with its role as electrophile. As the reaction proceeds the HOMO and LUMO orbitals interact which results in a decrease

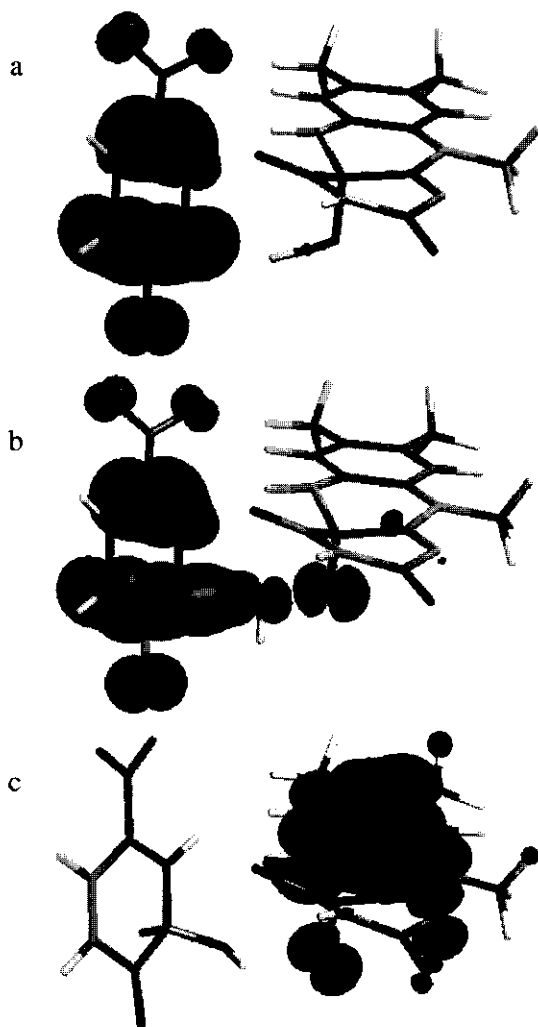


Figure 6.5 Isodensity surface (0.03 electrons/ \AA^3) of the highest occupied molecular orbital in a) the reactant state, b) the transition state and c) the product state obtained by *in vacuo* single point calculations (AM1) on the geometries of the QM atoms obtained in the QM/MM reaction pathway calculation with dianionic 4-hydroxybenzoate.

of the HOMO energy and an increase in the LUMO energy (Figure 6.4a). Meanwhile, the HOMO density gradually moves from the substrate to the cofactor (Figures 6.5b and 6.5c) and the LUMO density moves from the cofactor to the substrate (see discussion below).

For comparison, Figure 6.4b shows the Walsh diagram for the calculated reaction pathway with protonated substrate. In this case, the HOMO, HOMO-1 and HOMO-2

are all located around the carboxylate moiety of the substrate and have no density on the reacting C3' atom. Analysis of the orbital distributions indicates that the HOMO-3, which initially has density on C3', is the true interacting orbital. Comparison of Figures 6.4a and 6.4b shows that the energy gap between the interacting HOMO-3 and LUMO for the reaction with protonated substrate (7.7 eV) is almost twice as large as the energy gap between the interacting HOMO and LUMO for the reaction with unprotonated substrate (4.3 eV) which is in line with the difference in energy barrier observed.

Activation of the cofactor

Figures 6.6a and 6.6b present the changes in bond distances, along the calculated reaction pathway, of the bonds included in the present reaction coordinate, i.e. the bond to be broken, Op-Od, and the bond to be formed, Od-C3'. These figures show a notable feature of the approximate reaction pathway model. Following the initial approach of Od and C3', from $\text{Od-C3}' = 3.0 \text{ \AA}$ to 2.0 \AA , elongation of the Op-Od bond is observed. This elongation occurs prior to the further shortening of the Od-C3' distance (from 2.0 to 1.4 \AA), corresponding to the actual Od-C3' bond formation. The lengthening of the peroxide oxygen-oxygen bond might be rationalized through analysis of the changes in the distribution of the LUMO in the substrate-cofactor complex. Figure 6.6c shows that the LUMO, initially located on the flavin ring, has no density on the atom that actually attacks the substrate, i.e. at the distal oxygen (Od) of the peroxide moiety. The analysis of the LUMO distribution along the reaction pathway shows that this expected LUMO density at the distal oxygen occurs at an intermediate phase in the reaction where $-0.6 < r < -0.2$. Figure 6.6b shows that the most effective lengthening of the peroxide bond occurs at $-0.6 < r < -0.4$. The comparison of the Figures 6.6b and 6.6c suggests that the early lengthening of the peroxide oxygen-oxygen bond in the reaction pathway is related to the increase of LUMO density on Od. The subsequent interaction with the HOMO on the substrate C3' on the distal oxygen may then lead to the formation of the C3'-Od bond. It should be noted that the equilibrium peroxide oxygen-oxygen bond calculated by AM1 is about 0.16 \AA shorter than the bond length obtained at the HF/6-31+G(d) or B3LYP/6-31+G(d) levels and that the orbital contributions predicted by AM1 are slightly different from those predicted at higher levels. Furthermore, the true reaction may involve interactions of the groundstate configuration with other electronic configurations which is not included in the present model. The effect of lengthening the peroxide O-O bond on the LUMO density on Od should therefore be interpreted in a qualitative manner only.

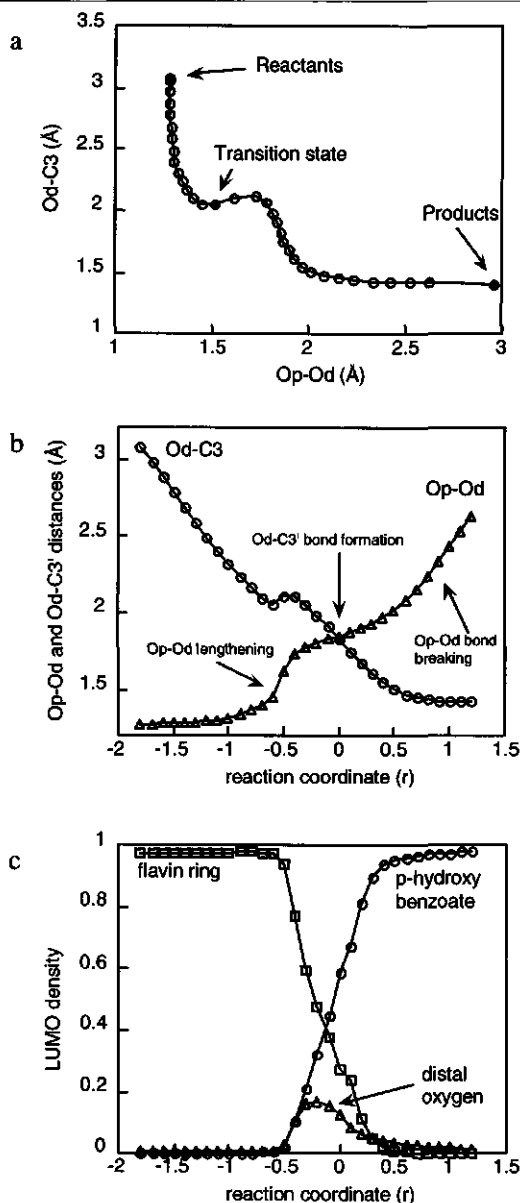


Figure 6.6 a) The Od-C3' distance (forming bond) versus the Op-Od distances (breaking bond). The closed dots correspond to the reactant state (left-top), the transition state and the product state (right-bottom). b) The change in Op-Od and Od-C3' atomic distance as function of the reaction coordinate r . c) The LUMO densities on the substrate, flavin ring and distal oxygen as a function of the reaction coordinate r , obtained by in vacuo single point calculations (AM1) on the geometries of the QM atoms obtained in the QM/MM reaction pathway calculation with dianionic *p*-hydroxybenzoate.

The role of active site amino acids in the hydroxylation

Figure 6.7 presents the results of the energy decomposition analyses which indicate approximately the effects of active site amino acids and water molecules on the hydroxylation step. In Figure 6.7a, the QM/MM energy difference between the reactants and transition state structures of the present reaction pathway is plotted as a function of the amino acid residues included (in order of the distance between their centre of mass COM, and the distal oxygen of hydroperoxyflavin). Similarly, Figure 6.7b presents the QM/MM energy difference between the reactant and the product structures of the present reaction pathway as a function of the residues included.

Some residues with a significant effect on the QM/MM energy differences are labelled. Analysis of the different contributions to the QM/MM energies indicates that these strong effects of individual residues on the QM/MM energy is largely due to electrostatic interactions between the QM and MM atoms. Most of these residues, except for Tyr201, Pro293 and Asn300, are charged residues. For example, Arg214 is involved in binding the carboxylate moiety of the substrate (Schreuder *et al.*, 1989) and Arg44 forms a hydrogen bond with the negatively charged phosphate oxygen of the AMP moiety of the FAD cofactor (Schreuder *et al.*, 1989). All these charged residues influence the electrostatic field in which a net charge transfer of $-1e$ from the substrate to the cofactor occurs along the reaction coordinate. These electrostatic contributions either favour or disfavour the reaction depending on the charge and the position of the residue relative to the reactive (QM) atoms.

Neutral amino acid residues may influence the energy differences, presented in Figure 6.7, through steric or dipole interactions and hydrogen bonds. For example, Figure 6.7a indicates that Tyr201 increases the energy barrier. The role of Tyr201 in the present pathway model, which represents only one particular step in the reaction cycle, is to stabilise the negative charge on the deprotonated hydroxyl moiety of the substrate, via a hydrogen bond, in the C4a-hydroperoxyflavin-substrate complex. As the reaction proceeds this oxyanion formally becomes a carbonyl oxygen which is less charged and therefore the hydrogen bond with the tyrosine residue weakens. This effect is also reflected by the increase in the hydrogen bond distance between Tyr201 and O4' of the substrate along the reaction coordinate (Table 6.2).

As already mentioned on the basis of Table 6.2, the hydrogen bond distance between crystal water 717 and the proximal oxygen of the cofactor decreases along the reaction coordinate. Figure 6.7a and 6.7b indicate a stabilising effect of Wat717 on the transition state and the products relative to the reactants. This is clearly due to the increasing interaction more negatively charged of the peroxide moiety which becomes more negatively charged as the reaction proceeds (Table 6.3).

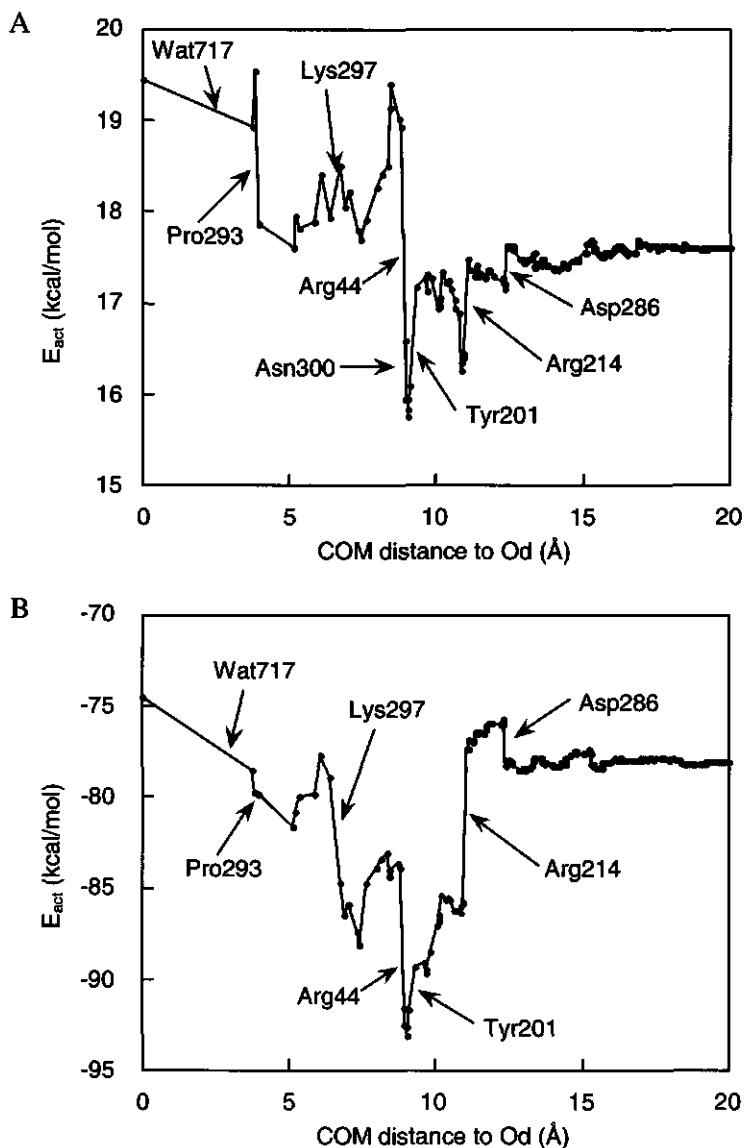


Figure 6.7 Amino acid decomposition analysis illustrating the effect of non-bonded Van der Waals and electrostatic interactions between protein residues and the substrate–flavin complex on the difference in the QM/MM energy term between a) the reactants and the transition state and b) the reactants and products of the present reaction pathway.

Perhaps the most interesting observation by comparison of Figures 6.7a and 6.7b is that Pro293 decreases the energy barrier (Fig. 6.7a), while it has almost no effect on the overall energy difference between reactants and products (Fig. 6.7b). In the

transition state of the reaction pathway model, the backbone carbonyl moiety of Pro293 is in close proximity to the Hd atom of the peroxide moiety (Table 6.2), which is being transferred from the peroxide cofactor to the substrate. In this transition state geometry, the Op-Od bond has been partially broken (Table 6.4, Figure 6.2b) and the total charge on the (somewhat isolated) OH moiety has become more positive (0.11 e, by Mulliken analysis, Table 6.3). The OH moiety appears to be stabilised by the interaction with the partial negative charge of -0.55 on the carbonyl oxygen of Pro293. This favourable electrostatic interaction is not (or less) present in the reactant and product geometries of the present reaction pathway (Table 6.2). Thus, the backbone carbonyl moiety of Pro293 appears to be responsible for a specific stabilisation of the transition state of the hydroxylation process.

It is interesting to note that, despite the fact that the most significant effects of individual residues have an electrostatic character, the total effect of all electrostatic interactions of the reacting QM system with the protein almost cancel out. The other contributions to the QM/MM energy differences, i.e. the QM/MM Van der Waals interactions and the purely MM energy terms account for 1.8 kcal/mol stabilisation of the transition state and 3.6 kcal/mol stabilisation of the products, relative to the reactants.

6.4 Discussion

The hydroxylation step in the reaction cycle of flavin dependent aromatic hydroxylases has been suggested to proceed via an electrophilic aromatic substitution mechanism (Maeda-Yorita and Massey, 1993). The present QM/MM reaction pathway calculations provide a model for this reaction step. Analysis of the charge distribution and the bond lengths along the reaction coordinate of this model for the hydroxylation step are in line with an electrophilic aromatic substitution type of mechanism and support the formation of a hydroxycyclohexadienone as the initial reaction product (Table 6.3 and 6.4). This hydroxycyclohexadienone may represent the so-called intermediate II observed in stopped flow studies with phenol hydroxylase (Maeda-Yorita and Massey, 1993).

Additional results of the reaction pathway analysis provide insight into the role of substrate and cofactor activation in catalysis by PHBH. Deprotonation of the hydroxyl moiety of the substrate has previously been proposed to lead to activation of the C3' atom towards a nucleophilic attack on the cofactor, which is essential for the hydroxylation step (Entsch *et al.*, 1976, Vervoort *et al.*, 1992). The results of the present reaction pathway calculations, performed for both protonated and deprotonated substrate, further strengthen the proposed mechanism for substrate activation by showing that this deprotonation of the hydroxyl moiety leads to a significantly lower

energy barrier for the hydroxylation reaction, corresponding to a $4 \cdot 10^3$ fold increase in the rate of hydroxylation. Furthermore, results obtained with protonated *p*-hydroxybenzoate appear to be comparable with results obtained with 4-fluorobenzoate. 4-Fluorobenzoate is known to bind to the active site and to stimulate formation of the C4a-hydroperoxyflavin intermediate (Spector and Massey, 1972), without subsequently being hydroxylated. This observation has been used as an argument to indicate the importance of substrate activation, which is impossible with the 4-fluorobenzoate lacking the hydroxyl moiety at C4'. The similarity observed between the reaction pathways obtained with the protonated substrate and 4-fluorobenzoate suggests that 4-fluorobenzoate is indeed a good model for an unactivated substrate.

In addition to substrate activation, activation of the C4a-hydroperoxyflavin cofactor for the hydroxylation reaction is of interest. Maeda-Yorita and Massey (Maeda-Yorita and Massey, 1993) and Vervoort et al. (Vervoort *et al.*, 1996) suggested that protonation of the distal oxygen is required for electrophilic reactivity of the peroxide moiety of the flavin cofactor intermediate. However, this protonation by itself appears insufficient to activate the cofactor, since virtually no LUMO density is present on the distal oxygen in the optimised substrate-C4a-hydroperoxyflavin complex. This illustrates that the electronic properties of the C4a-hydroperoxyflavin molecule are not completely characterized by its calculated equilibrium groundstate. The results suggest that an early lengthening of the O-O bond, accompanied by an increase in LUMO density at the protonated distal oxygen, enhances the C-O bond formation and the reaction to proceed. Since the reaction coordinate is an approximation to the intrinsic reaction coordinate and does not provide a dynamic description of the reaction, the question of how this lengthening actually proceeds remains open for discussion. Peroxide oxygen-oxygen bonds are relatively weak (with small force constants for the O-O stretching) (Cremer, 1983) which implies that an oxygen-oxygen lengthening of 0.2-0.3 Å can be expected to occur at low vibrational levels and at physiological temperatures.

It is interesting that the variation in the energy barriers appears to depend on the relative stability of the reactants and products, rather than on specific interactions in the transition state (Figure 6.3). A similar conclusion can be drawn from a previous study showing a linear correlation of calculated differences in heats of formation between the *p*-hydroxybenzoate and the hydroxycyclohexadienone with the natural logarithm of the experimental rate constants for the overall conversion of a series of fluorinated substrates (Vervoort and Rietjens, 1996). The energy differences of -78 kcal/mol between reactants and products, obtained in the present QM/MM calculation for deprotonated substrate, would suggest this step to be very exothermic in the context of the overall reaction cycle. However, since no experimental reaction

enthalpy is available for the individual hydroxylation step to validate this value, it may not be interpreted in an absolute way. The energy decomposition analysis (Figure 6.7b) indicates that the overall effect of the protein on the energy of the reaction may be relatively small and that the largest contribution comes from the QM system only. Comparison of AM1 with higher level calculations on the reactants and products in vacuum (Table 6.1) shows that AM1 overestimates the difference in energy of the reactants and products by about 20%. Furthermore, as also indicated in the note to Table 6.1, the QM energy difference may not be interpreted as enthalpy difference in an absolute way.

An important observation from the amino acid decomposition analysis is that the backbone carbonyl oxygen atom of Pro293 functions as a hydrogen bond acceptor for the hydrogen on the distal oxygen of the hydroperoxyflavin, specifically in the transition state geometry, which results in a lowering of the activation barrier. Proline residues, due to their cyclic structure, have a pronounced influence on backbone conformation. The presence of two neighbouring proline residues at the positions 292 and 293, may indicate that a specific backbone orientation is essential in this part of the protein structure. In relation to the present results it is tempting to speculate that the presence of the proline residues and their specific effect on the backbone orientation contributes to the catalytic effect of the backbone carbonyl of Pro293 on the hydroxylation step.

Further results of the residue analysis indicate that Tyr201 stabilises the substrate in an activated deprotonated state. It should be noticed that the essential role of Tyr201 in catalysis, as experimentally determined by mutating it to a phenylalanine, is to enhance the deprotonation of the enzyme bound substrate (Entsch *et al.*, 1991, Eschrich *et al.*, 1993). This deprotonation process is not included in the present reaction coordinate and, therefore, in the present reaction pathway, the only effect of Tyr201 is an increase of the activation barrier due to the fact that it stabilises the reactant state more than the transition state.

The crystal water 717 has a driving effect on the hydroxylation reaction in the present model. It stabilises the increasing negative charge on the proximal oxygen of the cofactor as the reaction proceeds. At the end of this reaction step, this water would be in a good position to be possibly involved in protonation of this oxygen.

Overall, the enzyme lowers the energy barrier of the calculated reaction pathway for hydroxylation by only 1.8 kcal/mol. This value, obtained by comparing the barrier for the whole system with the barrier for the QM atoms alone (Figure 6.7b), is an approximate and qualitative indication for the catalytic effect of the enzyme on the hydroxylation step in the reaction cycle. Nevertheless, it indicates that the stabilisation of the transition state in the hydroxylation step by PHBH could be relatively small.

Altogether, the present results corroborate the idea that the catalytic mechanism of the hydroxylation step by PHBH is mainly based on activating the substrate and cofactor. More specifically, the role of the enzyme appears to be to provide an environment which enables the coexistence of the activated deprotonated substrate (e.g. through the hydrogen bond of Tyr201) and the protonated C4a-hydroperoxyflavin, in the right orientation. The subsequent hydroxylation reaction, via an electrophilic attack of the cofactor on the substrate, readily proceeds, apparently with a small additional stabilisation of the transition state, e.g. by Pro293.

Acknowledgements

This work was supported by The Netherlands Organization for Scientific Research (NWO). A. J. M. is an EPSRC Advanced Research Fellow, and was previously a Wellcome Trust International Prize Travelling Research Fellow (grant no. 041229). We thank Prof. Martin Karplus for his support and we acknowledge the referees for their useful comments on the manuscript.

7

Modelling flavin- and substrate substituent effects on the activation barrier and rate of oxygen transfer by *p*-hydroxybenzoate hydroxylase

Lars Ridder, Bruce A. Palfeý, Jacques Vervoort and Ivonne M.C.M. Rietjens

Accepted for publication in *FEBS Letters*

The simulation of enzymatic reactions, using computer models, is becoming a powerful tool in the most fundamental challenge in biochemistry: to relate the catalytic activity of enzymes to their structure. In the present study, various computed parameters were correlated with the natural logarithm of experimental rate constants for the hydroxylation of various substrate derivatives catalysed by wild-type PHBH as well as for the hydroxylation of the native substrate (*p*-hydroxybenzoate) by PHBH reconstituted with a series of 8-substituted flavins. The following relative parameters have been calculated and tested: a) energy barriers from combined quantum mechanical/molecular mechanical (AM1/CHARMM) reaction pathway calculations, b) gas-phase reaction enthalpies (AM1) and c) differences between the HOMO and LUMO energies of the isolated substrate and cofactor molecules (AM1 and B3LYP/6-31+G(d)). The gas-phase approaches yielded good correlations, as long as similarly charged species are involved. The QM/MM approach resulted in a good correlation, even including differently charged species. This indicates that the QM/MM model accounts quite well for the solvation effects of the active site surroundings, which vary for differently charged species. The correlations obtained demonstrate quantitative structure activity relationships (QSAR) for an enzyme catalysed reaction including, for the first time, substitutions on both substrate and cofactor.

7.1 Introduction

p-Hydroxybenzoate hydroxylase (PHBH, EC 1.14.13.2) is a flavin dependent monooxygenase catalysing the 3-hydroxylation of 4-hydroxybenzoate. Such hydroxylations are important in the degradation pathway of a wide range of natural abundant compounds (e.g. lignin) as well as industrial aromatic pollutants. The first steps in the reaction cycle involve binding of the substrate, followed by a two electron reduction of the flavin cofactor by NADPH and the incorporation of molecular oxygen to form the C4a-peroxyflavin intermediate. Upon protonation of the distal oxygen of this C4a-peroxyflavin, the C4a-hydroperoxyflavin is formed which performs an electrophilic attack on the substrate (Figure 7.1). The substrate is activated for this electrophilic attack through deprotonation of its C4-hydroxyl moiety (Entsch *et al.*, 1976, Ridder *et al.*, 1999a, Vervoort *et al.*, 1992). The initial products formed are a hydroxy-cyclohexadienone (Entsch *et al.*, 1976, Ridder *et al.*, 1999a), which is converted to 3,4-dihydroxybenzoate through keto-enol tautomerism, and a deprotonated hydroxyflavin, which, after protonation and release of a water molecule, converts to oxidised flavin.

The attack of the C4a-hydroperoxyflavin on the substrate is an essential step in the reaction cycle. In previous studies this step was proposed to be rate limiting on the basis of correlations of the natural logarithm of the rate constant for overall conversion of 4-hydroxybenzoate and four of its fluorinated analogues by *p*-hydroxybenzoate hydroxylase (Husain *et al.*, 1980) with the calculated (gas-phase) nucleophilic reactivity of these substrates (Vervoort *et al.*, 1992), as well as with the calculated activation energies for hydroxylation (Ridder *et al.*, 1998).

In a recent study, the kinetics of PHBH with a series of modified flavin cofactors have been investigated and the results were shown to correlate with Hammett substituent parameters (Ortiz-Maldonado *et al.*, 1999). A drawback of the Hammett constants is that they cannot predict the combined effect of substitutions on both substrate or cofactor. The present study investigates whether several computational approaches to quantify effects of substitutions, used in previous studies on substrate analogues (Ridder *et al.*, 1998, Vervoort and Rietjens, 1996, Vervoort *et al.*, 1992), are able to account for variations in both substrate and cofactor within one theoretical model. A combined quantum mechanical/molecular mechanical (QM/MM) approach (Field *et al.*, 1990, Ridder *et al.*, 1999a, Ridder *et al.*, 1998), including the electrostatic and steric influence of the protein environment on the activation barrier, is compared to gas-phase approaches in which the isolated reacting species are investigated. The results of the various approaches are compared to experimental results, including

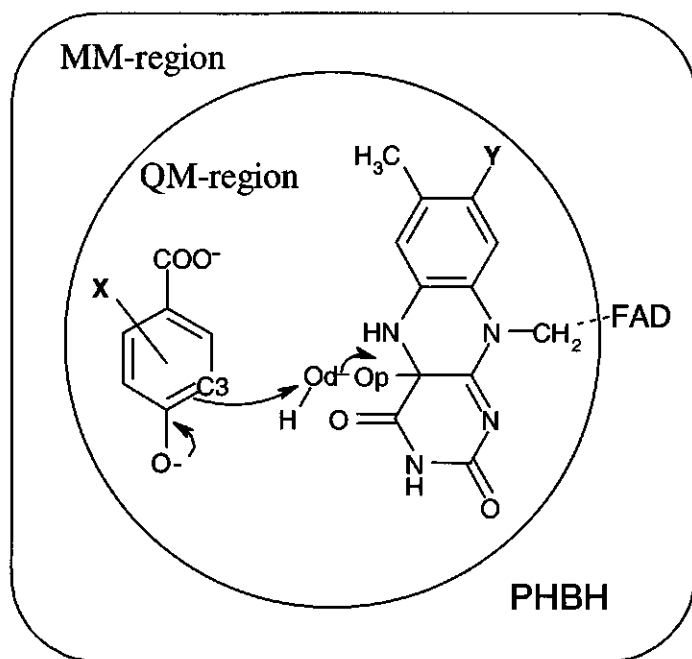


Figure 7.1 Schematic representation of the QM/MM model used to simulate the electrophilic attack of C4a-hydroperoxyflavin on dianionic *p*-hydroxybenzoate, for various the various substrates and cofactors listed in Table 7.1.

experiments with substituted substrates and the above mentioned experiments with substituted flavins (Ortiz-Maldonado *et al.*, 1999).

7.2 Methods

Stopped-flow experiments

2-Fluoro-4-hydroxybenzoate, 3-fluoro-4-hydroxybenzoate, 2,5-difluoro-4-hydroxybenzoate and 3,5-difluoro-4-hydroxybenzoate were synthesised as described previously (Eppink *et al.*, 1997b). 2,3,5,6-Tetrafluoro-4-hydroxybenzoate was purchased from Aldrich. Rate constants for the hydroxylation of the fluorinated substrates by PHBH were obtained by mixing the anaerobic dithionite-reduced enzyme-substrate complex with buffer containing O_2 in a stopped-flow apparatus (Entsch *et al.*, 1991). Reactions were in 50 mM potassium phosphate, pH 6.5, 1 mM EDTA, at 4 degrees. Flavin absorbance or fluorescence changes were analysed as described previously, in order to obtain the rate constant for hydroxylation (Entsch *et al.*, 1991).

Calculation of activation barriers and reaction energies within the protein environment

Approximate activation barriers and reaction energies for the hydroxylation step with a series of substituted substrates and flavin cofactors (Table 7.1) were calculated essentially as described previously (Ridder *et al.*, 1999a). Reaction pathway calculations were performed on the basis of the crystal structure of PHBH in complex with substrate, using a QM/MM potential implemented in CHARMM (Field *et al.*, 1990). The dianionic substrate and the C4a-hydroperoxyflavin cofactor were treated with semiempirical AM1 molecular orbital theory (Dewar and Zebisch, 1988, Dewar *et al.*, 1985), whereas the steric and electronic influence of the surrounding enzyme was included by treating all other atoms molecular mechanically (Brooks *et al.*, 1983). The AM1 parameters for sulphur, provided with CHARMM 24b1 were replaced by the more recently optimised values (Dewar and Yuan, 1990). All 330 crystal waters were included in the model, using the HBUILD routine of CHARMM to add hydrogen atoms. The 8-hydroxy-FAD and the 8-mercapto-FAD have been indicated to be deprotonated (Ortiz-Maldonado *et al.*, 1999) and were treated as such in the QM/MM model.

A reaction coordinate r for the electrophilic attack of the C4a-hydroperoxyflavin on the substrate (Figure 7.1) was defined as the difference between Op-Od interatomic distance and the Od-C3 interatomic distance, i.e. $r = d(\text{Op-Od}) - d(\text{Od-C3})$. To simulate the reaction pathway, this reaction coordinate parameter was changed in small steps (0.1 Å) leading to the splitting of the Op-Od bond and the formation of the Od-C3 bond. All atoms within 6 Å distance of any QM atom were optimised for the reactant, product and all intermediate states of the reaction pathway calculated. This definition of the optimised region around the active site is slightly different from the definition used previously (Ridder *et al.*, 1999a). This modification in the protocol was needed to ensure enough flexibility in the model to properly accommodate the different substituents at the 8-position of the flavin cofactor.

Gas-phase calculations

AM1 calculations were performed on the various reactant and product molecules separately, in gas-phase. The results were compared to the QM/MM reaction pathway calculations to obtain insight into the contributions of the chemical properties of the reactants themselves to the variation in the reaction and activation energies with the different substrates and cofactors within the protein. In order to test the ability of the semiempirical AM1 method to account for the effect of substituents on the reactivity of the substrate and cofactors, the variation in the calculated AM1 energies of the highest occupied molecular orbital (HOMO) of the various *p*-hydroxybenzoates and

Table 7.1 Energy differences between reactant and product complexes $\Delta E_{\text{reaction}}$, and between reactant and transition state complexes ΔE_{act} , of the calculated QM/MM reaction pathways for the various substrates and 8-substituted cofactors. Columns 6-8 present the reaction coordinate values corresponding to the approximate transition states r_{TS} , the gas-phase energy differences between the separate reactant and product molecules and the experimental rate constants for the hydroxylation step (k_{hydrox}). Rate constants for 7-13 were taken from Ortiz-Maldonado *et al.* (1999).

	substrate	8-substituted flavin	$\Delta E_{\text{reaction}}$ (kcal/mol)	ΔE_{act} (kcal/mol)	r_{TS} (Å)	$\Delta E_{\text{gas-phase}}$ (kcal/mol)	k_{hydrox} (s ⁻¹)
1	PHB	-CH ₃	-70.10	17.67	-0.50	-139.97	48
2	2-F-PHB	-CH ₃	-66.88	18.05	-0.50	-134.14	39
3	3-F-PHB	-CH ₃	-65.68	18.57	-0.49	-132.88	39
4	2,5-F ₂ -PHB	-CH ₃	-62.43	18.95	-0.48	-127.17	17.3
5	3,5-F ₂ -PHB	-CH ₃	-61.30	20.71	-0.48	-127.28	2.4
6	2,3,5,6-F ₄ -PHB	-CH ₃	-55.10	22.15	-0.45	-119.48	1.4
7	PHB	-H	-67.43	17.72	-0.50	-140.24	40
8	PHB	-NH ₂	-70.19	17.69	-0.50	-139.24	156
9	PHB	-N(CH ₃) ₂	-70.73	17.84	-0.51	-140.66	34
10	PHB	-SCH ₃	-70.41	18.04	-0.50	-141.24	151
11	PHB	-Cl	-69.18	17.50	-0.49	-143.31	300
12	PHB	-O ⁻	-57.86	21.24	-0.46	-71.619	7.6
13	PHB	-S ⁻	-63.35	19.60	-0.48	-81.050	14.3

the lowest unoccupied molecular orbital (LUMO) of the C4a-hydroperoxyflavin molecules were compared to B3LYP/6-31+G(d)//HF/6-31G(d) results.

In the case of *ab initio* methods, calculations on anions require basis sets including diffuse functions. Semiempirical methods, such as AM1, use minimal basis sets which do not include diffuse functions. Nevertheless, AM1 often performs well on anionic systems and appears to be able to correct for the increased electron repulsion through the parameterisation (Dewar *et al.*, 1985). The competence of AM1 for even a dianionic system is illustrated by a recent comparison of AM1 with *ab initio* methods for dianionic oxaloacetate, for which the AM1 results were similar to those from *ab initio* calculations including diffuse functions and performed even better than *ab initio* methods without diffuse functions (Mulholland and Richards, 1998a).

7.3 Results

Comparison of calculated QM/MM energy changes with experimental rate constants

Continuous reaction pathways were obtained for the various substrates and cofactors listed in Table 7.1. The reaction pathways are comparable with respect to the geometrical changes and the energy profiles (Ridder *et al.*, 1999a). The energy

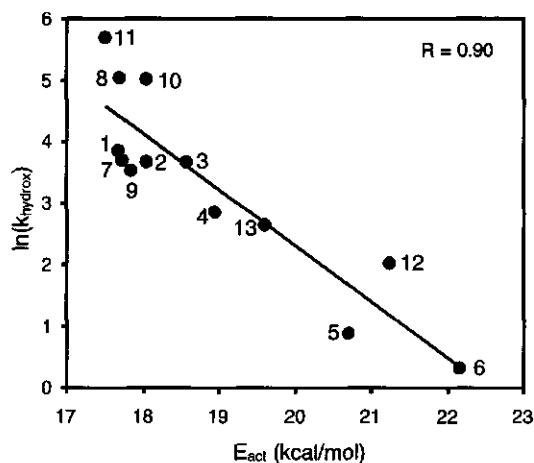


Figure 7.2 Linear correlation between the logarithm of the experimental rate constants for the hydroxylation step (k_{hydrox}) and the energy barriers obtained from the QM/MM reaction pathway calculations with the various substrate and cofactors. Labels according to Table 7.1.

barriers, i.e. the energy differences between the initial energy minimum and the energy maximum of the different profiles, vary between 17.5 and 22.2 kcal/mol (Table 7.1).

Comparison of these calculated QM/MM energy barriers with the natural logarithm of the experimental rate constants for the hydroxylation step shows a good linear correlation with a coefficient of 0.90 (Figure 7.2). In an earlier study a correlation was found between the calculated QM/MM barrier for hydroxylation and the logarithm of the overall k_{cat} for conversion of the substituted substrates only (Ridder *et al.*, 1998). The present results show a similar correlation with the rate constants of the individual hydroxylation step. More importantly, the correlation is now extended to include substitutions on the cofactor as well.

Brønsted relationship between calculated energy barriers and the calculated energy change of the reaction

The approximate transition states were found at slightly different values of the reaction coordinate r , as listed in Table 7.1. Generally the transition state is found at higher (i.e. less negative) values of r as the energy barrier increases and the reaction energy becomes less negative. This suggests that an increase in activation energy corresponds to a more product-like transition state, in line with the Brønsted and Marcus (Marcus, 1968) theories of energy profiles. An excellent linear Brønsted correlation is observed between the calculated QM/MM energy barriers and the calculated QM/MM energy change from reactants to products for all substrate and cofactor combinations studied (Figure 7.3).

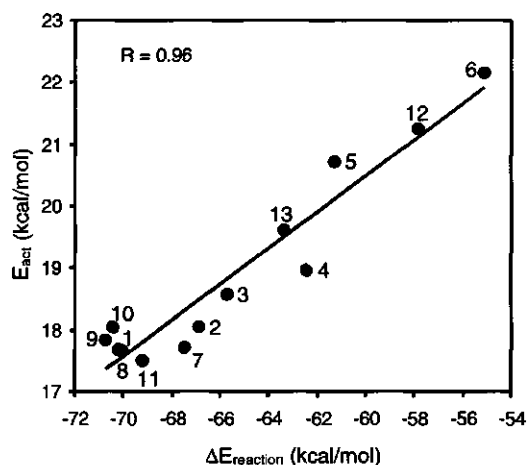


Figure 7.3 Linear Brønsted correlation between the energy barriers (E_{act}) and the reaction energies ($\Delta E_{reaction}$) of the QM/MM reaction pathway calculations. Labels according to Table 1.

The effect of the active site surroundings on the energy of the reaction

In order to estimate the effect of the protein environment on the energetics of the reaction, AM1 calculations were performed to optimise the separate substrate and cofactor molecules in the gas-phase. Figure 7.4 shows that the calculated QM/MM energy differences between the reactant and product states within the active site of PHBH correlate strongly ($r = 0.96$) with the energy differences between the separate reactant and product molecules in gas-phase, except for the deprotonated 8-hydroxy-FAD and the 8-mercapto-FAD. With the deprotonated cofactors, the energy differences between the separate reactant and product molecules in gas-phase are less negative than the gas-phase energy differences obtained for the other substrate-cofactor combinations. This may well be explained by the fact that the C4a-hydroxyflavin product molecule is one atomic charge unit more negative than the C4a-hydroperoxyflavin reactant molecule. In the case of the deprotonated 8-hydroxy-FAD and the 8-mercapto-FAD cofactors, the total charge changes from -1e to -2e upon the electrophilic reaction with the substrate, instead of changing from 0 to -1e as with the other cofactors. The dianionic 8-hydroxy-C4a-hydroxyflavin and the 8-mercapto-C4a-hydroxyflavin product molecules are relatively unstable due to electronic repulsion.

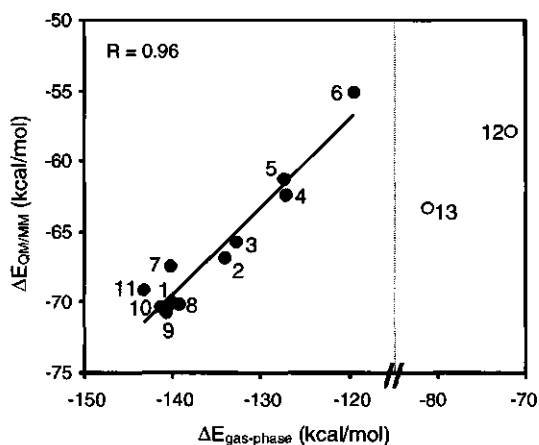


Figure 7.4 Linear correlation between the reaction energies of the QM/MM reaction pathways and the AM1 energy difference between the separate reactants and products in gas-phase. Labels according to Table 7.1. The open circles represent the differently charged hydroxyflavin and mercaptoflavin cofactors which are not included in the linear regression.

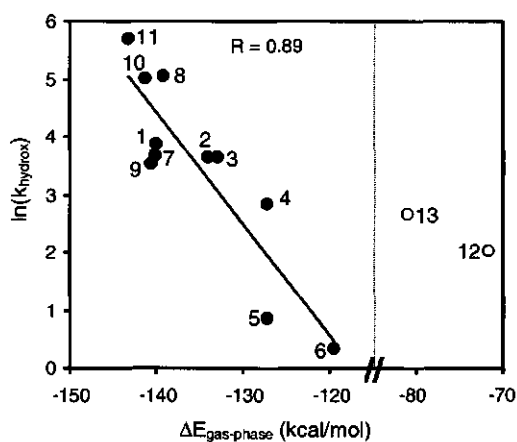


Figure 7.5 Linear correlation between experimental rate constants for the hydroxylation step (k_{hydrox}) and the AM1 energy difference between the separate reactants and products in gas-phase. Labels according to Table 7.1. The open circles represent the differently charged hydroxyflavin and mercaptoflavin cofactors which are not included in the linear regression.

However, this strong effect present in gas-phase does not appear to be representative for the situation within the active site of PHBH. Figure 7.5 shows that the calculated gas-phase energy differences for the reactions with the 8-hydroxy-flavin and the 8-mercapto-flavin cofactors deviate from the Brønsted correlation with the experimental rate constants. Within the actual protein surroundings, the strong effect of the negatively charged substituents on the reaction energy may in fact be reduced by

solvation effects, including specific hydrogen bond interactions, polarisation and dielectric screening. The reaction pathway calculations show that this effect of the protein surrounding is accounted for to a quite reasonable extent in the QM/MM calculations. This can be concluded from the fact that the barriers calculated with 8-hydroxy-FAD and the 8-mercapto-FAD fit well in the correlations shown in Figure 7.2 and 7.3. In the QM/MM models, based on the crystal structure, the 8-substituents are surrounded by a couple of water molecules. In case of 8-hydroxy-FAD and the 8-mercapto-FAD these water molecules clearly interact with (and stabilise) the negatively charged 8-O⁻ and the 8-S⁻ substituents. This could be one of the features of the active site environment reducing the unfavourable influence of the negative substituents on the reaction energy.

Nevertheless, the correlation in Figure 7.4 indicates that, for the similarly charged substrates and cofactors, the variation in the QM/MM calculated energy change of the reaction (and thus the variation in the activation energies (Figure 7.3) and the rate constants for hydroxylation (Figure 7.2)) is dominated by the chemical properties of the substrates and cofactors themselves and not by specific interactions with the PHBH environment. Thus, the calculated energy difference between separate reactant and products molecules in gas-phase is a good parameter for the enzyme activity with different substrates and cofactors (Figure 7.5) as long as they have the same molecular charge.

Prediction of activation energies and rate constants on the basis of the difference between HOMO and LUMO energies in the reactant complex.

In a previous study it was argued that the electrophilic attack of the C4a-hydroperoxyflavin on the substrate could be described in terms of frontier orbital interactions (Ridder *et al.*, 1999a). This implies that the variation in activation energies should relate to the differences in the HOMO and LUMO energies of the substrate and cofactor respectively, in accordance with frontier orbital theory (Fleming, 1976). A high HOMO energy indicates a strong nucleophilic reactivity, whereas a low LUMO energy indicates a strong electrophilic reactivity. Figures 7.6a and 7.6b show a correlation of the natural logarithm of the rate constants for hydroxylation with the energy difference between the HOMO of the substrate molecules and the LUMO of the C4a-hydroperoxyflavin molecules in gas-phase calculated with the AM1 and B3LYP/6-31+G(d)//HF/6-31G(d) methods. The correlations show that the gas-phase HOMO and LUMO energies are indeed useful parameters to quantify the nucleophilic and electrophilic reactivity of substrate and cofactor respectively and to predict variation in activation barriers, as long as similarly charged species are involved. The LUMO energies for the 8-hydroxy- and the 8-mercapto- cofactor molecules are

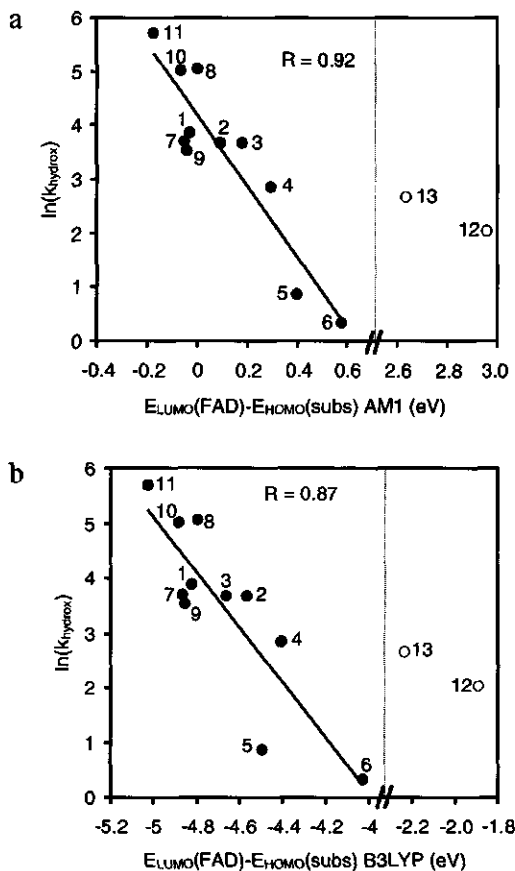


Figure 7.6 Linear correlation between the experimental rate constants for the hydroxylation step (k_{hydrox}) and the difference between the HOMO energy of the isolated substrate and the LUMO energy of the isolated cofactor calculated in gas-phase using a) AM1 and b) B3LYP/6-31+G(d)//HF/6-31G(d). Labels according to Table 7.1. The open circles represent the differently charged hydroxyflavin and mercaptoflavin cofactors which are not included in the linear regression.

significantly higher than the LUMO energies for the other cofactors, indicating that the negative charged substituents on the flavin ring in these cases make the C4a-hydroperoxyflavin less reactive as electrophile. In analogy to the effects described above, the influence of the negative O^- and S^- substituents seems to be smaller within the active site compared to the situation in vacuum, probably due to differential solvation effects for differently charged species.

The similarity between the correlations obtained with the AM1 orbital energies and with the more sophisticated B3LYP calculations indicates that AM1 accounts quite accurately for the effect of the substituents on the nucleophilic and electrophilic

reactivities the *p*-hydroxybenzoate substrate and the C4a-hydroperoxyflavin cofactor respectively.

7.4 Discussion

Computer simulations of enzymatic reactions are becoming increasingly important in understanding the catalytic activity of enzymes on the basis of their structure. Ultimately, it should be possible to predict a rate constant of a given enzymatic reaction on the basis of a 3-dimensional structure of the enzyme and substrate (Warshel, 1992). The present study compares gas-phase models and QM/MM models of an enzymatic reaction, with respect to their ability to explain the variation in the rate of a specific enzyme catalysed reaction step, with different cofactors and substrates. For the model enzyme of the present study, *p*-hydroxybenzoate hydroxylase, the structure activity relationships described so far, either involve only structural variations in the substrate, e.g. a series of fluorinated *p*-hydroxybenzoate homologues (Husain *et al.*, 1980, Ridder *et al.*, 1998, Vervoort *et al.*, 1992), or include only structural variations in the flavin cofactor (Ortiz-Maldonado *et al.*, 1999). To our knowledge the present study demonstrates for the first time a quantitative structure activity relationship (QSAR) for an enzyme catalysed reaction including substitutions on both substrate and cofactor.

It is demonstrated that the reactivities of cofactor and substrate can be predicted on the basis of several parameters. First, the HOMO and LUMO energies calculated for the substrates and cofactors respectively, both at the AM1 and B3LYP/6-31+G(*d*) levels, are shown to be useful parameters for their reactivity (Figures 7.6a and 7.6b). Second, the calculated energy differences between the separate reactant and product molecules, using AM1, appear to be a useful parameter as well (Figure 7.5). However, these two parameters are valid only for similarly charged species, as illustrated by deviations from the correlations (Figures 7.4 to 7.6) for the negatively charged 8-hydroxy-flavin and the 8-mercapto-flavin cofactors. The effect of replacement of a neutral group by a charged substituent on molecular properties in a solvent or protein environment often can not be correctly predicted by parameters derived from gas-phase calculations, due to differential solvation effects on differently charged compounds (Ridder *et al.*, 1999b).

A powerful feature of the QM/MM method is that it includes to a certain extent the solvation effects on the reacting species within the active site of the enzyme. It is promising to observe that, without further calibration of the standard parameters involving the interaction of the MM solvent molecules with negatively charged oxygen and sulfur atoms in the QM system, this differential solvation effect is modelled reasonably accurate in the present QM/MM model.

Acknowledgements

This study was especially dedicated to the event of awarding the title "scientist of the year 1998" at Ann Arbor Michigan University to prof. dr. Vincent Massey. We thank dr. Adrian J. Mulholland from the University of Bristol for his support. This work was funded by the Netherlands Organisation for Scientific Research (NWO).

8

A quantum mechanical/molecular mechanical study of the hydroxylation of phenol and halogenated derivatives by phenol hydroxylase

Lars Ridder, Adrian J. Mulholland, Ivonne M.C.M. Rietjens and Jacques Vervoort

Accepted for publication in *The Journal of the American Chemical Society*

A combined quantum mechanical and molecular mechanical (QM/MM) method (AM1/CHARMM) was used to investigate the mechanism of the aromatic hydroxylation of phenol by a flavin dependent phenol hydroxylase (PH), an essential reaction in the degradation of a wide range of aromatic compounds. The model for the reactive flavin intermediate (C4a-hydroperoxyflavin) bound to PH was constructed on the basis of the crystal structure of the enzyme-substrate complex. A potential energy surface (PES) was calculated as a function of the reaction coordinates for hydroxylation of phenol (on C6) and for proton transfer from phenol (O1) to an active site base Asp54 (OD1). The results support a reaction mechanism in which phenol is activated through deprotonation by Asp54, after which the phenolate is hydroxylated through an electrophilic aromatic substitution. *Ab initio* test calculations were performed to verify these results of the QM/MM model. Furthermore, the variation in the calculated QM/MM activation energies for hydroxylation of a series of substrate derivatives was shown to correlate very well ($R = 0.98$) with the natural logarithm of the experimental rate constants for their overall conversion by PH (25°C, pH 7.6). This correlation validates the present QM/MM model and supports the proposal of an electrophilic aromatic substitution mechanism in which the electrophilic attack of the C4a-hydroperoxyflavin cofactor on the activated (deprotonated) substrate is the rate limiting step at 25°C and pH 7.6. The correlation demonstrates the potential of the QM/MM technique for predictions of catalytic activity on the basis of protein

structure. Analysis of the residue contributions identified a catalytic role for the backbone carbonyl of a conserved proline residue, Pro364, in specific stabilization of the transition state for hydroxylation. A crystal water appears to assist in the hydroxylation reaction by stabilizing the deprotonated C4a-hydroxyflavin product. Comparison of the present results with previous QM/MM results for the related *p*-hydroxybenzoate hydroxylase (Ridder *et al.* (1998) *J. Am. Chem. Soc.* 120, 7641-7642) identified common mechanistic features, providing detailed insight into the relationship between these enzymes.

8.1 Introduction

Flavin-dependent phenol hydroxylase (PH) catalyzes the hydroxylation of phenol, specifically at the *ortho* position. This is an important step in the microbial degradation pathway of aromatic compounds including lignin, the main constituent of wood and therefore one of the most abundant natural compounds. The PH reaction leads to the formation of catechol, the aromatic ring of which can be cleaved subsequently by catechol dioxygenases. In addition to the parent substrate, various substituted phenols are hydroxylated by the enzyme, including fluorophenols, chlorophenols, aminophenols, nitrophenols, dihydroxybenzenes and cresol (Gaal and Neujahr, 1979, Neujahr and Gaal, 1973, Peelen *et al.*, 1993). As a result, the reactions catalyzed by PH are also important for degradation and detoxification of a large group of aromatic pollutants of industrial origin.

Transient kinetic studies on phenol hydroxylase (Maeda-Yorita and Massey, 1993) support a catalytic mechanism which proceeds via a C4a-hydroperoxyflavin intermediate, formed by reduction of the FAD cofactor by NADPH, subsequent incorporation of molecular oxygen and protonation of the resulting peroxide moiety. The C4a-hydroperoxyflavin is proposed to react with the phenolic substrate via an electrophilic aromatic substitution mechanism, which results in the formation of a cyclohexadienone as the initial reaction product. The cyclohexadienone is converted non-enzymatically into the catechol product via keto-enol tautomerisation.

Theoretical analysis on the basis of the 3D structure of enzymes is becoming increasingly important in providing additional information on their reaction mechanisms (Bash *et al.*, 1991, Harrison *et al.*, 1997, Mulholland and Richards, 1997, Ridder *et al.*, 1998, Warshel and Levitt, 1976). A powerful approach to simulate reactions within large molecular systems is to combine quantum mechanical (QM) and molecular mechanical (MM) methods. The combined QM/MM approach allows the part of the system directly involved in the chemical reaction to be described by quantum mechanics, while the environment is described by (computationally cheaper) molecular mechanics. In previous studies (Ridder *et al.*, 1999a, Ridder *et al.*, 1998), a

QM/MM approach was used to study the aromatic hydroxylation step catalysed by *p*-hydroxybenzoate hydroxylase (PHBH), which is both structurally and functionally related to PH (Enroth *et al.*, 1998, Maeda-Yorita and Massey, 1993). This QM/MM simulation provided detailed insight into the electrophilic attack of the C4a-hydroperoxyflavin on the substrate, catalyzed by PHBH, and the importance of substrate activation through its deprotonation.

In the present study, the equivalent step in the reaction cycle of phenol hydroxylase is investigated by a similar QM/MM approach, which was shown to be valid and useful for this type of reaction in our previous study (Ridder *et al.*, 1998). The first objective was to study the reaction mechanism of hydroxylation in PH in atomic detail, in order to identify some important catalytic features. A comparison of the results obtained for PH with the results of our earlier studies on the homologous step in PHBH, provides detailed insight into some remarkable similarities as well as some intriguing differences in the structure and function of these two related enzymes. The second objective was to investigate whether the energy barriers calculated for the hydroxylation step by PH could be related quantitatively to experimental overall rate constants for the conversion of a series of phenol derivatives (Peelen *et al.*, 1995). Such correlations are important in order to assess the potential of QM/MM simulations for the prediction of the catalytic properties of enzymes with respect to new substrates (Ridder *et al.*, 1998).

An unanswered question about the mechanism of hydroxylation in PH is how substrate is activated towards reaction with the hydroperoxyflavin cofactor, the key step in the hydroxylation process. For PHBH, a number of studies (Entsch *et al.*, 1976, Entsch *et al.*, 1991, Eschrich *et al.*, 1993) provide evidence for a mechanism in which the hydroxyl moiety of the substrate becomes deprotonated upon binding, which activates the substrate for an electrophilic attack by the hydroperoxyflavin (Vervoort *et al.*, 1992). The previous simulation of this reaction step indicated that deprotonation of the substrate does indeed significantly lower the activation energy for hydroxylation (Ridder *et al.*, 1999a, Ridder *et al.*, 1998). It is not clear whether a similar deprotonation of substrate occurs in phenol hydroxylase. NMR experiments indicate that the substrate binds to the oxidized PH enzyme in its neutral protonated state (Peelen *et al.*, 1993). However, in analogy to results previously obtained for PHBH (Ridder *et al.*, 1999a), the protonation state of the phenol substrate in the active site is expected to influence its susceptibility to electrophilic attack by the C4a-hydroperoxyflavin. The crystal structure of PH indicates the presence of a hydrogen bond between the hydroxyl moiety of the substrate and a potential active site base, Asp54. Therefore, the possibility of proton transfer from the phenol substrate molecule to Asp54 was included in the present simulation.

8.2 Methods

Construction of the protein model for the C4a-hydroperoxyflavin intermediate of phenol hydroxylase

The crystal structure of PH from *Trichosporon cutaneum* in complex with substrate (Enroth *et al.*, 1998) (PDB entry code 1FOH) contains two homodimers in the oxidized state. In each dimer, two different conformations of the flavin are present in the two monomers: the so-called “in” (chains C and D) and “out” (chains A and B) conformations. In analogy to the conclusions of a number of studies on PHBH (Gatti *et al.*, 1994, Schreuder *et al.*, 1994) the “in” conformation is expected to represent the conformation in which substrate hydroxylation can occur. Therefore, one of the subunits in the crystal structure in which the flavin is in the “in” conformation (chain C) was chosen as the starting point for building a model for the C4a-hydroperoxyflavin intermediate of phenol hydroxylase in complex with substrate. The model consists of all protein atoms of the monomer, in which the active site is completely buried, being inaccessible to solvent (Enroth *et al.*, 1998). The flavin cofactor, present in the crystal structure in its oxidized state, was modified to the reactive C4a-hydroperoxyflavin intermediate by superimposing coordinates of an experimentally derived model of the C4a-hydroperoxyflavin (Bolognesi *et al.*, 1978, Schreuder *et al.*, 1990, Vervoort *et al.*, 1986), using the algorithm of Kabsch (1976) implemented by Schreuder *et al.* (1990). Forty crystallographic waters within a 16Å sphere around the distal oxygen of the C4a-hydroperoxyflavin were included. Hydrogen atoms on these crystal water molecules were built using the HBUILD routine in the CHARMM molecular mechanics package version 24b1 (Brooks *et al.*, 1983).

Several ionizable side chains are present in the active site of phenol hydroxylase. A correct assignment of charges on ionizable side chains is of importance for the QM/MM calculations described below, since the MM point charges on the active site residues influence the electron density distribution in the QM system. Information about protonation states of ionizable side chains is not directly provided by the crystal structure. Histidine side chains are a particular problem as their pK_a s can have values below and above the physiological pH, dependent on the specific protein environment. The protonation states of four active site histidines in PH, His116, His189, His 360 and His362, were selected essentially as described by Mulholland and Richards (1997). The various possible protonation states were tested by applying (QM/MM) geometry optimization as described in the next section. All four histidines remained closest to their crystal coordinates when treated in their neutral states with a proton on NE1. These neutral states were therefore used in further calculations. Likewise, when

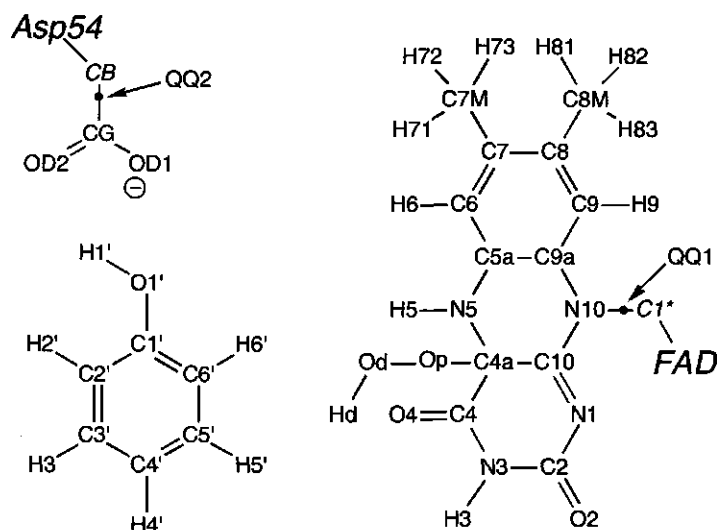


Figure 8.1 The QM region of the present model for the C4a-hydroperoxyflavin intermediate in the PH reaction cycle. MM atoms are labeled in italics. Bonds crossing the QM/MM boundary are replaced by bonds to link atoms (Field *et al.*, 1990) (OO1 and OO2) in the QM system.

Lys365 was treated as protonated the structure became altered due to a hydrogen bond interaction (not present in the crystal structure) between the charged Lys365 sidechain and the backbone carbonyl of Pro364. Lys365 was therefore modeled in its neutral state in which it kept its crystal structure conformation (being a good hydrogen bond acceptor for Tyr336).

An essential aspect of the model is the protonation state of phenol and of two residues with which the hydroxyl group of phenol appears to form hydrogen bonds, Asp54 and Tyr289. It is expected that these hydrogen bonds in the crystal structure of the oxidized state are also present during the hydroxylation step. Furthermore, the presence of the two hydrogen bonds would imply that (only) one of the three ionizable groups could be deprotonated. Since Asp54 is potentially the most acidic of the three, the electronic state with a deprotonated Asp54 (Figure 8.1) is chosen as the initial model. However, deprotonation of phenol would be expected to activate this substrate for the electrophilic attack by the C4a-hydroperoxyflavin. Therefore the possibility of proton transfer from phenol to Asp54 was included in the subsequent simulations. For this reason, part of the Asp54 side chain was included in the quantum mechanical region of the model as described in the next section. The situation in which Tyr289 is deprotonated is considered unlikely to be of physiological relevance and was therefore not investigated.

Application of the AM1/CHARMm 22 QM/MM potential

The flavin ring, the substrate and the side chain of Asp54 (a total of 49 atoms) were treated quantum mechanically (Figure 8.1) with the semiempirical (closed shell) AM1 Hamiltonian (Dewar and Zebisch, 1988, Dewar *et al.*, 1985). All other atoms of the complete subunit (6628) were treated at the molecular mechanical level, using the CHARMm 22 'united atom' forcefield (polar hydrogens only) (Quanta, 1993). Two types of non-bonded interactions between the QM and MM atoms are accounted for. Classical VDW terms are used to include steric effects between the QM and MM atoms, whereas the electronic interactions are accounted for by including the MM point charges (as atomic 'cores') in the Hamiltonian for the QM system (Field *et al.*, 1990). The covalent bonds, which cross the boundary between the QM and MM regions, were treated by introducing so-called link atoms, which are included in the QM system as hydrogen atoms (Field *et al.*, 1990). The same QM/MM potential has been applied successfully in previous studies (Mulholland and Richards, 1997, Ridder *et al.*, 1999a, Ridder *et al.*, 1998). The QM region has a total charge of -1e. A dielectric constant of 1.0 was used for all MM and QM/MM electrostatic interactions. An approximate correction for dielectric screening in the MM and QM/MM electrostatic interactions was applied through the use of a 16Å non-bonded cut-off. The electrostatic interactions were smoothly scaled down to zero over the 11Å to 16Å distance range, using a group based switching function (Brooks *et al.*, 1983). All QM atoms were treated as one group to ensure that the MM charges were treated consistently in the Hamiltonian for the QM atoms, i.e. all QM atoms 'feel' the same MM charges.

Throughout the calculations, all atoms within a 16Å active site region around the distal oxygen of the C4a-hydroperoxyflavin (Od) were optimized. Atoms outside this active site region were fixed. All non-hydrogen atoms within a 14Å to 16Å buffer-zone were harmonically restrained to their crystal coordinates with force constants based on model average *B*-factors (Brooks and Karplus, 1989) and scaled from zero at 14Å to maximum at 16Å away from the centre of the active site region (Od).

Definition of the reaction coordinates for hydroxylation and proton transfer to Asp54

The two reaction coordinates of interest are the electrophilic attack of the distal oxygen Od of the C4a-hydroperoxyflavin on the substrate and the proton transfer from O1' of the substrate to OD1 of Asp54. In the PH structure, the position of one of the two carbon centers (C6' in the crystal structure) ortho with respect to the hydroxyl moiety is much closer to the distal oxygen of the flavin peroxide than the other (C2' in the crystal structure). Therefore, only the electrophilic attack of the distal oxygen of the cofactor on the C6' position of the substrate was assumed to be of physiological

relevance. The reaction coordinate for this step (r_{OH}) was defined as $r_{\text{OH}} = d(\text{Op-Od}) - d(\text{Od-C6'})$. This definition of the reaction coordinate is identical to the definition successfully used in a previous simulation of the equivalent reaction step in PHBH (Ridder *et al.*, 1999a). The reaction coordinate (r_{H}) for proton transfer from the O1' of phenol to OD1 of Asp54 was defined as $r_{\text{H}} = d(\text{O1'-H1'}) - d(\text{H1'-OD1})$. The reaction coordinate restraints were applied using the CHARMM RESDistance command (Eurenius *et al.*, 1996).

Initial optimization and solvation of the simulation system

The QM/MM model was initially optimized with the two reaction coordinates of interest harmonically restrained ($k = 5000 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$) to their midpoint values, i.e. $r_{\text{OH}} = r_{\text{H}} = 0$, which is equivalent to $d(\text{Op-Od}) = d(\text{Od-C6'})$ and $d(\text{O1'-H1'}) = d(\text{H1'-OD1})$. This required 1281 steps of ABNR minimization (gradient $< 0.01 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$). The structure thus obtained represents a midpoint of the potential energy surface to be calculated as described in the next section. This midpoint structure was first solvated as follows. 15 Water molecules were added by superimposing an equilibrated box of TIP3P waters and removing all waters further than 16 \AA away from the centre of the active site region or having an oxygen atom within 2.6 \AA of another non-hydrogen atom. To equilibrate the solvent (including crystal waters) within the protein, 26 ps of stochastic boundary molecular dynamics (SBMD) (Brooks and Karplus, 1989) were performed at 300K with the protein atoms fixed (at their optimized positions) and with MM point charges on the QM atoms derived from a fit to the (AM1) electrostatic potential around the isolated QM atoms (Spartan 5.1). Cooling from 300K to 0K was performed during another 6 ps of SBMD. The solvation procedure (superimposing an equilibrated box of TIP3P waters and removing all waters further than 16 \AA away from the centre of the active site region or having an oxygen atom within 2.6 \AA of another non-hydrogen atom) was then repeated, resulting in the addition of three more water molecules. In a next step, 250 steps of ABNR minimization were applied to optimize the positions of the water molecules within the (frozen) protein.

This solvated system was again fully optimized (gradient $< 0.01 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$, requiring 1046 steps of ABNR minimization) with all the active site atoms free to move, using the boundary and reaction coordinate restraints and the QM/MM potential as described above. The complete model consisted of substrate, the C4a-hydroperoxide form of FAD, all 656 amino acid residues, 40 crystal water molecules and 18 newly built water molecules.

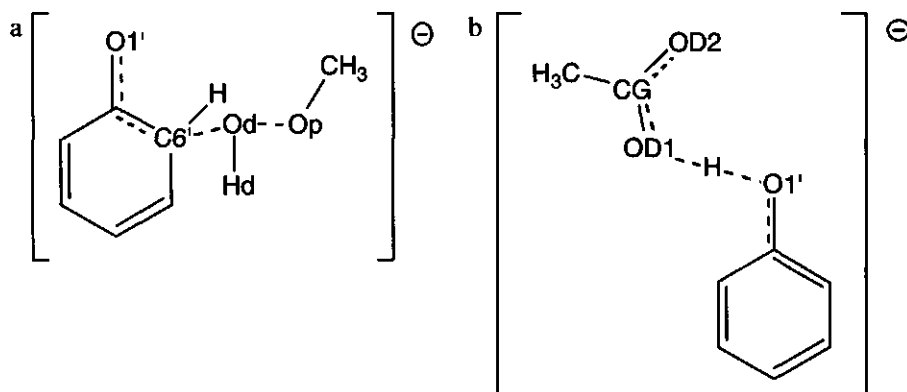


Figure 8.2 Transition state models a) for the hydroxylation reaction in which the C4a-hydroperoxyflavin cofactor is replaced by methylperoxide, and b) for proton transfer from phenol to Asp54.

Calculation of a potential energy surface

Starting from this solvated QM/MM model a potential energy surface was calculated on a grid by restraining the two reaction coordinates. The reaction coordinate for hydroxylation (r_{OH}) was harmonically restrained with a force constant of $k = 5000 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$, and was varied between -2.0 \AA and $+2.0 \text{ \AA}$ in steps of 0.2 \AA . The reaction coordinate for proton transfer (r_H) was harmonically restrained, using the same force constant as used for restraining the hydroxylation reaction, and was varied between -1.0 \AA and $+1.4 \text{ \AA}$. All water atoms were harmonically restrained to their initial optimized positions, using a mass weighted force constant of $0.1 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ to prevent discontinuous changes in the solvent configuration. At each gridpoint, up to 303 steps of ABNR minimization were performed, starting from the geometry of a neighboring grid point, until the gradient became less than $0.02 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$. The approximate stationary points on the surface, i.e. minima (corresponding to intermediates) and saddle points (corresponding to approximate transition states), were determined more precisely by performing additional geometry optimizations with only one (for saddle points) or neither (for minima) of the reaction coordinates restrained. The reaction coordinate values that correspond to the approximate transition states $r(TS)$ were determined to 0.01 \AA precision.

Validation of the AM1 method

The present QM/MM model treats the substrate, the flavin ring and the carboxylate moiety of the active site base, Asp54 quantum mechanically (Figure 8.1), using the semiempirical AM1 method. To test the adequacy of this semiempirical method for the

reactions of interest, its performance for small model systems was compared to higher level *ab initio* methods. For the hydroxylation reaction a model system was investigated in which the C4a-hydroperoxyflavin is replaced by methyl peroxide (Figure 8.2a). The second reaction of interest, proton transfer from phenol to Asp54, was tested with a small model system consisting of phenol and acetic acid (Figure 8.2b). Reactants, transition states and products were optimized in the gas-phase using Gaussian98 (Frisch *et al.*, 1998) and Spartan 5.1. The AM1 results are compared to results from *ab initio* HF and MP2 calculations and density functional B3LYP calculations using the 6-31+G(*d*) basis set.

Amino acid decomposition analysis

In order to identify important active site residues, an energy decomposition analysis was performed, with a procedure similar to those used previously (Bash *et al.*, 1991, Cunningham *et al.*, 1997, Lyne *et al.*, 1995, Mulholland and Richards, 1997, Ridder *et al.*, 1999a). The contributions of individual MM residues to the total QM/MM energy barriers, and to the energy differences between the reactant and product states, were determined for the proton transfer as well as the hydroxylation reaction. The analyses start from the energy difference between the reactant and the transition states, or the reactant and the product states, without the MM protein environment included. Then, one by one, the amino acid residues are included in the energy difference calculation in order of increasing distance between their centre of mass (COM) and the distal oxygen of the hydroperoxyflavin. The effect of an (MM) residue on the total QM/MM energy difference gives an approximate and qualitative indication of its influence on the energy of the reaction in the active site of PH.

Calculation of energy barriers for substituted phenols

In addition to the calculations with the native substrate, relevant points on the potential energy surface, i.e. the phenolate and transition state complexes, were calculated for a series of 15 phenol derivatives with fluoro- and chloro-substituents, which are known to be substrates for PH (Peelen *et al.*, 1995). These calculations started from the relevant structures obtained for phenol, in which phenol was replaced by the substituted substrates. QM/MM energy minimization was applied using the same geometry restraints as used for phenol. For the transition state, the reaction coordinate r_{OH} was varied to find the approximate transition state to 0.01 Å accuracy with respect to r_{OH} (similar to the procedure with the native substrate). The energy barriers were calculated as the energy difference between the reactant and transition state structures thus obtained. For the asymmetrically substituted substrates these calculations were performed for the two possible orientations.

Correlation of calculated energy barriers with experimental rate constants

The calculated QM/MM energy barriers were compared with experimental rate constants for enzymatic conversion of the different phenols by PH (Peelen *et al.*, 1995)*. Based on the Arrhenius equation, a linear correlation of the natural logarithm of the experimental rate constants with the calculated barriers is expected, if the hydroxylation step is indeed rate limiting:

$$k_{cat} = Ae^{-E_{act}/RT} \leftrightarrow \ln k_{cat} \sim -E_{act}$$

In case of the asymmetric substrates two different orientations of the substrate, and therefore two different reactions are relevant. For these substrates, the average values of the two calculated energy barriers were used in the correlation with the logarithm of the experimental rate constants for overall conversion (Peelen *et al.*, 1995). A more sophisticated approach to calculate apparent energy barriers, based on the calculated barriers for the two orientations and on experimental product ratios, is given in the Supporting Information.

8.3 Results

Potential energy surface

Figure 8.3 presents the QM/MM potential energy surface as a function of both reaction coordinates, obtained from the grid of geometry optimizations. From the potential energy surface six approximate stationary points were identified and determined more precisely by performing additional geometry optimizations: the phenol complex, the transition state for proton transfer (TS_{proton}), the phenolate complex, the transition states for hydroxylation of phenol (TS_{phenol}) and phenolate ($TS_{phenolate}$) and the cyclohexadienone product complex (Figure 8.4).

The product complex, in which H1' is on Asp54, is the most stable state on the calculated energy surface. This indicates that the proton H1' does indeed shift from phenol to Asp54 somewhere in the overall reaction. It is apparent that the potential energy surface does not support a concerted mechanism, i.e. a mechanism in which hydroxylation and proton transfer occur at the same time. This would be the case if a single transition state (e.g. a saddle point in the center region of the potential energy surface in Figure 8.3) connected the phenol-C4a-hydroperoxyflavin complex to the deprotonated cyclohexadienone-C4a-hydroxyflavin complex. Instead, two stepwise reaction pathways are possible. One possible reaction pathway is the hydroxylation of

* Four of the substrate derivatives investigated by Peelen *et al.* were not included in the present study. The $K_M(\text{NADPH})$ values with these substrates appeared to be significantly higher than with the other substrates which makes the k_{cat} values derived from these experiments (with $[\text{NADPH}] < K_M(\text{NADPH})$) inaccurate.

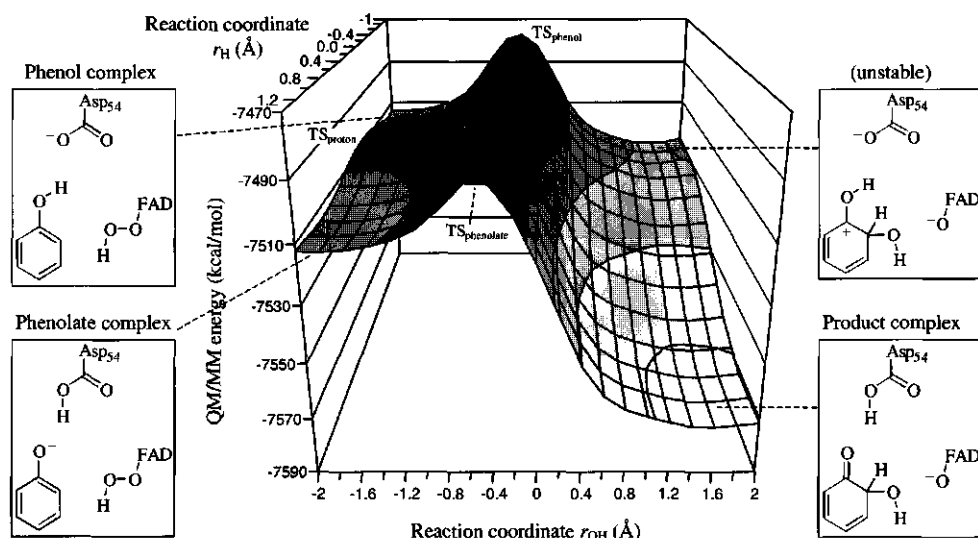


Figure 8.3 QM/MM potential energy surface as a function of the two reaction coordinates $r_{OH} = d(Op-Od) - d(OD-C6')$ and $r_H = d(O1-H1) - d(H1-OD1_{Asp54})$, obtained by interpolation of the energies of all optimized intermediate structures (using Microsoft Excel 97).

the neutral phenol, directly followed by transfer of a proton from the product to Asp54. In the alternative reaction pathway, first the proton H1' from the hydroxyl moiety of phenol is transferred to Asp54, after which hydroxylation of the phenolate takes place. Table 8.1 presents the energies, relative to the phenol complex, and some important atomic distances in the various optimized structures. It can be seen from Table 8.1 that direct hydroxylation of phenol corresponds to an energy barrier of about 36 kcal/mol. It is apparent from Figure 8.3 that the direct hydroxylation of phenol itself does not result in an intermediate (i.e. a minimum) with the H1' still on O1' (represented by the scheme labeled 'unstable' in Figure 8.3). Rather, the positively charged cyclohexadienol is very unstable and the H1' proton would spontaneously shift to Asp54, without crossing an energy barrier. The alternative reaction pathway first encounters an energy barrier for proton transfer of only 6 kcal/mol after which the barrier for hydroxylation is 24 kcal/mol. The hydroxylation of phenolate (after the H1' proton has been transferred to Asp54) does result in a stable product. Analysis of the bond lengths (Table 8.1) and the Mulliken charge distribution (results in the Supporting Information) in the QM system for the various intermediate structures indicates that this product of the hydroxylation of phenolate is a cyclohexadienone. This follows, first of all, from a decrease in the C2'-C3', C4'-C5' and C1'-O1' bond distances and a lengthening of the other substrate C-C bonds (Table 8.1, columns 5-7).

Second, the significant decrease in the negative charge on O1' (from -0.60e to -0.35e) indicates its transition from a deprotonated hydroxyl oxygen in the phenolate complex to a carbonyl oxygen in the product complex. In total, -1e charge is transferred from the substrate to the flavin cofactor. The resulting formal charge on the proximal oxygen is delocalized over the flavin ring. Altogether the results for the hydroxylation of phenolate are in agreement with an electrophilic aromatic substitution type of mechanism.

Table 8.1 The relative energy (kcal/mol) of the various intermediates and approximate transition states on the calculated potential energy surface for the QM/MM model, and a selection of interatomic distances (Å) between QM atoms and between QM and MM atoms. Energies are given relative to the energy of the phenol complex. The distances subject to reaction coordinate restraints are shown in bold face. Atom labeling as in Figure 8.1.

	Phenol complex	TS _{phenol}	TS _{proton}	Phenolate complex	TS _{phenolate}	Product complex
energy	0	36.17	6.0	-0.48	23.22	-59.40
C4a-Op	1.48	1.38	1.48	1.48	1.40	1.31
Op - Od	1.28	1.76	1.28	1.28	1.63	2.94
Od - C6'	3.11	1.99	3.08	3.10	2.13	1.41
C6' - C1'	1.41	1.42	1.42	1.43	1.44	1.53
C1' - C2'	1.41	1.42	1.42	1.43	1.45	1.47
C2' - C3'	1.39	1.38	1.38	1.38	1.37	1.35
C3' - C4'	1.40	1.41	1.40	1.40	1.42	1.45
C4' - C5'	1.40	1.38	1.40	1.40	1.38	1.34
C5' - C6'	1.39	1.42	1.39	1.38	1.41	1.49
C1' - O1'	1.36	1.34	1.32	1.29	1.27	1.24
O1' - H1'	0.99	1.00	1.22	1.92	1.98	2.14
H1' - OD1	1.87	1.83	1.19	0.99	0.99	0.98
CG - OD1	1.27	1.27	1.30	1.34	1.34	1.35
CG - OD2	1.27	1.27	1.25	1.24	1.24	1.23
Tyr289-H - O1'	1.88	1.85	1.87	1.80	1.82	2.06
Pro364-O - Hd	2.60	1.99	2.79	3.07	2.14	2.89
Wat168-H - Op	2.10	2.09	2.11	2.12	2.09	2.01
Asp54-H - O4	2.23	2.17	2.10	2.04	1.99	1.99
Gly55-H - O4	2.44	2.19	2.32	2.31	2.16	2.13
Met370-H - N1	2.16	2.16	2.15	2.17	2.18	2.12
Asn371-H - O2	2.77	2.71	2.75	2.82	2.80	2.65
Asn371-HD - O2	2.25	2.11	2.22	2.25	2.13	2.08

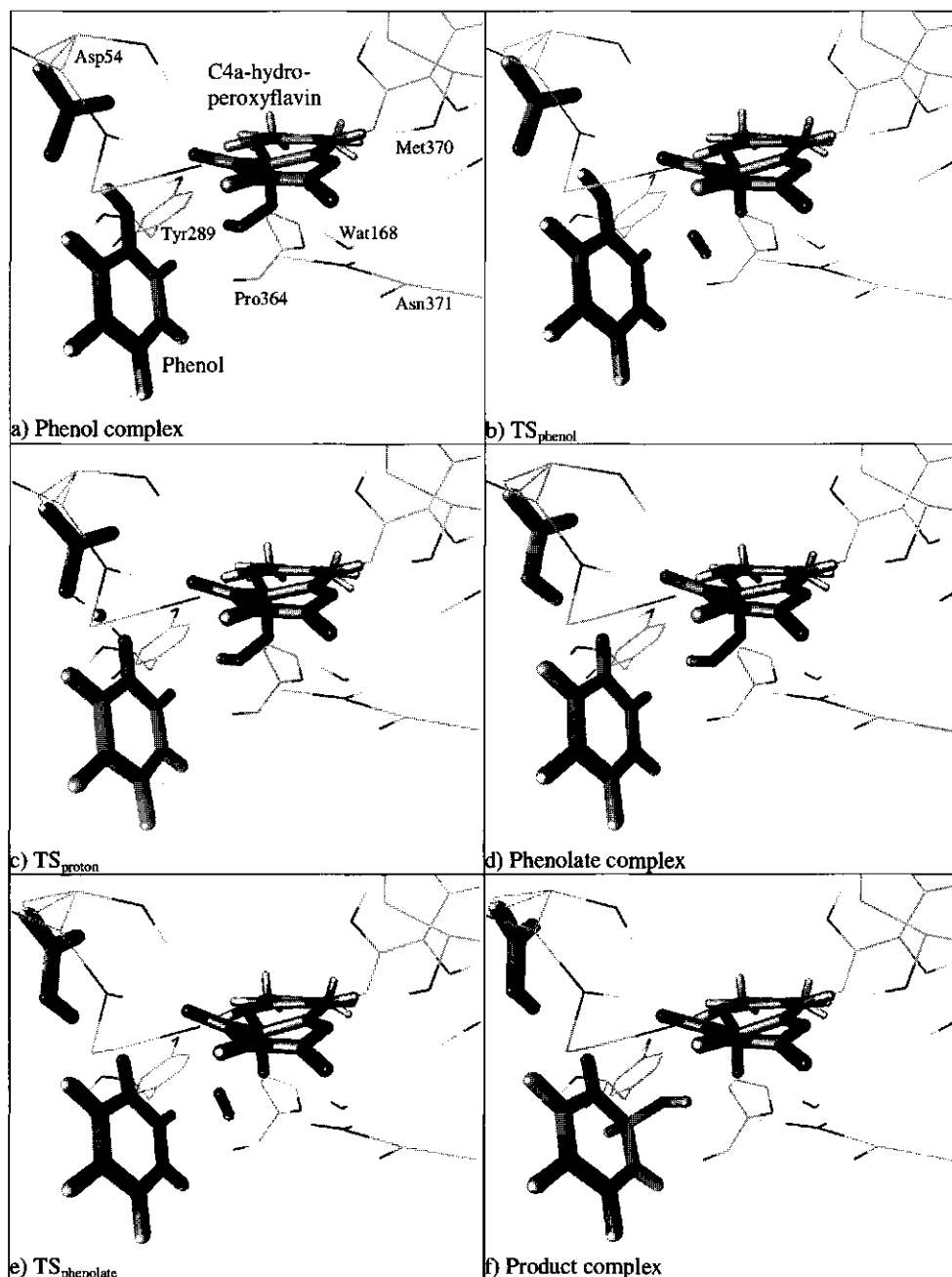


Figure 8.4 Optimized active site structures (showing QM atoms with thick bonds, MM atoms in thin lines using VMD (1996) of a) the phenol–C4a-hydroperoxyflavin complex, b) the transition state for direct hydroxylation of phenol (TS_{phenol}), c) the transition state for proton transfer to Asp54 (TS_{proton}), d) the phenolate–C4a-hydroperoxyflavin complex, e) the transition state for hydroxylation of the phenolate ($TS_{phenolate}$) and f) the product of hydroxylation.

Table 8.2 Mulliken charges on the QM atoms within the protein environment in the various stationary points on the potential surface. In brackets the effect of the protein (in hundredths of an atomic charge unit), calculated as the charge within the protein (as shown) minus the charge in vacuum, obtained from single point calculation of the isolated QM atoms in vacuum (not shown). Although these effects are generally small they indicate that the protein induces polarization in various parts of the QM region. Atom labeling as in Figure 8.1. Dotted lines indicate how the total charges for cofactor, substrate and Asp54 were obtained. Note: Reasonable charges for the link atoms (QQ1 and QQ2) and nearby atoms were obtained, which do not differ dramatically from the charges in vacuum. This indicates that the introduction of link atoms, to connect the QM and MM regions, has not resulted in unrealistic polarization of the system.

atom	Phenol complex	TS _{phenol}	TS _{proton}	Phenolate complex	TS _{phenolate}	Product complex
N1	-0.33 (-4)	-0.35 (-3)	-0.33 (-3)	-0.33 (-3)	-0.34 (-3)	-0.38 (-3)
C2	0.38 (0)	0.39 (0)	0.38 (0)	0.38 (0)	0.39 (0)	0.40 (0)
O2	-0.34 (-1)	-0.36 (-1)	-0.34 (-1)	-0.33 (-1)	-0.36 (-1)	-0.40 (-1)
N3	-0.39 (0)	-0.39 (0)	-0.39 (0)	-0.38 (0)	-0.39 (0)	-0.40 (0)
H3	0.28 (0)	0.26 (1)	0.28 (1)	0.28 (0)	0.26 (1)	0.24 (1)
C4	0.31 (2)	0.30 (1)	0.32 (2)	0.33 (1)	0.32 (2)	0.27 (1)
O4	-0.29 (-8)	-0.35 (-8)	-0.30 (-8)	-0.34 (-8)	-0.36 (-8)	-0.46 (-7)
C4A	0.11 (1)	0.19 (0)	0.10 (1)	0.10 (1)	0.18 (0)	0.30 (0)
N5	-0.26 (-1)	-0.28 (-2)	-0.26 (-1)	-0.25 (-1)	-0.27 (-2)	-0.30 (-2)
H5	0.23 (-1)	0.22 (-1)	0.23 (-1)	0.22 (-1)	0.23 (-1)	0.19 (-1)
C5A	-0.01 (-4)	0.01 (-4)	-0.01 (-4)	-0.02 (-5)	0.00 (-4)	0.03 (-5)
C6	-0.14 (1)	-0.15 (1)	-0.14 (0)	-0.14 (0)	-0.15 (0)	-0.17 (1)
H6	0.14 (-1)	0.14 (-1)	0.14 (-2)	0.14 (-1)	0.14 (-1)	0.13 (-1)
C7	-0.07 (-1)	-0.07 (-2)	-0.07 (-2)	-0.07 (-1)	-0.07 (-1)	-0.08 (-2)
C7M	-0.19 (-2)	-0.18 (-1)	-0.19 (-1)	-0.19 (-2)	-0.19 (-1)	-0.18 (-1)
H71	0.10 (2)	0.09 (2)	0.10 (2)	0.10 (2)	0.09 (2)	0.09 (2)
H72	0.08 (0)	0.08 (0)	0.08 (0)	0.08 (0)	0.08 (0)	0.07 (0)
H73	0.09 (1)	0.09 (1)	0.09 (1)	0.09 (1)	0.09 (1)	0.08 (1)
C8	-0.02 (6)	-0.04 (6)	-0.02 (6)	-0.01 (6)	-0.03 (6)	-0.05 (7)
C8M	-0.23 (-6)	-0.22 (-5)	-0.23 (-6)	-0.23 (-6)	-0.22 (-5)	-0.21 (-6)
H81	0.13 (4)	0.12 (5)	0.13 (5)	0.12 (5)	0.12 (5)	0.11 (4)
H82	0.14 (6)	0.13 (6)	0.14 (6)	0.14 (6)	0.13 (6)	0.13 (7)
H83	0.11 (2)	0.10 (2)	0.10 (2)	0.11 (2)	0.10 (2)	0.10 (2)
C9	-0.16 (-2)	-0.16 (-3)	-0.16 (-2)	-0.16 (-2)	-0.17 (-2)	-0.17 (-3)
H9	0.18 (4)	0.17 (4)	0.18 (4)	0.18 (4)	0.17 (4)	0.16 (4)
C9A	0.00 (-3)	0.00 (-2)	0.01 (-3)	0.01 (-2)	0.00 (-2)	0.00 (-3)
N10	-0.22 (3)	-0.22 (2)	-0.22 (3)	-0.22 (3)	-0.22 (2)	-0.23 (2)
QQ1	0.29 (3)	0.28 (3)	0.29 (3)	0.29 (3)	0.28 (3)	0.26 (3)
C10	0.22 (1)	0.20 (1)	0.22 (1)	0.21 (1)	0.20 (1)	0.16 (1)
Op	-0.19 (-3)	-0.42 (-3)	-0.19 (-2)	-0.18 (-2)	-0.38 (-3)	-0.68 (-4)
Od	-0.18 (-2)	-0.12 (-1)	-0.18 (-2)	-0.19 (-2)	-0.12 (-1)	-0.33 (-1)
Hd	0.24 (2)	0.23 (4)	0.25 (2)	0.26 (1)	0.23 (3)	0.27 (1)

Table 8.2 continued

atom	Phenol complex	TS _{phenol}	TS _{proton}	Phenolate complex	TS _{phenolate}	Product complex
C1'	0.12 (-2)	0.24 (-2)	0.20 (-1)	0.26 (0)	0.30 (0)	0.25 (1)
C2'	-0.20 (1)	-0.22 (0)	-0.26 (1)	-0.31 (1)	-0.28 (-1)	-0.22 (-2)
H2'	0.19 (0)	0.21 (0)	0.15 (0)	0.12 (1)	0.13 (1)	0.17 (1)
C6'	-0.19 (2)	-0.22 (1)	-0.26 (2)	-0.33 (2)	-0.34 (0)	0.02 (0)
H6'	0.15 (3)	0.16 (1)	0.13 (2)	0.12 (2)	0.13 (1)	0.12 (0)
C3'	-0.10 (0)	-0.05 (1)	-0.09 (0)	-0.06 (0)	-0.06 (1)	-0.06 (2)
H3'	0.13 (2)	0.15 (2)	0.12 (2)	0.10 (2)	0.11 (2)	0.15 (2)
C5'	-0.12 (-2)	-0.08 (0)	-0.09 (-1)	-0.07 (-2)	-0.06 (0)	-0.16 (0)
H5'	0.13 (3)	0.15 (4)	0.12 (3)	0.10 (2)	0.12 (3)	0.17 (3)
C4'	-0.22 (-1)	-0.23 (-2)	-0.27 (-2)	-0.32 (-2)	-0.30 (-2)	-0.17 (-2)
H4'	0.11 (0)	0.13 (0)	0.10 (0)	0.09 (0)	0.10 (0)	0.14 (1)
O1'	-0.35 (-5)	-0.31 (-7)	-0.53 (-6)	-0.64 (-7)	-0.56 (-8)	-0.35 (-6)
H1'	0.31 (0)	0.34 (0)	0.35 (0)	0.31 (0)	0.30 (0)	0.27 (0)
QQ2	0.00 (2)	0.02 (2)	0.08 (2)	0.15 (2)	0.16 (2)	0.18 (2)
CG1	0.28 (0)	0.28 (0)	0.28 (0)	0.27 (1)	0.27 (1)	0.27 (0)
OD1	-0.60 (4)	-0.61 (4)	-0.48 (3)	-0.32 (2)	-0.31 (2)	-0.31 (1)
OD2	-0.64 (-5)	-0.64 (-6)	-0.54 (-5)	-0.44 (-5)	-0.43 (-4)	-0.41 (-3)
totals						
cofactor	0.00	-0.42	0.00	-0.01	-0.37	-0.99
substrate	-0.04	0.26	-0.68	-0.96	-0.72	-0.01
Asp54	-0.96	-0.95	-0.66	-0.03	-0.02	-0.01

Active site interactions

Various results obtained from the present QM/MM model provide information about the interactions between the QM and MM regions, and how these interactions change during reaction processes. Table 8.1 (lower section) lists some important interatomic distances between QM and MM atoms, which represent hydrogen bonds in the various intermediates. Changes in the hydrogen bond distances indicate changes in the strength of these interactions, which are often caused by the changes in the electronic structure (e.g. charge distribution; Table 8.2) of the QM system. More quantitative information about the interactions between MM residues and the QM system was obtained from the results of the energy decomposition analyses, presented in Figures 8.5 and 8.6. These analyses provide approximate values for the effects of individual amino acids and water molecules on the change in energy upon transfer of the H1' proton to Asp54 (Fig. 8.5) and on the energy barrier (Fig. 8.6a) and the overall energy change (Fig. 8.6b) for hydroxylation of phenolate. The QM/MM energy differences between the phenol and phenolate complexes (Fig. 8.5), between the phenolate complex and the TS_{phenolate} (Fig. 8.6a) and between the phenolate and product complexes (Fig. 8.6b), are plotted as a function of the amino acids included (in order of the distance of their

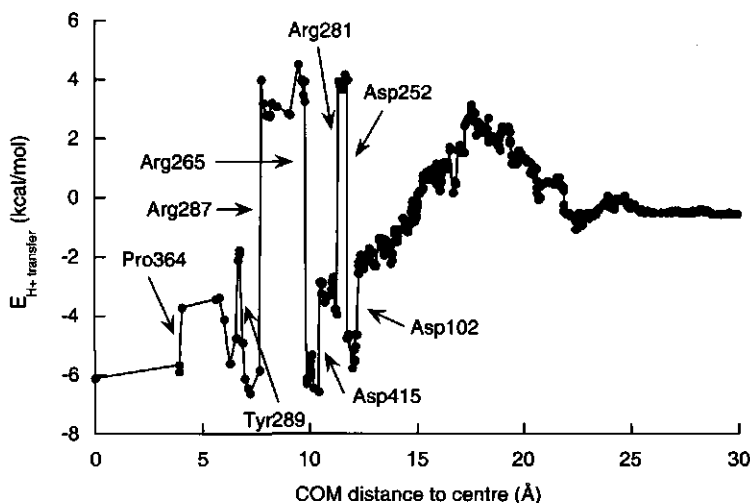


Figure 8.5 Decomposition analysis illustrating the contribution of individual MM residues to the QM/MM energy difference between the phenol C4a-hydroperoxyflavin complex and the phenolate C4a-hydroperoxyflavin complex, i.e. $E(\text{phenolate complex}) - E(\text{phenol complex})$.

centre of mass (COM) to the centre of the simulation system). Combining the data from Tables 8.1 and 8.2 and Figures 8.5 and 8.6, a number of catalytic features of PH with respect to hydroxylation and proton transfer to Asp54 are identified.

First, a hydrogen bond interaction appears to exist between the hydroxyl hydrogen of Tyr289 and O1' of phenol in the initial phenol complex (Table 8.1). This hydrogen bond is expected to be important for binding the substrate. Upon proton transfer to Asp54 leading to the formation of the phenolate complex, the hydrogen bond distance between Tyr289 and the O1' of the substrate decreases from 1.88 \AA to 1.80 \AA , indicating a stabilization of the negative charge on O1' (which increases from $-0.35e$ to $-0.60e$ according to Mulliken analysis). This interpretation is in line with the effect of Tyr289 on the energy change upon proton transfer as observed in the energy decomposition analysis for this step presented in Figure 8.5. The analysis indicates a stabilization of the phenolate complex by Tyr289 by about 3 kcal/mol, relative to the phenol complex. During subsequent hydroxylation of the phenolate, the hydrogen bond distance between Tyr289 and O1' of the substrate increases, which can be explained by the fact that the negatively charged hydroxyl oxygen ($-0.60e$ by Mulliken analysis) formally becomes a neutral carbonyl oxygen ($-0.35e$ by Mulliken analysis) which forms a weaker hydrogen bond with Tyr289. This is in accordance with Figure 8.6, which indicates that Tyr289 increases the energy barrier for hydroxylation by about 3 kcal/mol (Fig. 8.6a) and makes the overall energy change of the reaction approximately 7 kcal/mol less negative (Fig. 8.6b).

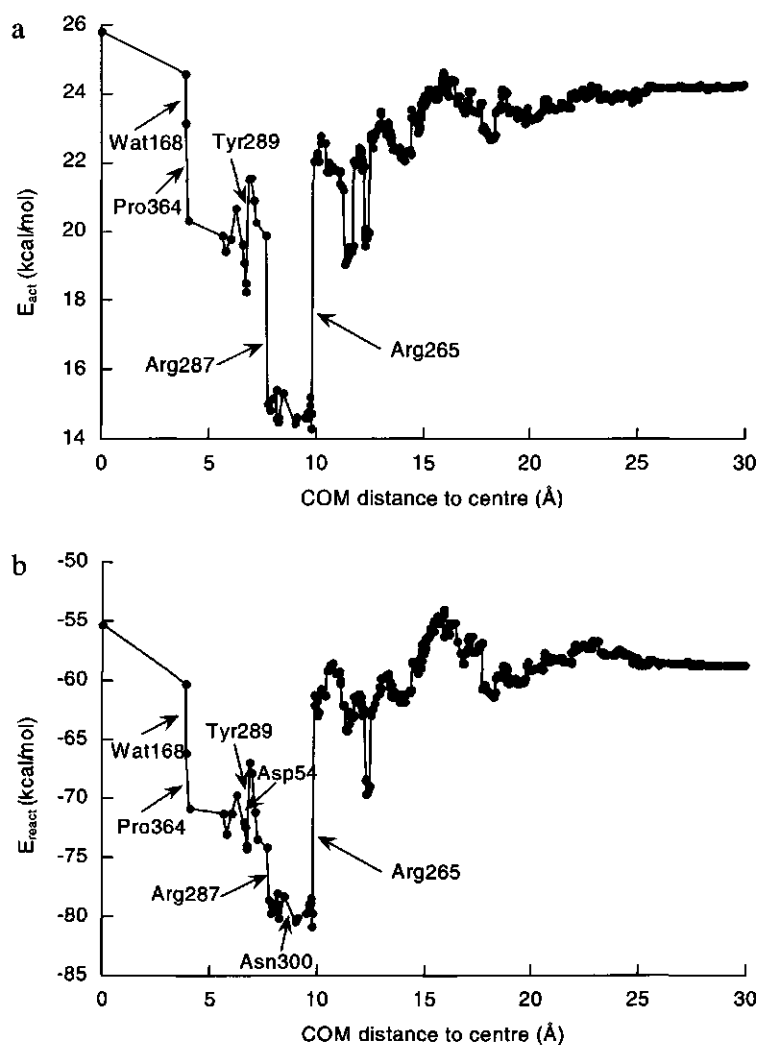


Figure 8.6 Decomposition analysis illustrating the contribution of individual MM residues to the QM/MM energy difference a) between the phenolate C4a-hydroperoxyflavin complex and the transition state for hydroxylation, i.e. $E_{act} = E(TS_{phenolate}) - E(phenolate\ complex)$, and b) between the phenolate C4a-hydroperoxyflavin complex and the product of hydroxylation, i.e. $E_{react} = E(product\ complex) - E(phenolate\ complex)$.

A second remarkable interaction can be identified between the backbone carbonyl oxygen of Pro364 and the Od-Hd moiety. This interaction appears to exist specifically in the transition state for hydroxylation. Table 8.1 indicates that the interatomic distance between Hd and the carbonyl oxygen of Pro364 is much smaller in both transition states for hydroxylation compared to the reactant and product complexes. By

comparing Figures 8.4d, e and f it appears that the Od-Hd moiety rotates almost 180 degrees during its transfer from cofactor to phenolate. In the transition state the OH moiety points towards the backbone oxygen of Pro364. Through this interaction, which is observed only in the transition state, the (partially) negatively charged carbonyl oxygen of Pro364 may stabilize the partial positive charge on Hd in the transition state (0.23e by Mulliken analysis). The energy decomposition analysis for the hydroxylation step indeed indicates that Pro364 lowers the energy barrier for the hydroxylation step by about 3 kcal/mol.

A third important QM/MM interaction involves a crystal water molecule, Wat168. This water molecule remains in the position it occupies in the crystal structure during the initial equilibration of the solvent. This solvent molecule donates a hydrogen bond to the proximal oxygen of the peroxide moiety (Op) of the flavin cofactor. The Wat168-Op hydrogen bond becomes shorter upon hydroxylation of the phenolate, indicating that it increases in strength. (Table 8.1, columns 5 to 7). This suggests that Wat168 plays a role in stabilizing the negative charge on the proximal oxygen Op of the cofactor (the Mulliken charge on Op increases from -0.19e in the phenolate complex to -0.33e in the C4a-hydroxyflavin product). The stabilizing effect of Wat168 on the C4a-hydroxyflavin anion is also apparent from the decomposition analysis which indicates that Wat168 makes the hydroxylation reaction 6 kcal/mol more favorable (Figs. 8.6a and 8.6b).

Five other hydrogen bond interactions, between N1, O2 and O4 of the flavin ring and polar hydrogen atoms of residues Asp54, Gly55, Met370 and Asn371, appear to increase in strength (i.e. the distances become shorter, Table 8.1) upon hydroxylation. Thus, these interactions may stabilize, to some extent, the increasing negative charge on the flavin ring, especially on the N1, O2 and O4 atoms (details in the Supporting Information), upon formation of the deprotonated C4a-hydroxyflavin product of the hydroxylation step. Accordingly, small favorable effects (i.e. making the reaction energy more negative) of the residues Asp54, Met370 and Asn371 on the hydroxylation step are observed in the energy decomposition analysis (Fig. 8.6b).

The strongest effects on the energy changes upon proton transfer and hydroxylation are observed from charged residues (Figures 8.5 and 8.6). Considerable redistributions of charge accompany the proton transfer and the hydroxylation reactions. Therefore, the effects (mainly electrostatic) of charged residues depend on the formal charge and the position of the residue with regard to the QM system in which charge redistributions occur. These strong effects, at relatively long distances (e.g. Asp102, Asp252 and Asp415), indicate that the non-bonded cut-off distance must be chosen with care. The non-bonded cut-off function applied in the present model scales the interactions down to zero over a relatively broad range (between 11.0Å and 16.0Å).

This seems, in retrospect, to be a reasonable treatment of the non-bonded interactions in this system.

Overall, the energy change upon proton transfer is almost zero within the complete protein model, whereas it is -6 kcal/mol for the QM atoms only (Figure 8.5). This indicates that the electrostatic environment of the enzyme favors the phenol complex with Asp54 deprotonated, relative to the phenolate complex. The energetics of the hydroxylation are relatively little affected by the protein (Figures 8.6a+b).

Gas-phase calculations on small model systems for the hydroxylation reaction and for the proton transfer from phenol to Asp54.

A possible source of inaccuracy in the calculated potential energy surface (Figure 8.3) is the fact that it is based on the semiempirical AM1 method. In order to validate the use of AM1 for the QM system, and to obtain insight into how the energy surface may change on going to a higher level of theoretical treatment, gas-phase calculations were performed on small model systems for the two reaction coordinates. The performance of AM1 for these models was tested by comparison to results at higher levels of quantum chemical theory. First, transition state optimizations were performed on the model system for hydroxylation as shown in Figure 8.2a. Some quantitative differences between the results at the various levels of theory are observed. Table 8.3 presents the interatomic distances for the bonds which are broken (Op-Od) and formed (Od-C6') in the calculated transition states at different levels of theory. Comparison of the geometries optimized at the AM1, HF/6-31+G(d), MP2/6-31+G(d) and B3LYP/6-31+G(d) levels indicates that AM1 underestimates the Op-Od bond distance in the TS by about 0.2 Å, whereas it performs well for the Od-C6' distance. The AM1 TS results for the Op-Od and Od-C6' distances (1.68 Å and 1.98 Å, Table 8.3) differ slightly from the corresponding values in the approximate QM/MM transition state for hydroxylation of phenolate (1.63 Å and 2.13 Å, Table 8.1 - column 6), possibly indicating that the flavin ring does affect the reactive properties of the peroxide moiety (as would be expected).

Table 8.3 also presents the TS energies relative to the separate reactants, calculated at different levels of theory. The MP4 or B3LYP methods are the highest levels of theory used and the results obtained at these levels serve as the reference values. Although the AM1 result appears better than the HF/6-31+G(d) outcome, AM1 overestimates the activation energy for hydroxylation, by possibly more than 10 kcal/mol. This may explain the apparent discrepancy between the absolute value of the QM/MM energy barrier and the experimental rate constant. Neglecting entropic contributions, the experimental rate constant of $5 \cdot 10^2 \text{ min}^{-1}$ would correspond to a barrier of about 16 kcal/mol. The difference between this expected barrier and the calculated QM/MM

Table 8.3 Results of transition state optimizations in vacuum for the model system for hydroxylation presented in Figure 8.2a. ΔH_{TS} is the enthalpy of the transition state relative to the total enthalpy of the separate (optimized) reactants. Note: The HF, MP2 and MP4 enthalpies were calculated from the corresponding energies by applying an unscaled thermal correction based on HF/6-31+G(d) frequency calculations. Likewise, the B3LYP enthalpies are based on B3LYP/6-31+G(d) frequency calculations.

Transition states:	d(Op-Od) (Å)	d(Od-C6') (Å)	ΔH_{TS} (kcal/mol)	Imaginary frequency
AM1	1.68	1.98	25.12	725i
HF/6-31+G(d)	1.87	1.92	49.28	983i
B3LYP/6-31+G(d)	1.93	1.97	9.84	391i
B3LYP/6-311+G(2d,p) // B3LYP/6-31+G(d)	id.	id.	10.02	
MP2/6-31+G(d)	1.86	1.99	5.06	
MP4/6-31+G(d) // MP2/6-31+G(d)	id.	id.	11.92	

barrier of 23.7 kcal/mol, is within the error of AM1 indicated above on the basis of the test calculations.

The AM1 method was also tested for the proton transfer from phenol to acetate (as a model for Asp54) in the gas-phase. Geometry optimizations were performed at the AM1 and HF/6-31+G(d) levels for the phenol-acetate complex, the transition state (represented in Figure 8.2b) and the phenolate-acetic acid complex. The energies of the various species at the AM1, HF/6-31+G(d), MP2/6-31+G(d) // HF/6-31+G(d) and B3LYP/6-31+G(d) // HF/6-31+G(d) levels are presented in Table 8.4, relative to the energy of the phenol-acetate complex. AM1 predicts the phenol-acetate complex to be less stable than the phenolate-acetic acid complex. In contrast, the *ab initio* and B3LYP methods predict the phenol-acetate complex to be more stable than the phenolate-acetic acid complex, as would be expected on the basis of the fact that acetic acid is a stronger acid than phenol. The error of about 6 kcal/mol in AM1 seems to be due to a known overestimation of the proton affinity of carboxylate by AM1 (Mulholland and Richards, 1998b).

It should be noted that in the QM/MM model Asp54 is represented by formate in the QM system rather than by acetate. At the AM1 level the proton affinity of acetate (354.8 kcal/mol) and formate (355.1 kcal/mol) are very similar and thus, for reasons of efficiency, formate was chosen to represent Asp54 in the QM/MM model. At the MP2/6-31+G(d) level the proton affinity of acetate (346.57 kcal/mol) differs by 4.7 kcal/mol from the proton affinity of formate (341.87 kcal/mol) and was chosen to be a more realistic representation of Asp54 in the gas-phase model. However, because of the similar proton affinities of formate and acetate at the AM1 level, the comparison of the results for the acetate complex at various levels can be extrapolated to the QM/MM

Table 8.4 Relative energies in kcal/mol of the phenol acetate complex, the transition state for proton transfer (Figure 8.2b) and the phenolate acetic acid complex in the gas-phase at the AM1, HF/6-31+G(d), MP2/6-31+G(d)//HF/6-31+G(d) and the B3LYP/6-31+G(d)//HF/6-31+G(d) levels.

	phenol complex	TS	phenolate complex
AM1	0	4.56	-0.73
HF/6-31+G(d)	0	7.66	6.96
MP2/6-31+G(d)	0	4.29	6.21
//HF/6-31+G(d)			
B3LYP/6-31+G(d)	0	2.18	4.67
//HF/6-31+G(d)			

model. Thus, the comparison of AM1 with the higher levels of theory in Table 8.4 suggests that the proton transfer from phenol to Asp54 in the active site of PH may be somewhat more unfavorable than it appears from the calculated potential energy surface, due to an overestimation of the proton affinity of Asp54 in the QM/MM model.

In summary, the results of the test calculations in gas-phase suggest some corrections to the potential energy surface presented in Figure 8.3 and the energies given in Table 8.1. The phenolate complex may be about 6 kcal/mol more unfavourable relative to the phenol complex. Furthermore, the energy barrier for the subsequent hydroxylation may be lower. From a qualitative point of view, however, the conclusions drawn from the potential energy surface remain unchanged. The approximate transition state for hydroxylation of the phenolate in the QM/MM model (formed upon proton transfer) is 13 kcal/mol lower than the transition state for direct hydroxylation of the phenol. Even if a correction of 6 kcal/mol on the energy change upon proton transfer from phenol to Asp54 were to be extrapolated to both transition states for hydroxylation (which might be an overcorrection) the potential energy surface would still favor a reaction mechanism involving initial proton transfer based on a residual difference of about 7 kcal/mol between the transition states of hydroxylation of the phenol and phenolate complexes.

Computation of energy barriers with halogenated substrates

In subsequent calculations, the energy barrier for hydroxylation of the phenolate complex, after initial proton transfer, was also determined for a number of substituted substrates. The structures of the phenolate and transition state complexes calculated for the substrate analogs were in all cases comparable to those found for the native phenolate. The energy barriers and reaction coordinate values of the approximate transition states for the various substrates are listed in Table 8.5. For asymmetric substrates the calculations were performed for the two possible binding orientations in

Table 8.5 Experimental rate constants for overall conversion of a series of halogenated phenols by phenol hydroxylase (Peelen *et al.*, 1995), and calculated energy barriers $E_{act(1)}$ for the electrophilic attack of the C4a-hydroperoxyflavin on the C6' of the phenolate substrates. In case of asymmetric substrates a second energy barrier E_{act2} is calculated for the attack of the C4a-hydroperoxyflavin on the C6' of the substrates with the alternative substituent positions (in brackets, column 1), representing the reaction with the substrates in the alternative binding orientation. Columns 4 and 6 present the reaction coordinate value at the transition state, $r_{OH}(TS)$, with $r_{OH} = d(Op-Od) - d(Od-C6')$. The last column presents the apparent energy barriers, $E_{act,app}$, derived using equation (3) assuming an equal contribution of the two substrate orientations.

substrate	k_{cat} (min^{-1})	$E_{act(1)}$ (kcal/ mol)	$r_{OH}(TS)$ (Å)	E_{act2} (kcal/ mol)	$r_{OH}(TS)$ (Å)
phenol	470	23.70	(-0.50)		
2-F-phenol (6-F)	150	25.68	(-0.49)	26.32	(-0.50)
3-F-phenol (5-F)	280	24.69	(-0.49)	25.35	(-0.49)
4-F-phenol	470	24.58	(-0.49)		
2,3-F ₂ -phenol (5,6-F ₂)	90	26.35	(-0.47)	27.23	(-0.49)
2,4-F ₂ -phenol (4,6-F ₂)	190	25.98	(-0.47)	27.52	(-0.47)
2,5-F ₂ -phenol (3,6-F ₂)	150	27.25	(-0.47)	27.29	(-0.48)
3,4-F ₂ -phenol (4,5-F ₂)	390	25.56	(-0.47)	26.09	(-0.47)
3,5-F ₂ -phenol	280	26.37	(-0.47)		
2,3,4-F ₃ -phenol (4,5,6-F ₃)	150	26.91	(-0.44)	28.56	(-0.46)
2,3,5-F ₃ -phenol (3,5,6-F ₃)	70	28.09	(-0.45)	28.42	(-0.47)
3,4,5-F ₃ -phenol	210	27.26	(-0.45)		
2,3,5,6-F ₄ -phenol	50	30.04	(-0.44)		
2,3,4,5,6-F ₅ -phenol	40	30.88	(-0.41)		
3-Cl,4-F-phenol (5-Cl,4-F)	190	24.89	(-0.45)	26.52	(-0.47)
4-Cl,3-F-phenol (4-Cl,5-F)	270	24.44	(-0.45)	24.85	(-0.47)

the active site. Overall, the substituents increase the energy barrier. This is likely to be due to the electronegative character of the substituents, which withdraw electron density from the aromatic ring, lowering the reactivity of the substrate for electrophilic attack by C4a-hydroperoxyflavin (Peelen *et al.*, 1995). Fluorine substituents on the reacting C6' centre seem to have a stronger effect on the barrier than fluorine substituents on other positions. The fluorine substituents on this position may affect the energy involved in changing the hybridization of C6' from sp^2 to sp^3 and in loss of aromaticity upon hydroxylation; another possibility is that a fluorine substituent at this position sterically hinders the reaction. The substitutions (except for those on C6') cause a shift of the TS geometry towards the product geometry, i.e. $r_{OH}(TS)$ becomes less negative, which corresponds a decrease in $Od-C6'$ (varying between 2.09 Å and 2.14 Å) distance and an increase in $Op-Od$ distance (varying between 1.62 Å and 1.70 Å).

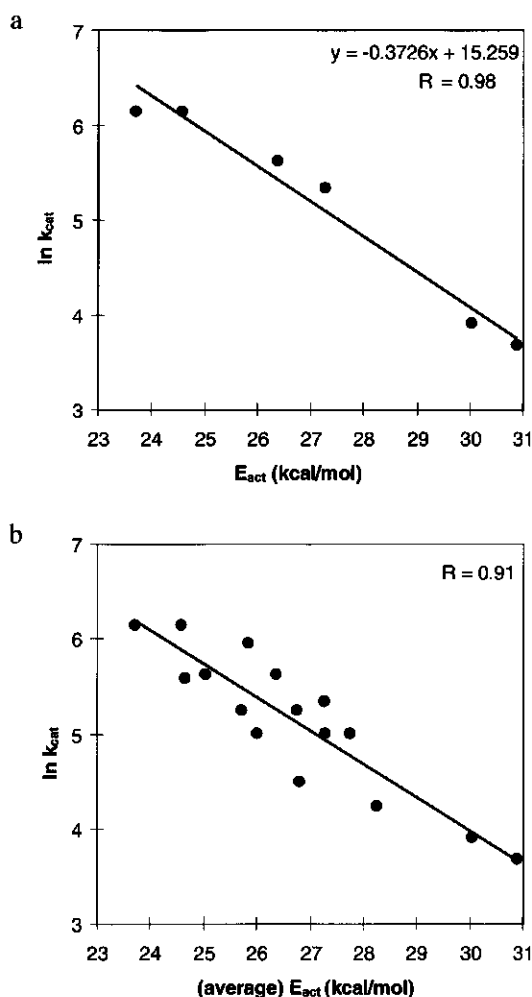


Figure 8.7 Linear correlations between the natural logarithm of the experimental rate constants (min^{-1}) for overall conversion of symmetric phenol derivatives (Peelen *et al.*, 1995) and the (average) calculated energy barriers (E_{act}) obtained for a) the symmetric substrates only and b) all substrates.

Correlation of calculated barriers with experimental rate constants

Based on the hypothesis that the electrophilic attack of the C4a-hydroperoxyflavin on the substrate is rate limiting in the overall reaction cycle of PH, the calculated energy barriers for the different substrates were compared to the experimental rate constants for their overall conversion by PH (Peelen *et al.*, 1995). For the symmetric substrates, indeed an excellent linear correlation ($R = 0.98$) was found between the logarithm of the experimental rate constants and the calculated energy barriers for the reaction of

Table 8.6 Fractions of substrate in the orientation corresponding to substituent labeling given in the full names of the substrates (f_1) and in the orientation corresponding to substituent labeling given in brackets (f_2), derived on the basis of equation (6) using experimental product ratios (Peelen *et al.*, 1995) and rate constants (not shown) for the reactions in both orientations predicted from the calculated energy barriers and the correlation in Figure 8.8a. Apparent activation barriers $E_{act,app}$ were calculated using equation (5).

substrate	f_1	f_2	$E_{act,app}$ (kcal/mol)
2-F-phenol (6-F)	0.66	0.34	25.83
3-F-phenol (5-F)	0.66	0.34	24.85
2,3-F ₂ -phenol (5,6-F ₂)	0.84	0.16	26.43
2,4-F ₂ -phenol (4,6-F ₂)	0.69	0.31	26.18
2,5-F ₂ -phenol (3,6-F ₂)	0.34	0.66	27.28
3,4-F ₂ -phenol (4,5-F ₂)	0.92	0.08	25.59
2,3,4-F ₃ -phenol (4,5,6-F ₃)	0.98	0.02	26.92
2,3,5-F ₃ -phenol (3,5,6-F ₃)	0.60	0.40	28.20
3-Cl,4-F-phenol (5-Cl,4-F)	0.93	0.07	24.93
4-Cl,3-F-phenol (4-Cl,5-F)	0.93	0.07	24.46

the C4a-hydroperoxyflavin with the substrate (Figure 8.7a). For the asymmetric substrates slightly different energy barriers are obtained for the two possible orientations in the active site (Table 8.5). Figure 8.7b shows that the average energy barriers for the asymmetric substrates, together with the energy barriers for the symmetric substrates, also correlate well ($R = 0.91$) with the natural logarithm of the experimental rate constants for overall hydroxylation.

A more sophisticated approach to correlate the calculated barriers with experimental rate constants for the asymmetric substrates was tested, based on the following derivation. The rate constants for the asymmetric substrates are in fact a weighted summation of two different rate constants depending on the fraction of substrate present in each orientation, f_1 and f_2 :

$$k_{cat} = f_1 \cdot k_1 + f_2 \cdot k_2 \quad (1)$$

Consistent with the Arrhenius equation,

$$k = Ae^{-E_{act}/RT} \leftrightarrow \ln k_{cat} \sim -E_{act} \quad (2)$$

the rate constants k_1 and k_2 in (1) can be replaced by exponentials of the energy barriers E_{act1} and E_{act2} respectively:

$$k_{cat} = A(f_1 e^{-E_{act1}/RT} + f_2 e^{-E_{act2}/RT}) \quad (3)$$

In line with the linear relationship in (2) this can be written as:

$$\ln k_{cat} \sim -E_{act,app} = RT \ln(f_1 e^{-E_{act1}/RT} + f_2 e^{-E_{act2}/RT}) \quad (4)$$

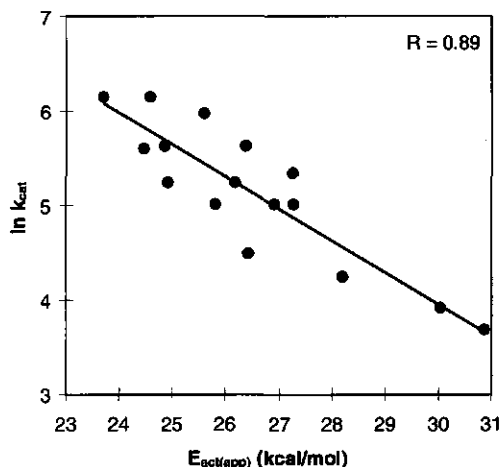


Figure 8.8 Linear correlation between the natural logarithm of the experimental rate constants for overall conversion of symmetric and asymmetric phenol derivatives (Peelen *et al.*, 1995) and the calculated energy barriers (E_{act} and $E_{act,app}$). The apparent energy barriers ($E_{act,app}$) were derived on the basis of equation (3).

It was attempted to derive the occupancies of the two substrate orientations f_1 and f_2 from the ratio of the two products formed p_1/p_2 , determined previously using ^{19}F -NMR (Peelen *et al.*, 1995), and the ratio of both conversion rates k_1/k_2 .

$$\frac{p_1}{p_2} = \frac{k_1}{k_2} \frac{f_1}{f_2} \quad \text{thus:} \quad \frac{f_1}{f_2} = \frac{p_1}{p_2} \frac{k_2}{k_1} \quad (5)$$

The correlation obtained for the symmetric substrates (Figure 8.7a) can be used to predict the rate constants k_1 and k_2 for the asymmetric substrates on the basis of the calculated energy barriers in the two orientations (Table 8.5). The values thus obtained for f_1 and f_2 (using: $f_1 + f_2 = 1$) are presented in Table 8.6. These values could be introduced into equation (3), with $T=298\text{K}$, and the resulting apparent energy barriers (Table 8.6) were correlated with the logarithm of the experimental rate constants (Figure 8.8). The correlation obtained in this second approximation ($R = 0.89$) is similar to the correlation obtained with the average energy barriers. Table 8.5 indicates that the differences in energy barriers are relatively small ($< 20\%$ of the overall variation in the energy barriers). The fact that the correlation obtained with the more sophisticated approach is not better than the one obtained with the simply averaged energy barriers may indicate that the differences in the calculated energy barriers for the two different substrate orientations do not contain significant information on the regioselectivity of the reaction.

8.4 Discussion

The potential energy surface obtained in the present study is based on an energy minimization approach, which is also referred to as adiabatic mapping (McCammon and Harvey, 1987). This approach does not include contributions of the protein dynamics, which may involve multiple conformational substates (Frauenfelder *et al.*, 1991, Kurzynski, 1998). Instead, it is based on a single conformation, which is expected to be representative for the reacting enzyme. The conformation used is derived from the crystal structure of the oxidized enzyme-substrate complex, with the flavin in the "in" position (Enroth *et al.*, 1998). In analogy to the case of PHBH, the reactive C4a-hydroperoxyflavin is in a similar position (Schreuder *et al.*, 1990) and the protein structure is not expected to change significantly upon hydroxylation (Ridder *et al.*, 1999a, Schreuder *et al.*, 1989, Schreuder *et al.*, 1988b).

The observed correlation, between the calculated QM/MM energy barriers and the logarithm of the experimental rate constants of overall conversion by PH, indicates that the QM/MM model provides an appropriate description of the rate limiting step (at 25°C and pH 7.6) in the reaction cycle of PH. The analysis of the changes in geometry and charge, along the reaction coordinate for hydroxylation of the phenolate, indicates that the (validated) QM/MM model is in line with an electrophilic aromatic substitution which results in the formation of a hydroxy-cyclohexadienone as the initial product. Although the calculated energy barriers for hydroxylation might be somewhat too high, as suggested by the gas-phase results on the small model system (Table 8.2) and in line with a previous conclusion for PHBH (Ridder *et al.*, 1998), the results show that the barriers are useful in a relative way. The correlation for a broad range of substrates, including not only fluorinated phenols, but also chlorinated phenols, holds out the possibility of predicting conversion rates of other substrates by phenol hydroxylase (which might be significant information with respect to their aerobic degradability). It shows the potential utility of the QM/MM method for quantitative structure activity relationship (QSAR) studies.

The present simulation provides new insight into the activation of the phenolic substrate by proton transfer from the hydroxyl moiety of the substrate to a potential active site base, Asp54. NMR experiments indicate that the substrate binds to the oxidized PH enzyme in its neutral protonated state (Peelen *et al.*, 1993). Apparently, the active site base Asp54, which appears to form a hydrogen bond with the hydroxyl moiety of the substrate on the basis of the crystal structure, does not deprotonate the substrate at that stage in the reaction cycle. The QM/MM model suggests that, further on in the reaction cycle, the electrophilic attack of the C4a-hydroperoxyflavin on the substrate occurs after a proton has been transferred from phenol to Asp54.

The first question that one may address is at what point in the reaction cycle the proton is transferred from phenol to Asp54. If the QM/MM potential energy surface obtained in the present simulation study (Figure 8.3) is corrected for an overestimation (by up to approximately 6 kcal/mol) of the Asp54 proton affinity by AM1, which appears from the comparison to higher level calculations (Table 8.4), the phenolate complex would be almost 6 kcal/mol higher in energy than the phenol complex. This suggests that, in the C4a-hydroperoxyflavin intermediate of the PH reaction cycle, the proton may still be preferentially present on the phenolic substrate. However, although the initial proton transfer step might be energetically unfavorable, the lower energy barrier of the subsequent electrophilic attack seems to more than compensate for this. In kinetic terms this means that, although only a small fraction of the substrate is in a deprotonated state, the subsequent hydroxylation reaction proceeds so much faster in this deprotonated state that this pathway prevails over the hydroxylation of the substrate in its protonated state.

This idea about substrate activation in PH, i.e. that it occurs just before the hydroxylation takes place, is different from the mechanism of PHBH, in which substrate is deprotonated directly upon binding in the active site (before the flavin cofactor is reduced) (Entsch *et al.*, 1991, Eschrich *et al.*, 1993, Schreuder *et al.*, 1994). Thus, the second question that comes up is what would be the advantage of keeping phenol neutral during the first steps in the reaction cycle of PH (whereas in PHBH substrate is deprotonated directly after binding) ? Related to the fact that the hydroxyl moiety of phenol is the only polar group of the molecule to be involved in orientating the substrate in the active site, one could speculate that the substrate might be more specifically and/or efficiently, bound (and orientated) by the two hydrogen bonds with Asp54 and Tyr289 when phenol is present in its neutral state. In PHBH, the *p*-hydroxybenzoate substrate is orientated by hydrogen bonding of the carboxyl moiety as well, which makes the role of the hydroxyl moiety in substrate binding/orientation less important. In this context, it is interesting to note that, by structural alignment of the phenol hydroxylase and *p*-hydroxybenzoate hydroxylase crystal structures, the hydroxyl moiety of phenol in phenol hydroxylase binds at the site where the carboxyl moiety (instead of the hydroxyl moiety) of *p*-hydroxybenzoate binds in *p*-hydroxybenzoate hydroxylase.

In combination with previous work (Ridder *et al.*, 1999a), the results of the present study allow a detailed comparison of the catalytic features of the active sites of the two related enzymes, PH and PHBH, with respect to the hydroxylation step. The observed changes in charge distribution and bond distances upon hydroxylation are very similar in the present simulation of PH and the previous simulation of PHBH, which supports the proposal that the mechanism of hydroxylation is similar in both enzymes. In

addition, some common active site interactions appear to play a role in both hydroxylation reactions. A crystal water molecule is present at a similar position in the active sites of both enzymes structures. In the simulations of both enzymes, this water molecule donates a hydrogen bond to the proximal oxygen (Op) of the C4a-hydroperoxyflavin cofactor, which results in a stabilization of the increasing negative charge on Op. Catalytic effects of bound water molecules, by stabilizing transient negative charges, have been recently reported for other enzymes too (Mulholland and Richards, 1997, Zheng and Ornstein, 1997), which may indicate their general importance. A second comparison can be made for the role of a tyrosine (Tyr289 in PH and Tyr201 in PHBH) which forms a hydrogen bond with the hydroxyl moiety of the substrate. In PHBH, Tyr201 is known to be directly involved in deprotonating the substrate (Entsch *et al.*, 1991, Eschrich *et al.*, 1993), as the first step in a proton channel to transport the proton out of the active site (Gatti *et al.*, 1996, Schreuder *et al.*, 1994). The present study indicates that in PH substrate may be deprotonated by Asp54 rather than Tyr289. Furthermore, the position of Tyr289 in the active site of PH is completely different from the position of Tyr201 in PHBH. (It is, however, comparable in position to Tyr222 in PHBH, which binds the substrate carboxylate.) Despite these important differences, one effect of Tyr289 in PH and Tyr201 in PHBH seems to be similar, based on the comparison of the present and previous QM/MM simulations respectively: it appears to stabilize the activated substrate by forming a hydrogen bond with the deprotonated hydroxyl moiety. In both simulations, this hydrogen bond becomes weaker along the reaction coordinate of the hydroxylation step, which results in a slight unfavourable effect on this reaction step.

A third common feature is the catalytic function of the backbone carbonyl moiety of a conserved proline residue, i.e. Pro293 in PHBH and Pro364 in PH, which is similarly orientated towards the FAD cofactor in both structures. The positions of these proline residues are also identical in the (structural) alignment of both proteins (Enroth *et al.*, 1998). The (partial) negatively charged carbonyl oxygen specifically stabilizes the transition state by interacting with the somewhat positively charged Od-Hd moiety in the process of being transferred from the cofactor to the substrate. It is an interesting question whether the catalytic feature of the backbone carbonyl observed in the simulation studies on PHBH and PH could be of a more general significance. A backbone carbonyl oxygen has a relatively large (partial) negative charge due to π -interaction with the free electron pair on the nitrogen through the peptide bond. The backbone carbonyl moiety is therefore a suitable candidate for stabilizing positively charged groups. Furthermore, the advantage of a (carbonyl) backbone group is that its position and orientation is relatively well defined (i.e. restricted). This allows the carbonyl moiety to be tightly oriented, such that it can specifically stabilize a transition

state, as appears to be the case in the present simulation study. One could argue about the significance of a proline residue being present at this specific position in an active site loop, in both PH and PHBH. It is interesting to note that this proline residue is part of the FAD fingerprint motif in the flavoprotein monooxygenases (Dimarco *et al.*, 1993, Eppink *et al.*, 1997a). It is tempting to speculate that the suggested catalytic role of the backbone carbonyl could be one of the reasons for the conservation of a proline residue at this position. The proline residue, due to its cyclic structure, might induce and restrain a specific conformation of the backbone at this position and thus keep the backbone carbonyl moiety in the right orientation for its stabilizing interaction in the transition state.

Acknowledgements

We would like to thank Willem J. H. van Berkel for useful discussions. AJM is an EPSRC Advanced Research Fellow. This work was supported by the Netherlands Organisation for Scientific Research (NWO).

9

Summary and conclusions

9.1 Summary

Theoretical simulations are becoming increasingly important for our understanding of mechanisms of enzymes. The aim of the research presented in this thesis is to contribute to this development by applying various computational methods to three enzymes of the β -ketoadipate pathway, and to validate the models obtained by means of quantitative structure-activity relationships (QSAR). The models and the resulting QSARs can provide valuable mechanistic information about the relevant (rate-limiting) steps in the reaction cycles of the enzymes studied (**Chapter 1**).

Two of the enzymes that have been studied in this thesis, are flavin-dependent monooxygenases: *para*-hydroxybenzoate hydroxylase (PHBH) from *Pseudomonas fluorescens*, and phenol hydroxylase (PH) from *Trichosporon cutaneum*. These enzymes catalyse the *ortho*-hydroxylation of *para*-hydroxybenzoate and phenol, leading to the formation of respectively catechol and protocatechuate, key intermediates in the microbial degradation of many aromatic compounds. Once the catechol or protocatechuate is formed, the aromatic ring can be cleaved between the two hydroxyl-substituted carbon atoms. This intradiol cleavage is catalysed by another enzyme studied in this thesis, catechol-1,2-dioxygenase (1,2-CTD), and by the homologous protocatechuate-3,4-dioxygenase (3,4-PCD), respectively.

Chapters 2 and 3 describe some theoretical principles and practical aspects of the computational approaches used in this thesis. Pure quantum mechanical calculations can be used to calculate properties of substrates and cofactors, or somewhat larger active site models, in vacuum. One useful parameter for enzyme activity, which can be obtained from QM calculations, is the energy of the relevant frontier orbital of the reacting molecule. This frontier orbital can be either the highest occupied molecular

orbital (HOMO) representing the nucleophilic reactivity of a reactant or the lowest unoccupied molecular orbital (LUMO) representing its electrophilic reactivity. Another useful gas-phase property is the energy difference between the reactants and the products of the reaction step of interest, which, according to the Brønsted relationship, is related to the activation barrier. The results of these (gas-phase) calculations may be used as parameters that can be correlated linearly with the logarithm of the experimental rate constants of enzymatic conversion of series of related compounds.

An important step forward, that has been made in this thesis, is the use of a combined quantum mechanical/molecular mechanical (QM/MM) method. Using this method, the quantum mechanical (reaction pathway) calculation on the reacting compounds could be performed within the actual environment of the protein. The surrounding protein atoms are calculated at a molecular mechanical (MM) level. Their electrostatic and steric effects on the quantum mechanical system is accounted for by the inclusion of the MM atomic charges into the Hamiltonian of the QM atoms, and by the addition of a Van der Waals interaction term respectively.

Catechol dioxygenase

The reaction mechanism of catechol-1,2-dioxygenase from *Pseudomonas putida* has been studied by means of a QSAR approach based on gas-phase molecular orbital calculations (**Chapter 4**). Catechol-1,2-dioxygenase catalyses intradiol cleavage of the aromatic ring of catechol by incorporating both oxygen atoms of molecular oxygen. In addition to the native catechol, this enzyme converts several C4-substituted catechol derivatives. In this study, 4-methyl-, 4-fluoro-, 4-chloro-, 4-bromo, 4,5-difluoro- and 4-chloro,5-fluoro-catechol were obtained biosynthetically from the corresponding phenols by using the enzyme phenol hydroxylase. The overall rate constants for the conversion of these catechols by catechol-1,2-dioxygenase were determined by steady-state kinetic experiments at various oxygen concentrations and saturating catechol concentrations.

The crucial step in the reaction mechanism of the enzyme catalysed reaction was considered to be the nucleophilic attack of the substrate on the oxygen molecule. Therefore, the experimental results were compared to calculated energies of the HOMO of the catechol substrate, representing its nucleophilic reactivity. According to the reaction cycle proposed in the literature, this step occurs after bidentate binding of the dianionic substrate to the non-heme iron in the active site of the enzyme. The effect of this binding on the electron density of the substrate was estimated on the basis of a computational model of the complete iron-ligand-substrate complex. The result of this calculation, based on density functional theory, which generally provides relatively

accurate results for transition metal systems, indicated that the charge on the substrate, being close to $-1e$ instead of $-2e$, is partly delocalised over the complex. Based on this result, the calculations on the isolated substrate molecules were performed with the monoanionic substrates, to mimic their charge relevant for the situation in the active site. The HOMO energies obtained from these calculations were shown to correlate with the logarithm of the experimental rate constants for their conversion by catechol-1,2-dioxygenase. This indicates that the rate-limiting step in the overall reaction cycle involves a nucleophilic reaction of the substrate. Thus, the reaction of the substrate with molecular oxygen may indeed be rate limiting. However, it can not be excluded that one of the other steps in the reaction cycle also depends on the nucleophilic reactivity of the substrate and is influenced by the effect of the substituents on the most reactive electrons within the substrate molecule. Therefore, additional calculations were performed for the different substrates to obtain energies of the various intermediates in the reaction cycle. The relative energy changes, calculated for the individual reaction steps, were also compared to the experimental rate constants with the various substrates. The correlations obtained indicated that, in addition to the nucleophilic attack on molecular oxygen, the subsequent step in the reaction cycle, i.e. rearrangement to yield a cyclic anhydride intermediate, may also be the rate-limiting step. Furthermore, these results excluded the last two steps in the reaction cycle from being rate limiting.

The results for catechol-1,2-dioxygenase from *Pseudomonas putida* were also compared to the data from two different types of catechol-1,2-dioxygenase, a normal (type I) and a chloro-catechol dioxygenase (type II), from *Pseudomonas sp.* B13. It could be argued that the difference in substrate preference between both types of catechol dioxygenases is related to a differential effect of the substituents on the rate of oxygen affinity binding by the two enzymes, rather than on the rate-limiting step.

p-Hydroxybenzoate hydroxylase

In chapters 5, 6 and 7, a QM/MM model has been developed and validated for the hydroxylation step catalysed by *p*-hydroxybenzoate hydroxylase (PHBH), a flavin dependent monooxygenase. It was first investigated whether the energy barriers obtained from QM/MM reaction pathway calculations could be used to explain the variation in the overall rate constants for the conversion of a series of fluorinated substrates by PHBH (Chapter 5). Reaction pathways were calculated for the proposed rate-limiting step in the reaction cycle: the electrophilic attack of the C4a-hydroxyperoxyflavin cofactor intermediate on the substrate. The energy profiles calculated for this reaction step with the various substrates yielded barriers of different heights. A correlation was found between the natural logarithm of the experimental

overall rate constants for conversion of the fluorinated substrates by PHBH and the QM/MM calculated energy barriers. This correlation with overall rate constants supports that the electrophilic attack of the C4a-hydroxyperoxyflavin on the substrate is indeed the rate-limiting step in the reaction cycle.

The correlation also indicates that the QM/MM model provides a realistic description of the hydroxylation step, as it accounts correctly for the effect of substrate substituents on the rate of hydroxylation. This was the basis for a further and more detailed analysis of the QM/MM model, which is described in **chapter 6**. This chapter gives insight into the mechanisms of substrate and cofactor activation to facilitate the (rate limiting) electrophilic attack of the C4a-hydroxyperoxyflavin cofactor intermediate on the substrate. Deprotonation of the substrate is shown to significantly lower the calculated energy barrier for this reaction step. This effect on the reaction could be interpreted in terms of frontier orbital theory. Substrate deprotonation increases the energy of the relevant HOMO-electrons and their density on the reacting C3 atom of the substrate, leading to a lower activation barrier according to frontier orbital theory. Furthermore, some insight was obtained in the reactivity of the C4a-hydroperoxyflavin cofactor. A lengthening of the peroxide oxygen-oxygen bond is observed at an intermediate stage in the calculation reaction pathway, accompanied by an increase of LUMO density on the reacting distal oxygen. This suggests a role for oxygen-oxygen bond lengthening in activating the cofactor for the electrophilic attack on the substrate. Finally, the QM/MM model allowed the analysis of the energetic effect of the individual amino acid residues on the hydroxylation reaction. The results suggest catalytic effects of the backbone carbonyl moiety of Pro293, by a specific stabilisation of the transition state, and of a (crystal) water molecule (Wat717), which stabilises the negative charge arising on the proximal oxygen of the flavin cofactor.

In **chapter 7**, the different QSAR-parameters mentioned above are used for an extended QSAR for the hydroxylation step in PHBH, including substitutions on the substrate as well as on the cofactor. The QSARs described in this chapter are based on rate constants for the individual hydroxylation step (instead of overall rate constants), derived from stopped-flow experiments. This study revealed that the gas-phase parameters, i.e. the HOMO-LUMO energy difference and the gas-phase reaction energy, yield reasonable correlations with the experimental rate constants obtained for various similarly charged substrates and flavin cofactors. The negatively charged 8-hydroxy-flavin and the 8-mercapto-flavin cofactors, however, deviated from these QSARs. This is likely to be due to a reduction of the effect of the negatively charged substituents on the reaction energy by a shielding and stabilising effect of surrounding residues, which is not accounted for in the gas-phase calculations. In contrast, the QM/MM calculations, which do account for this effect in an approximate way,

allowed the inclusion of all substrate-cofactor combinations investigated, i.e. including the charged 8-hydroxy-flavin and the 8-mercapto-flavin cofactors, into one correlation. This illustrates an important advantage of the QM/MM approach over calculations in vacuum, for studies on solvated systems including reactions within the surroundings of an enzyme.

Phenol hydroxylase

Chapter 8 presents a QM/MM study on phenol hydroxylase (PH). As for PHBH, the hydroxylation step, proposed to be rate limiting in the reaction cycle of PH, has been simulated for a series of halogenated substrate derivatives. The energy barriers obtained correlate well with the logarithm of the experimental rate constants for overall conversion by PH. This correlation supports that the electrophilic attack of the C4a-hydroperoxyflavin on the substrate is the rate-limiting step in the reaction cycle at pH 7.6 and 25°C. An additional mechanistic question addressed in this study, is the protonation state of the substrate during hydroxylation. Substrate deprotonation has not been established for PH as firmly as it has been for PHBH. Proton transfer from phenol to a potential active site base, Asp54, has been investigated by calculating a 2-dimensional potential energy surface for the two reaction coordinates, i.e. hydroxylation and proton transfer. This potential energy surface suggests that proton transfer prior to hydroxylation is the most favourable mechanism. This indicates that in the PH reaction, substrate deprotonation is important as well. However, this deprotonation in PH seems to occur at a later stage in the reaction cycle (e.g. just before hydroxylation), whereas in PHBH it occurs directly upon binding of the substrate to the active site. The QM/MM model was further analysed to provide insight into the effect of the protein environment on the simulated reaction steps. Important catalytic effects on the hydroxylation step, i.e. of a proline carbonyl moiety and of a crystal water in the active site of PH, were similar to those found for PHBH.

9.2 Conclusions

The present thesis has resulted in validated computational models for three different enzyme reactions. The models provide parameters that correlate linearly with the logarithm of experimental rate constants for enzymatic conversion of series of substrates.

QSARs based on frontier orbital energies and reaction energies, give insight into the nature of the reaction step or can be used to identify the rate-limiting step in a given reaction cycle. However, these gas-phase parameters do not account for effects of the active site, such as screening and stabilisation of localised charges (which may be subject to redistribution along the reaction coordinate).

The influence of the protein environment can be accounted for in a more complete way, by means of the combined quantum mechanical / molecular mechanical (QM/MM) technique. The reaction pathways calculated with this QM/MM method provide more detailed insight into the mechanism of the simulated reactions. They allow the characterisation of the unstable intermediates (e.g. the hydroxy-cyclohexadienone intermediate in PHBH and PH) and reveal the role of individual protein residues in catalysing the reaction. This mechanistic information is often difficult to obtain from experiment, because of the short lifetimes of many reaction intermediates.

The energy barriers obtained from QM/MM reaction pathway calculations, at present, can not be used as absolute activation energies. However, this thesis clearly demonstrates that the calculated QM/MM energy barriers are very useful as relative parameters for enzyme activity, showing a linear correlation with the logarithm of experimental rate constants for the enzymatic conversion. The QM/MM technique potentially allows QSARs for broader ranges of structural variation, since it takes specific interactions between the reactants and the active site into account. This is demonstrated by the comparison of the various QSARs presented in chapter 7. The QM/MM parameters allow the inclusion of differently charged substituents, where the gas-phase parameters fail due to the absence of the stabilising effects of the surroundings.

All together, the research presented in this thesis has made a new contribution to the development and validation of computational models that can be used to address a major challenge in the present field of biochemistry, i.e. to obtain insight into enzymatic reaction mechanisms and enzyme activity on the basis of the structure of enzyme and substrate(s). Special emphasis has been on the application and validation of the QM/MM technique in the context of a QSAR approach. The investigations of this thesis provide a first survey of the possibilities of the QM/MM method with respect to the prediction of biochemical activity, taking explicitly into account the influence of the active site surroundings.

Samenvatting en conclusies

Samenvatting

Theoretische simulaties worden steeds belangrijker bij het leren begrijpen van de werking van enzymen. Het doel van het onderzoek in dit proefschrift is bij te dragen aan deze ontwikkeling door verschillende computermethoden toe te passen op drie enzymen van de β -ketoacidroute, en de modellen te valideren aan de hand van kwantitatieve structuur-activiteits relaties (QSAR). De verkregen modellen en QSARs leveren belangrijke informatie over het reactiemechanisme van de te bestuderen (snelheidsbepalende) stap in de reactiecyclus van de enzymen (**hoofdstuk 1**).

Twee van de enzymen die in dit proefschrift worden onderzocht zijn de flavine-afhankelijke monooxygenases *para*-hydroxybenzoeaat hydroxylase (PHBH) van *Pseudomonas fluorescens*, en fenol hydroxylase (PH) van *Trichosporon cutaneum*. Deze enzymen katalyseren de ortho-hydroxylering van respectievelijk *para*-hydroxybenzoeaat en fenol, waarbij respectievelijk catechol en protocatechuaat worden gevormd. Deze producten zijn centrale intermediaren in de afbraak van veel aromatische verbindingen. Als catechol of protocatechuaat eenmaal is gevormd wordt de benzeen-ring gesplitst tussen de twee hydroxyl-gesubstitueerde koolstof atomen. Deze zogenaamde intradiol splitsing wordt gekatalyseerd door respectievelijk catechol-1,2-dioxygenase (1,2-CTD), het derde enzym dat in dit proefschrift bestudeerd wordt, en protocatechuaat-3,4-dioxygenase (3,4-PCD), dat homoloog is aan 1,2-CTD.

Hoofdstukken 2 en 3 beschrijven theoretische principes en praktische aspecten van de computer-methodes die in dit proefschrift gebruikt worden. Quantum mechanische (QM) berekeningen kunnen worden gebruikt om eigenschappen van substraten en cofactoren, of nog iets uitgebreidere modellen van katalytische centra, te berekenen. Een bruikbare parameter voor enzymactiviteit die quantum mechanisch berekend kan

worden is de energie van de relevante frontier orbitaal van het reagerende molecuul. Deze frontier orbitaal kan de hoogst bezette moleculaire orbitaal (HOMO) zijn, die de nucleofiele reactiviteit van een reactant bepaalt, of de laagst onbezette moleculaire orbitaal (LUMO), die de electrofiele reactiviteit bepaalt. Een tweede bruikbare parameter is het energieverval tussen de reactanten en de producten van de betreffende reactiestap. Dit energieverval is, volgens de Brønsted theorie, gerelateerd aan de activeringsenergie. De uitkomsten van deze kwantummechanische berekeningen in de gasfase kunnen worden gebruikt als parameters voor een lineaire correlatie met de natuurlijke logaritme van de experimentele snelheidsconstanten voor de enzymatische omzetting van een serie homologe verbindingen.

In dit proefschrift is een belangrijke volgende stap gezet in het QSAR-onderzoek door het toepassen van een gecombineerde kwantummechanische/moleculair mechanische (QM/MM) methode. Met deze methode kunnen de kwantummechanische eigenschappen (en reactiepaden) van de reagerende verbindingen worden berekend in de eiwitomgeving. De omliggende eiwitatomen worden daarbij berekend op moleculair mechanische wijze. De elektrostatische en sterische effecten van deze eiwitatomen op het kwantummechanische gedeelte worden in de berekening meegenomen door enerzijds de moleculair mechanische atoomladingen toe te voegen aan de Hamiltoniaan van het quantummechanische systeem, en anderzijds door het toevoegen van Van der Waals interacties tussen de QM en MM atomen.

Catechol dioxygenase

Het reactiemechanisme van catechol-1,2-dioxygenase uit *Pseudomonas putida* is bestudeerd aan de hand van een QSAR benadering op basis van moleculaire orbitaal berekeningen in gasfase (**Hoofdstuk 4**). Catechol-1,2-dioxygenase katalyseert *ortho*-splitsing van de aromatische ring van catechol via het inbouwen van beide atomen van een molecuul zuurstof. Naast catechol, zet dit enzym ook verscheidene C4-gesubstitueerde catecholderivaten om. In dit onderzoek werden 4-methyl-, 4-chloor-, 4-broom-, 4,5-difluor- en 4-chloor-5-fluorcatechol verkregen, via de omzetting van de overeenkomstige fenolen met het enzym fenol hydroxylase. De snelheidsconstanten voor de omzetting van deze catecholderivaten door catechol-1,2-dioxygenase werden bepaald doormiddel van steady-state kinetiek metingen onder variërende zuurstof-concentraties en verzadigende catechol concentraties.

De cruciale stap in het reactiemechanisme van catechol-1,2-dioxygenase is de nucleofiele aanval van het substraat op een molecuul zuurstof. Daarom werd de energie van de hoogst bezette orbitaal (HOMO) van het catecholmolecuul, die een maat is voor de nucleofiele reactiviteit, gekozen als theoretische parameter voor een vergelijking met de experimentele snelheidsconstanten. Volgens de in de literatuur

voorgestelde reactiecyclus vindt deze nucleofiele aanval plaats nadat het di-anionische substraat op bidentate wijze gebonden is aan het ijzer atoom in het katalytische centrum van het enzym. Het effect van deze bidentate binding op de verdeling van de electronendichtheid op het substraatmolecuul is geschat op basis van een berekening aan het volledig ijzer-ligand-substraat complex. Deze berekening is uitgevoerd met een dichtheids-functionaal methode, die in het algemeen goede resultaten geeft voor systemen met overgangsmetalen. Het resultaat van de berekening gaf aan dat de lading op het substraat ongeveer $-1e$ is, in plaats van $-2e$, en dus gedeeltelijk gedelocaliseerd wordt over het gehele complex. Op basis van dit resultaat zijn de berekeningen aan de afzonderlijke substraat moleculen uitgevoerd voor de mono-anionische vorm, om zo goed mogelijk de situatie in het katalytisch centrum te benaderen.

De berekende HOMO-energieën bleken goed te correleren met de logaritme van de experimenteel bepaalde snelheidsconstanten voor de omzetting van de catecholen door 1,2-CTD. Dit wijst erop dat de snelheidsbepalende stap in het reactiemechanisme een nucleofiele reactie van het substraat betreft, en dat dit inderdaad de reactie met zuurstof zou kunnen zijn. Het kan echter niet worden uitgesloten dat andere stappen in de reactiecyclus ook afhangen van de nucleofiele reactiviteit van het substraat en dus worden beïnvloed door de substituenten. Daarom zijn ook de energieën van de verschillende intermediären in de reactiecyclus voor de verschillende substraten berekend. De (relatieve) energieveranderingen die berekend zijn voor de afzonderlijke reactiestappen met de verschillende substraten zijn vergeleken met de experimentele snelheidsconstanten. De resulterende correlaties wijzen erop dat, naast de nucleofiele reactie met zuurstof, ook de daarop volgende stap in de reactiecyclus, d.w.z. de omlegging naar een cyclisch anhydride intermediair, ook snelheidsbepalend zou kunnen zijn. De laatste twee stappen in de reactiecyclus konden echter worden uitgesloten als snelheidsbepalende stap.

De resultaten voor catechol-1,2-dioxygenase van *Pseudomonas putida* zijn vergeleken met gegevens over twee verschillende typen catechol-1,2-dioxygenase, een "normale" (type I) en een chloorcatechol dioxygenase (type II), uit *Pseudomonas sp.* B13. Op basis van deze vergelijking kon worden beargumenteerd dat het verschil in substraatvoorkeur van beide typen catechol dioxygenase wordt veroorzaakt door een verschillend effect van de substituenten op de zuurstofbinding en niet door een verschillend mechanisme van de snelheidsbeperkende stap.

Para-hydroxybenzooat hydroxylase

In **hoofdstuk 5, 6 en 7** wordt een QM/MM model beschreven voor de hydroxyleringsstap in de reactiecyclus van *p*-hydroxybenzooat hydroxylase, een flavine-afhankelijk monooxygenase. Eerst is onderzocht of met behulp van dit model

berekende activeringsenergieën de variatie in snelheidsconstanten voor de omzetting van een serie gefluorideerde substraten door PHBH kunnen verklaren (**hoofdstuk 5**). Reactiepaden werden berekend voor de electrofiële aanval van de intermediaire C4a-hydroperoxyflavine op het substraat. Deze stap was voorgesteld als snelheidsbepalend in de reactie van PHBH. De activeringsenergieën, die volgden uit de berekende energieprofielen voor deze stap, zijn verschillend voor de diverse substraten. Deze berekende activeringsenergieën bleken te correleren met de natuurlijke logaritme van de experimentele snelheidsconstanten voor de omzetting door PHBH. Deze correlatie bevestigt de hypothese dat de electrofiële aanval van de C4a-hydroperoxyflavine op het substraat snelheidsbeperkend is in de reactiecyclus.

De correlatie bevestigt ook dat het QM/MM model een realistische beschrijving geeft van de hydroxyleringsstap, omdat het een juist effect van de substraat substituenten op de activeringsenergie laat zien. Op basis van deze bevestiging, is in **hoofdstuk 6** het QM/MM model gebruikt voor een verdere analyse van de reactie. Dit geeft inzicht in het mechanisme waarmee het substraat en de cofactor worden geactiveerd om de electrofiële reactie te vergemakkelijken. Deprotonering van het substraat blijkt een significante verlaging van de berekende activeringsenergie tot gevolg te hebben. Dit effect van substraatdeprotonering op de reactie kan worden verklaard met behulp van frontier-orbitaal theorie. Substraat deprotonering verhoogt zowel de energie als de dichtheid van de bij de reactie betrokken HOMO-elektronen op het reagerende C3 atoom van het substraat. Ook is een beter begrip gevormd van de reactiviteit van de C4a-hydroperoxyflavine cofactor. Halverwege het berekende reactiepad treedt een verlenging van de peroxide zuurstof-zuurstof binding op, die samengaat met een toename van de LUMO dichtheid op de reagerende distale zuurstof. Dit suggereert dat het verlengen van de peroxide binding belangrijk is voor het activeren van de cofactor voor een electrofiële aanval op het substraat. Verder is het QM/MM model gebruikt voor het analyseren van de energetische effecten van de afzonderlijke aminozuren in het katalytische centrum op de hydroxyleringsreactie. De resultaten wijzen op mogelijke katalytische effecten van de carbonyl groep van Pro293, die specifiek de overgangstoestand stabiliseert, en van een water molecuul (nr. 717 uit de kristal structuur), die de toenemende negatieve lading op de proximale zuurstof van de flavine cofactor stabiliseert.

In **hoofdstuk 7** worden verschillende parameters gebruikt voor een uitgebreidere QSAR voor de hydroxyleringsstap in PHBH, waarin niet alleen substituties op het substraat, maar ook op de cofactor worden meegenomen. De QSARs die in dit hoofdstuk worden beschreven zijn gebaseerd op snelheidsconstanten voor de afzonderlijke hydroxyleringsstap (in plaats van overall snelheden) die zijn verkregen uit stopped-flow experimenten. Dit onderzoek maakt duidelijk dat parameters

afkomstig van berekeningen in de gasfase, d.w.z. het energieverval tussen HOMO en LUMO en de energie van de reactie, correleren met de experimentele snelheidsconstanten zolang het gaat om substraten en cofactoren die onderling een gelijke lading hebben. De negatief geladen 8-hydroxyflavine en 8-mercaptoflavine cofactoren resulteerden echter in afwijkende punten op de QSAR. Dit moet waarschijnlijk toegeschreven worden aan een vermindering van het effect van deze negatieve substituenten door het afschermende en stabiliserende effect van de eiwitomgeving in het katalytische centrum, waarmee geen rekening wordt gehouden in de gasfase berekeningen. De QM/MM berekeningen daarentegen, die dit effect bij benadering wel meenemen, maakten het mogelijk alle onderzochte substraat-cofactor combinaties, inclusief de geladen 8-hydroxyflavine en 8-mercaptoflavine cofactoren, in één QSAR op te nemen. Dit is een duidelijke illustratie van het voordeel van de QM/MM benadering, ten opzichte van berekeningen in gasfase, voor berekeningen aan enzymen.

Fenol hydroxylase

Hoofdstuk 8 beschrijft een QM/MM model voor fenol hydroxylase (PH) uit *Trichosporon cutaneum*. Evenals voor PHBH is de hydroxyleringsstap gesimuleerd voor een serie van gehalogeneerde fenol derivaten. De berekende activeringsenergieën correleren goed met de logaritme van de experimentele snelheidsconstanten. Deze correlatie wijst erop dat de electrofiele aanval van de C4a-hydroperoxyflavine op het substraat snelheidsbeperkend is in de reactiecyclus bij een pH 7.6 en 25°C. Dit hoofdstuk gaat ook in op het effect van het wel of niet geprotoneerd zijn van het substraat tijdens de hydroxylering. De mogelijke rol van substraat-deprotonering is voor PH nog veel minder duidelijk dan voor PHBH. In het QM/MM model is de protonoverdracht van fenol naar de mogelijke katalytische base, Asp54, bestudeerd. Hiertoe is een 2-dimensionaal potentiaal-oppervlak berekend, als functie van de twee reactiecoördinaten voor hydroxylering en protonoverdracht. Het resultaat wijst erop dat het mechanisme, waarbij proton-overdracht plaats vindt voordat hydroxylering optreedt, energetisch het meest gunstig is. Deze deprotonering vindt in PH waarschijnlijk op een later moment in de reactiecyclus plaats (bijv. net voor de hydroxylering), dan in PHBH, waar het substraat direct na binding in het katalytisch centrum wordt gedeprotoneerd.

Een verdere analyse van het QM/MM model geeft inzicht in het effect van de eiwitomgeving op de berekende reactiestappen. Een aantal katalytische effecten op de hydroxyleringsstap, zoals die van een carbonylgroep van proline en van een (kristal-) water in de buurt van de reactanten, komen sterk overeen met die in PHBH.

Conclusies

Het onderzoek in dit proefschrift heeft geleid tot gevalideerde theoretische modellen van de snelheidsbeperkende reactiestappen van drie enzymen. Met deze modellen zijn parameters berekend die op lineaire wijze correleren met de logaritme van de experimentele snelheidsconstanten van de enzymatische omzetting van reeksen substraten.

De QSARs op basis van de parameters verkregen uit berekeningen in de gasfase, d.w.z. de energie van frontier-orbitalen en reactie-energieën, geven inzicht in de aard van de snelheidsbeperkende reactiestap en zijn een hulpmiddel bij het identificeren van de snelheidsbepalende stap in een reactiecyclus. De parameters houden echter geen rekening met omgevingseffecten, zoals het afschermen en stabiliseren van lokale ladingen (die kunnen veranderen tijdens de reactie) door residuen in het katalytisch centrum van een enzym.

Het effect van de gehele eiwit omgeving kan (bij benadering) worden meegerekend door middel van de gecombineerde quantummechanisch / moleculair mechanische methode. De reactiepaden die met deze QM/MM methode kunnen worden berekend geven een gedetailleerd inzicht in het mechanisme van de gesimuleerde reacties. Het maakt het mogelijk onstabiele reactie-intermediären (bijv. het hydroxycyclohexadienon in PHBH en PH) te karakteriseren en om de rol van afzonderlijke aminozuur-residuen in de katalyse te analyseren. Deze gedetailleerde informatie over het mechanisme van enzymatische reacties is vaak moeilijk af te leiden uit experimenten.

De reactiepaden die met de QM/MM methode berekend worden geven, met de huidige stand van zaken, geen absolute waarden voor de activeringsenergieën. Dit proefschrift laat echter duidelijk zien dat de met QM/MM berekende energie-barrières wel goed gebruikt kunnen worden als relatieve parameters voor enzymatische activiteit. Dit blijkt uit de correlaties tussen berekende activeringsenergieën en de logaritme van experimentele snelheidsconstanten voor enzymatische omzettingen. De QM/MM methode maakt het in principe mogelijk QSARs op te stellen met een grotere variatie aan verbindingen, omdat specifieke interacties tussen de reagerende verbindingen en de eiwit omgeving worden meegerekend. Dit blijkt uit de vergelijking van verschillende QSARs in hoofdstuk 7. Met de QM/MM parameters kunnen verbindingen (in dit geval cofactoren) met een verschillende totale lading in één QSAR opgenomen worden, terwijl dit met gasfase parameters sterke afwijkingen oplevert. Dit wordt veroorzaakt doordat in de gasfase parameters geen rekening wordt gehouden met het stabiliserende effect van de omgeving op de extra negatieve lading.

Het onderzoek in dit proefschrift heeft een nieuwe bijdrage geleverd aan de ontwikkeling en validatie van computermodellen die inzicht kunnen geven in enzymatische reactiemechanismen en enzymactiviteit op basis van de structuren van het enzym en de substraten. Hierbij heeft het accent gelegen bij de toepassing en validatie van de QM/MM methode in de context van een QSAR benadering. De studies in dit proefschrift vormen een eerste verkenning van de mogelijkheden van de QM/MM methode met betrekking tot de voorspelling van biochemische activiteit, waarbij de invloed van het katalytische centrum expliciet in de berekeningen wordt meegenomen.

Symbols and abbreviations

ABNR	Adopted Basis Newton-Raphson	LUMO	lowest unoccupied molecular orbital
AM1	Austin Model 1	MM	molecular mechanical
AMP	adenosine monophosphate	MNDO	modified neglect of differential overlap
B3LYP	Becke-3-Lee-Yang-Parr	MO	molecular orbital
BP86	Becke-Perdew '86	MP2	Møller-Plesset (second order)
CI	configuration interaction	MP4	Møller-Plesset (fourth order)
CHARMM	Chemistry at HARvard Molecular Mechanics	NADPH	nicotinamide adenine dinucleotide phosphate (reduced)
CoA	coenzyme A	NADP+	nicotinamide adenine dinucleotide phosphate (oxidised)
COM	centre of mass	NMR	nuclear magnetic resonance
CPU	central processing unit	3,4-PCD	protocatechuate-3,4-dioxygenase
1,2-CTD	catechol-1,2-dioxygenase	PDB	protein databank
3D	3-dimensional	PH	phenol hydroxylase
DFT	density functional theory	PHBH	<i>para</i> -hydroxybenzoate hydroxylase
DNA	deoxyribose nucleic acid	PM3	parametric method 3
E_{act}	activation energy	QM	quantum mechanical
E_{HOMO}	energy of the highest occupied molecular orbital	QM/MM	combined quantum mechanical / molecular mechanical
E_{LUMO}	energy of the lowest unoccupied molecular orbital	QSAR	quantitative structure activity relationship
EDTA	ethylene-diamine-tetra-acetate	r	reaction coordinate
eV	electron volt	R	regression coefficient
FAD	flavin adenine dinucleotide	RMSD	root-mean-squared difference RMSD
a-FAD	arabino flavin adenine dinucleotide	SBMD	stochastic boundary molecular dynamics
H	Hamiltonian	SCF	self-consistent field
H_f	heat of formation	TS	transition state
HF	Hartree-Fock	UHF	unrestricted Hartree-Fock
HOMO	highest occupied molecular orbital	VDW	Van der Waals
HPLC	high pressure liquid chromatography	ZPE	zero-point energy
kcal	kilo calory (= 4184 joules)		
LCAO	linear combination of atomic orbitals		

References

- Anderson, R. F., Patel, K. B. and Stratford, M. R. L. (1987) Absorption spectra of substrates for *p*-hydroxybenzoate hydroxylase following electrophilic attack of the $^{\circ}\text{OH}$ radical in the 3 position. *J. Biol. Chem.* 262, 17475-17479
- Anderson, R. F., Patel, K. B. and Vojnovic, B. (1991) Absorption spectra of radical forms of 2,4-dihydroxybenzoic acid, a substrate for *p*-hydroxybenzoate hydroxylase. *J. Biol. Chem.* 266, 13086-13090
- Atkins, P. W. (1983) *Molecular quantum mechanics*. University Press, Oxford
- Bash, P. A., Field, M. J., Davenport, R. C., Petsko, G. A., Ringe, D. and Karplus, M. (1991) Computer simulation and analysis of the reaction pathway of triosephosphate isomerase. *Biochemistry* 30, 5826-5832
- Boersma, M. G., Dinariyeva, T. Y., Middelhoven, W. J., Van Berkel, W. J. H., Doran, J., Vervoort, J. and Rietjens, I. M. C. M. (1998) ^{19}F Nuclear magnetic resonance as a tool to investigate microbial degradation of fluorophenols to fluorocatechols and fluoromuconates. *Appl. Environ. Microbiol.* 64, 1256-1263
- Bolognesi, M., Ghisla, S. and Incoccia, L. (1978) *Acta Crystallogr., Sect. B* 34, 821-828
- Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M. (1983) CHARMM: A program for macromolecular energy, minimization and dynamics calculations. *J. Comp. Chem.* 4, 187-217
- Brooks, C. L., III and Karplus, M. (1989) Solvent effects on protein motion and protein effects on solvent motion. *J. Mol. Biol.* 208, 159-181
- Cass, A. E. G., Ribbons, D. W., Rossiter, J. T. and Williams, S. R. (1987) Biotransformation of aromatic compounds. Monitoring fluorinated analogues by NMR. *FEBS lett.* 220, 353-357
- Chang, R. (1981) *Physical chemistry with applications to biological systems*. Macmillan Publishing Co., Inc., New York
- Cnubben, N. H. P., Peelen, S., Borst, J. W., Vervoort, J., Veeger, C. and Rietjens, I. M. C. M. (1994) Molecular orbital based quantitative structure-activity relationship for the cytochrome P450-catalyzed 4-hydroxylation of halogenated anilines. *Chem. Res. Toxicol.* 7, 590 - 598
- Cremer, D. (1983) General and theoretical aspects of the peroxide group. In: S. Patai (eds.) *The Chemistry of Functional Groups, Peroxides*. John Wiley & Sons Ltd, pp.
- Cunningham, M. A., Ho, L. L., Nguyen, D. T., Gillilan, R. E. and Bash, P. A. (1997) Simulation of the enzyme reaction mechanism of malate dehydrogenase. *Biochemistry* 36, 4800-4816

- Dagley, S. (1987) Lessons from biodegradation. *Ann. Rev. Microbiol.* 41, 1-23
- Detmer, K. and Massey, V. (1985) Effect of substrate and pH on the oxidative half-reaction of phenol hydroxylase. *J. Biol. Chem.* 260, 5998-6005
- Dewar, M. J. S. (1975) Quantum organic chemistry. *Science* 187, 1037-1044
- Dewar, M. J. S. and Thiel, W. (1977) Ground states of molecules. 38. The MNDO method. Approximations and parameters. *J. Am. Chem. Soc.* 99, 4899-4907
- Dewar, M. J. S. and Yuan, Y. (1990) AM1 Parameters for sulfur. *Inorg. Chem.* 29, 3881-3890
- Dewar, M. J. S. and Zoebisch, E. G. (1988) Extension of AM1 to the halogens. *J. Mol. Struct (Theochem)* 180, 1-21
- Dewar, M. J. S., Zoebisch, E. G., Healy, E. F. and Stewart, J. J. P. (1985) AM1: A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* 107, 3902-3909
- Dimarco, A. A., Averhoff, B. A., Kim, E. E. and Ornston, L. N. (1993) Evolutionary divergence of *pobA*, the structural gene encoding *p*-hydroxybenzoate hydroxylase in an *Acinetobacter calcoaceticus* strain well-suited for genetic analysis. *Gene* 125, 25-33
- Dorn, E. and Knackmuss, H. (1978) Chemical structure and biodegradability of halogenated aromatic compounds. *Biochem. J.* 174, 85-94
- Eckstein, J. W., Hastings, J. W. and Ghisla, S. (1993) Mechanism of bacterial bioluminescence: 4a,5-dihydroflavin analogs as models for luciferase hydroperoxide intermediates and the effect of substituents at the 8-position of flavin on luciferase kinetics. *Biochemistry* 32, 404-411
- Enroth, C., Neujahr, H., Schneider, G. and Lindqvist, Y. (1998) The crystal structure of phenol hydroxylase in complex with FAD and phenol provides evidence for a concerted conformational change in the enzyme and its cofactor during catalysis. *Structure* 6, 605-617
- Entsch, B., Ballou, D. P. and Massey, V. (1976) Flavin-oxygen derivatives involved in hydroxylation by *p*-hydroxybenzoate hydroxylase. *J. Biol. Chem.* 251, 2550-2563
- Entsch, B., Palfey, B. A., Ballou, D. P. and Massey, V. (1991) Catalytic function of tyrosine residues in *p*-hydroxybenzoate hydroxylase as determined by the study of site-directed mutants. *J. Biol. Chem.* 266, 17341-17349
- Entsch, B. and Van Berkel, W. J. H. (1995) Structure and mechanism of *p*-hydroxybenzoate hydroxylase. *FASEB J.* 9, 476-483
- Eppink, M. H., Schreuder, H. A. and Van Berkel, W. J. H. (1997a) Identification of a novel conserved sequence motif in flavoprotein hydroxylases with a putative dual function in FAD/NAD(P)H binding. *Protein Sci.* 6, 2454-2458
- Eppink, M. H. M., Boeren, S. A., Vervoort, J. and Van Berkel, W. J. H. (1997b) Purification and properties of 4-hydroxybenzoate 1-hydroxylase (decarboxylating), a novel flavin adenine dinucleotide - dependent monooxygenase from *Candida parapsilosis* CBS604. *J. Bacteriology* 179, 6680-6687
- Eppink, M. H. M., Schreuder, H. A. and Van Berkel, W. J. H. (1998) Interdomain binding of NADPH in *p*-hydroxybenzoate hydroxylase as suggested by kinetic, crystallographic and modeling studies of histidine 162 and arginine 269 variants. *J. Biol. Chem.* 273, 21031-21039
- Eppink, M. J. M., Van Berkel, W. J. H., Teplakov, A. and Schreuder, H. A. (1999) "Unactivated" *p*-hydroxybenzoate hydroxylase: Crystal structures of the free enzyme and the enzyme-benzoate complex.

- Eschrich, K., Van Der Bolt, F. J. T., De Kok, A. and Van Berkel, W. J. H. (1993) Role of Tyr201 and Tyr385 in substrate activation by *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens*. *Eur. J. Biochem.* 216, 137-146
- Eurenius, K. P., Chatfield, D. C., Brooks, B. R. and Hodoscek, M. (1996) Enzyme Mechanisms with Hybrid Quantum and Molecular Mechanical Potentials. I. Theoretical Considerations. *Int. J. Quantum Chem.* 60, 1189-1200
- Felton, R. H., Barrow, W. L., May, S. W., Sowell, A. L., Goel, S., Bunker, G. and Stern, E. A. (1982) EXAFS and Raman evidence for histidine binding at the active site of protocatechuate 3,4-dioxygenase. *J. Am. Chem. Soc.* 104, 6132-6134
- Field, M. J., Bash, P. A. and Karplus, M. (1990) A combined quantum mechanical and molecular mechanical potential for molecular dynamics simulations. *J. Comp. Chem.* 11, 700-733
- Fleming, I. (1976) *Frontier Orbitals and Organic Chemical Reactions*. John Wiley and sons., New York
- Foresman, J. B. and Frisch, A. E. (1996) *Exploring chemistry with electronic structure methods*. Gaussian, Inc., Pittsburgh, PA
- Frauenfelder, H., Sligar, S. G. and Wolynes, P. G. (1991) The energy landscapes and motions of proteins. *Science* 254, 1598-1603
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Gill, P. M. W., Johnson, B. G., Robb, M. A., Cheeseman, J. R., Keith, T., Petersson, G. A., Montgomery, J. A., Raghavachari, K., Al-Laham, M. A., Zakrzewski, V. G., Ortiz, J. V., Foresman, J. B., Cioslowski, J., Stefanov, B. B., Nanayakkara, A., Challacombe, M., Peng, C. Y., Ayala, P. Y., Chen, W., Wong, M. W., Andres, J. L., Replogle, E. S., Gomperts, R., Martin, R. L., Fox, D. J., Binkley, J. S., Defrees, D. J., Baker, J., Stewart, J. P., Head-Gordon, M., Gonzalez, C. and P. J. (1995) Gaussian94, Revision E.2. Gaussian, Inc., Pittsburgh PA
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., Zakrzewski, V. G., Montgomery, J. A. J., Stratmann, R. E., Burant, J. C., Dapprich, S., Millam, J. M., Daniels, A. D., Kudin, K. N., Strain, M. C., Farkas, O., Tomasi, J., Barone, V., Cossi, M., Cammi, R., Mennucci, B., Pomelli, C., Adamo, C., Clifford, S., Ochterski, J., Petersson, G. A., Ayala, P. Y., Cui, Q., Morokuma, K., Malick, D. K., Rabuck, A. D., Raghavachari, K., Foresman, J. B., Cioslowski, J., Ortiz, J. V., Stefanov, B. B., Liu, G., Liashenko, A., Piskorz, P., Komaromi, I., Gomperts, R., 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., Andres, J. L., Gonzalez, C., Head-Gordon, M., Replogle, E. S. and Pople, J. A. (1998) Gaussian98, Revision A.6. Gaussian, Inc., Pittsburgh PA
- Fujisawa, H. and Hayaishi, O. (1968) Protocatechuate 3,4-dioxygenase. *J. Biol. Chem.* 243, 2673-2681
- Fujiwara, M., Golovleva, L. A., Saeki, Y., Nozaki, M. and Hayaishi, O. (1975) Extradiol cleavage of 3-substituted catechols by an intradiol dioxygenase, pyrocatechase, from a *Pseudomonas*. *J. Biol. Chem.* 250, 4848-4855
- Fukui, K. (1981) The path of chemical reaction - the IRC approach. *Acc. Chem. Res.* 14, 363
- Fukui, K., Yonezawa, T. and Shingu, H. (1952) A molecular orbital theory of reactivity in aromatic hydrocarbons. *J. Chem. Phys.* 20, 722-725

- Gaal, A. and Neujahr, H. Y. (1979) Metabolism of phenol and resorcinol in *Trichosporon cutaneum*. *J. Bacteriology* 137, 13-21
- Gatti, D. L., Entsch, B., Ballou, D. P. and Ludwig, M. L. (1996) PH-dependent structural changes in the active site of *p*-hydroxybenzoate hydroxylase point to the importance of proton and water movements during catalysis. *Biochemistry* 35, 567-578
- Gatti, D. L., Palfey, B. A., Lah, M. S., Entsch, B., Massey, V., Ballou, D. P. and Ludwig, M. L. (1994) The mobile flavin of 4-OH Benzoate Hydroxylase. *Science* 266, 110-113
- Gonzalez, C. and Schlegel, H. B. (1989) An improved algorithm for reaction path following. *J. Chem. Phys.* 90, 2154-2161
- Grootenhuys, P. D. J. and Van Galen, P. J. M. (1995) Correlation of binding affinities with non-bonded interaction energies of thrombin-inhibitor complexes. *Acta Crystallogr. D* 51, 560-566
- Harayama, S., Kok, M. and Neidle, E. L. (1992) Functional and evolutionary relationships among diverse oxygenases. *Annu. Rev. Microbiol.* 46, 565-601
- Harrison, M. J., Burton, N. A. and Hillier, I. H. (1997) Catalytic mechanism of the enzyme papain: Predictions with a hybrid quantum mechanical/molecular mechanical potential. *J. Am. Chem. Soc.* 119, 12285-12291
- Hartsough, D. S. and Merz, K. M. (1995) Dynamic force field models: Molecular dynamics simulations of human carbonic anhydrase II using a quantum mechanical/molecular mechanical coupled potential. *J. Phys. Chem.* 99, 11266-11275
- Harwood, C. S. and Parales, R. E. (1996) The β -ketoadipate pathway and the biology of self-identity. *Annu. Rev. Microbiol.* 50, 553-590
- Hasford, J. J. and Rizzo, C. J. (1998) Linear free energy substituent effect on flavin redox chemistry. *J. Am. Chem. Soc.* 120, 2251-2255
- Hehre, W. J. (1997) A guide to molecular mechanics and molecular orbital calculations in Spartan. Wavefunction, Inc., Irvine
- Hosokawa, K. and Stanier, R. Y. (1966) Crystallization and properties of *p*-hydroxybenzoate hydroxylase from *Pseudomonas putida*. *J. Biol. Chem.* 241, 2453
- Howell, L. G., Spector, T. and Massey, V. (1972) Purification and properties of *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens*. *J. Biol. Chem.* 247, 4340
- Husain, M., Entsch, B., Ballou, D. P., Massey, V. and Chapman, J. P. (1980) Fluoride elimination from substrates in hydroxylation reactions catalyzed by *p*-hydroxybenzoate hydroxylase. *J. Biol. Chem.* 255, 4189-4197
- Jurema, M. W. and Shields, G. C. (1993) Ability of the PM3 quantum-mechanical method to model intermolecular hydrogen bonding between neutral molecules. *J. Comp. Chem.* 14, 89-104
- Kabsch, W. (1976) A solution for the best rotation to relate two sets of vectors. *Acta Crystallogr. A* 32, 922-923
- Kemal, C. and Bruice, T. (1979) Transfer of O₂ from a 4a-hydroperoxyflavin anion to a phenolate ion. A flavin catalyzed dioxygenation reaction. *J. Am. Chem. Soc.* 101, 1635-1638
- Kemal, C., Chan, T. W. and Bruice, T. C. (1977) Reaction of ³O₂ with dhydroflavins. 1. N^{3,5}-dimethyl-1,5-dihydrolumiflavin and 1,5-dihydroisoalloxazines. *J. Am. Chem. Soc.* 99, 7272-7286

- Keum, S. R., Gregory, D. H. and Bruice, T. C. (1990) Oxidation of aminophenols by 4a-hydroperoxy-5-ethylthymine anion: Flavoenzyme hydroxylase mechanism. *J. Am. Chem. Soc.* 112, 2711-2715
- King, E. L. and Altman, C. (1956) A schematic method of deriving the rate laws for enzyme-catalyzed reactions. *J. Phys. Chem.* 60, 1375-1378
- Klopman, G. (1968) Chemical reactivity and the concept of charge- and frontier-controlled reactions. *J. Am. Chem. Soc.* 90, 223-234
- Kobayashi, K., Katayama-Hirayama, K. and Tobita, S. (1997) Hydrolytic dehalogenation of 4-chlorobenzoic acid by an *Acinetobacter* sp. *J. Gen. Appl. Microbiol.* 43, 105-108
- Koch, H. F., Mishima, M., Zuilhof, H. (1998) Proton transfer between carbon acids and methoxide: studies in methanol, the gas phase and *ab initio* MO calculations. *Ber. Bunsenges. Phys. Chem.* 102, 567-572
- Kojima, Y., Fujisawa, H., Nakazawa, A., Nakazawa, T., Kanetsuna, F., Taniuchi, H., Nozaki, M. and Hayaishi, O. (1967) Studies on pyrocatechase. Purification and spectral properties. *J. Biol. Chem.* 242, 3270-3278
- Kurzynski, M. (1998) A synthetic picture of intramolecular dynamics of proteins. Towards a contemporary statistical theory of biochemical processes. *Progr. Biophys. Molec. Biol.* 69, 23-82
- Lipscomb, J. D. and Orville, A. M. (1992) Mechanistic aspects of dihydroxybenzoate dioxygenases. In: H. Sigel and A. Sigel (eds.) *Metal Ions in Biological Systems*. Marcel Dekker Inc., New York, pp. 243-298
- Lyne, P., Mulholland, A. J. and Richards, W. G. (1995) Insights into chorismate mutase catalysis from a combined QM/MM simulation of the enzyme reaction. *J. Am. Chem. Soc.* 117, 11345-11350
- Maeda-Yorita, K. and Massey, V. (1993) On the reaction mechanism of phenol hydroxylase: New information obtained by correlation of fluorescence and absorbance stopped flow studies. *J. Biol. Chem.* 268, 4134-4144
- Marcus, R. A. (1968) Theoretical relations among rate constants, barriers and Brønsted slopes of chemical reactions. *J. Phys. Chem.* 72, 891-899
- Marks, T. S., Wait, R., Smith, A. R. W. and Quirk, A. V. (1984) The origin of the oxygen incorporated during the dehalogenation/hydroxylation of 4-chlorobenzoate by an *Arthrobacter* sp. *Biochem. Biophys. Res. Commun.* 124, 669-674
- McCammon, J. A. and Harvey, S. C. (1987) *Dynamics of proteins and nucleic acids*. University Press, Cambridge
- Mulholland, A. J. and Karplus, M. (1996) Computer Modelling of Biological Molecules. *Biochem. Soc. Trans.* 24, 247-254
- Mulholland, A. J. and Richards, W. G. (1997) Acetyl-CoA enolization in citrate synthase: A quantum mechanical/molecular mechanical (QM/MM) study. *Proteins* 27, 9-25
- Mulholland, A. J. and Richards, W. G. (1998a) Calculations on the substrates of citrate synthase I. Oxaloacetate. *J. Mol. Struct. (Theochem)* 429, 13-21
- Mulholland, A. J. and Richards, W. G. (1998b) Modeling enzyme reaction intermediates and transition states: citrate synthase. *J. Phys. Chem. B* 102, 6635-6646

- Nakai, C., Horiike, K., Kuramitsu, S., Kagamiyama, H. and Nozaki, M. (1990) Three isozymes of catechol 1,2-dioxygenase (pyrocatechase), $\alpha\alpha$, $\alpha\beta$, and $\beta\beta$, from *Pseudomonas arvilla* C-1. *J. Biol. Chem.* 265, 660-665
- Nakai, C., Nakazawa, T., Nozaki, M. (1988) Purification and properties of catechol 1,2-dioxygenase (Pyrocatechase) from *Pseudomonas putida* mt-2 in comparison with that from *Pseudomonas arvilla* C-1. *Arch. Biochem. Biophys.* 267, 701-713
- Nakamura, S., Ogura, Y., Yano, K., Higashi, N. and Arima, K. (1970) Kinetic studies on the reaction mechanism of *p*-hydroxybenzoate hydroxylase. *Biochemistry* 9, 3235
- Neujahr, H. Y. and Gaal, A. (1973) Phenol hydroxylase from yeast. Purification and properties of the enzyme from *Trichosporon cutaneum*. *Eur. J. Biochem.* 35, 386-400
- Neujahr, H. Y. and Kjellén, K. G. (1978) Phenol hydroxylase from yeast. Reaction with phenol derivatives. *J. Biol. Chem.* 253, 8835-8841
- Ohlendorf, D. H., Orville, A. M. and Lipscomb, J. D. (1994) Structure of protocatechuate 3,4-dioxygenase from *Pseudomonas aeruginosa* at 2.15 Å resolution. *J. Mol. Biol.* 244, 586-608
- Ornston, L. N. and Stanier, R. Y. (1964) Mechanism of B-ketoadipate formation by bacteria. *Nature* 206, 1279-1283
- Ortiz Maldonado, M., Ballou, D. P. and Massey, V. (1997) Leaving group tendencies of 8-substituted flavin-C4a-alkoxides and the mechanism of hydroxylation catalyzed by *p*-hydroxybenzoate hydroxylase. In: (eds.) Flavins and flavoproteins. University of Calgary Press, Calgary, pp. 323-326
- Ortiz-Maldonado, M., Ballou, D. P. and Massey, V. (1999) Use of free energy relationships to probe the individual steps of hydroxylation of *p*-hydroxybenzoate hydroxylase: studies with a series of 8-substituted flavins. *Biochemistry* 38, 8124-8137
- Orville, A. M. and Lipscomb, J. D. (1989) Binding of isotopically labeled substrates, inhibitors, and cyanide by protocatechuate 3,4-dioxygenase. *J. Biol. Chem.* 264, 8791-8801
- Orville, A. M., Lipscomb, J. D. and Ohlendorf, D. H. (1997) Crystal structures and substrate analog complexes of protocatechuate 3,4-dioxygenase: Endogenous Fe^{3+} ligand displacement in response to substrate binding. *Biochemistry* 36, 10052-10066
- Palfey, B. A. and Massey, V. (1998) Flavin-dependent enzymes. In: M. Sinnott (eds.) Radical reactions and oxidation/reduction. Academic press limited, London, pp. 83-154
- Palfey, B. A., Moran, G. R., Entsch, B., Ballou, D. P. and Massey, V. (1999) Substrate recognition by "password" in *p*-hydroxybenzoate hydroxylase. *Biochemistry* 38, 1153-1158
- Peelen, S., Rietjens, I. M. C. M., Boersma, M. G. and Vervoort, J. (1995) Conversion of phenol derivatives to hydroxylated products by phenol hydroxylase from *Trichosporon cutaneum*. A comparison of regioselectivity and rate of conversion with calculated molecular orbital substrate characteristics. *Eur. J. Biochem.* 227, 284-291
- Peelen, S., Rietjens, I. M. C. M., Van Berkel, W. J. H., Van Workum, W. A. and Vervoort, J. (1993) 19F-NMR study on the pH-dependent regioselectivity and rate of the ortho-hydroxylation of 3-fluorophenol by phenol hydroxylase from *Trichosporon cutaneum*. Implications for the reaction mechanism. *Eur. J. Biochem.* 218, 345-353

- Peräkylä, M. and Pakkanen, T. A. (1993) *Ab initio* models for receptor-ligand interactions in proteins. 3. Model assembly study of the proton transfer in the hydroxylation step of the catalytic mechanism of *p*-hydroxybenzoate hydroxylase. *J. Am. Chem. Soc.* 115, 10958-10963
- Powlowski, J. B. and Dagley, S. (1985) *b*-Ketoacid pathway in *Trichosporon cutaneum* modified for methyl substituted metabolites. *J. Bacteriol.* 163, 1126-1135
- Quanta (1993) Molecular Simulations Inc., 200 Fifth Avenue, Waltham, MA.
- Que, L. (1989) The catechol dioxygenases. In: T. M. Loehr (eds.) *Iron carriers and iron proteins*. VCH, New York, pp. 467-524
- Que, L. and Epstein, R. M. (1981) Resonance Raman studies on protocatechuate 3,4-dioxygenase-inhibitor complexes. *Biochemistry* 20, 2545-2549
- Que, L. and Ho, R. Y. N. (1996) Dioxygen activation by enzymes with mononuclear non-heme iron active sites. *Chem. Rev.* 96, 2606-2624
- Que, L., Lipscomb, J. D., Zimmermann, R., Münck, E., Orme-Johnson, N. R. and Orme-Johnson, W. H. (1976) Mössbauer and EPR spectroscopy on protocatechuate 3,4-dioxygenase from *Pseudomonas aeruginosa*. *Biochim. Biophys. Acta* 452, 320-334
- Reineke, W. and Knackmuss, H. J. (1988) Microbial degradation of haloaromatics. *Ann. Rev. Microbiol.* 42, 263-287
- Ridder, L., Mulholland, A. J., Rietjens, I. M. C. M. and Vervoort, J. (1999a) Combined quantum mechanical and molecular mechanical reaction pathway calculation for aromatic hydroxylation by *p*-hydroxybenzoate-3-hydroxylase. *J. Mol. Graphics Mod.* 17, 163-175
- Ridder, L., Mulholland, A. J., Vervoort, J. and Rietjens, I. M. C. M. (1998) Correlation of calculated activation energies with experimental rate constants for an enzyme catalyzed aromatic hydroxylation. *J. Am. Chem. Soc.* 120, 7641-7642
- Ridder, L., Zuilhof, H., Vervoort, J. and Rietjens, I. M. C. M. (1999b) Computational methods in flavin research. In: S. K. Chapman and G. A. Reid (eds.) *Meth. Mol. Biol.: Flavoprotein Protocols*. Humana Press Inc., NJ, pp. 207-228
- Rietjens, I. M. C. M., Soffers, A.E.M.F., Hooiveld, G.J.E.J., Veeger, C., Vervoort, J. (1995) Quantitative structure-activity relationships based on computer calculated parameters for the overall rate of glutathione *S*-transferase catalyzed conjugation of a series of fluoronitrobenzenes. *Chem. Res. Toxicol.* 8, 481-488
- Sakurada, J., Sekiguchi, R., Sato, K. and Hosoya, T. (1990) Kinetic and molecular orbital studies on the rate of oxidation of monosubstituted phenols and anilines by horseradish peroxidase compound II. *Biochemistry* 29, 4093-4098
- Salem, L. (1968) Intermolecular orbital theory of the interaction between conjugated systems. *J. Am. Chem. Soc.* 90, 543-566
- Schlömann, M. (1994) Evolution of chlorocatechol catabolic pathways. *Biodegradation* 5, 301-321
- Schlömann, M., Fischer, P., Schmidt, E. and Knackmuss, H.-J. (1990) Enzymatic formation, stability and spontaneous reactions of 4-fluoromuconolactone, a metabolite of the bacterial degradation of 4-fluorobenzoate. *J. Bacteriol.* 172, 5119-5129
- Schreuder, H. A., Hol, W. G. and Drenth, J. (1988a) Molecular modeling reveals the possible importance of a carbonyl oxygen binding pocket for the catalytic mechanism of *p*-hydroxybenzoate hydroxylase. *J. Biol. Chem.* 263, 3131-3136

- Schreuder, H. A., Hol, W. G. J. and Drenth, J. (1990) Analysis of the active site of the flavoprotein *p*-hydroxybenzoate hydroxylase and some ideas with respect to its reaction mechanism. *Biochemistry* 29, 3101-3108
- Schreuder, H. A., Mattevi, A., Obmolova, G., Kalk, K. H., Hol, W. G., Van der Bolt, F. J. and Van Berkel, W. J. (1994) Crystal structures of wild-type *p*-hydroxybenzoate hydroxylase complexed with 4-aminobenzoate, 2,4-dihydroxybenzoate, and 2-hydroxy-4-aminobenzoate and of the Tyr222Ala mutant complexed with 2-hydroxy-4-aminobenzoate. Evidence for a proton channel and a new binding mode of the flavin ring. *Biochemistry* 33, 10161-10170
- Schreuder, H. A., Prick, P. A. J., Wierenga, R. K., Vriend, G., Wilson, K. S., Hol, W. G. J. and Drenth, J. (1989) Crystal structure of the *p*-hydroxybenzoate hydroxylase-substrate complex refined at 1.9 Å resolution: Analysis of the enzyme-substrate and enzyme-product complexes. *J. Mol. Biol.* 208, 679-696
- Schreuder, H. A., Van der Laan, J. M., Hol, W. G. and Drenth, J. (1988b) Crystal structure of *p*-hydroxybenzoate hydroxylase complexed with its reaction product 3,4-dihydroxybenzoate. *J. Mol. Biol.* 199, 637-648
- Schreuder, H. A., Van der Laan, J. M., Hol, W. G. J. and Drenth, J. (1991) The structure of *p*-hydroxybenzoate hydroxylase. In: F. Müller (eds.) *Chemistry and Biochemistry of Flavoproteins*. CRC Press, Boca Raton, pp. 31-64
- Schreuder, H. A., Van der Laan, J. M., Swarte, M. B. A., Kalk, K. H., Hol, W. G. J. and Drenth, J. (1992) Crystal structure of the reduced form of *p*-hydroxybenzoate hydroxylase refined at 2.3 Å resolution. *Proteins: struct. func. genet.* 14, 178-190
- Shu, L., Chiou, Y. M., Orville, A. M., Miller, M. A., Lipscomb, J. D. and Que, L., Jr. (1995) X-ray absorption spectroscopic studies of the Fe(II) active site of catechol 2,3-dioxygenase. Implications for the extradiol cleavage mechanism. *Biochemistry* 34, 6649-6659
- Soffers, A. E. M. F., Ploemen, J. H. T. M., Moonen, M. J. H., Wobbes, T., Van Ommen, B., Vervoort, J., Van Bladeren, P. J. and Rietjens, I. M. C. M. (1996) Regioselectivity and quantitative structure-activity relationships for the conjugation of a series of fluoronitrobenzenes by purified glutathione *S*-transferase enzymes from rat and man. *Chem. Res. Toxicol.* 9, 638 - 646
- Spartan 5.1, Wavefunction Inc, 18401 Von Karman Avenue, Suite 370, Irvine, CA 92612
- Spector, T. and Massey, V. (1972) Studies on the effector specificity of *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens*. *J. Biol. Chem.* 247, 4679
- Stewart, J. J. P. (1989) Optimization of parameters for semiempirical methods I. Method. *J. Comp. Chem.* 10, 209-220
- Szabo, A. and Ostlund, N. S. (1982) *Modern quantum chemistry*. Macmillan Publishing Co., Inc., New York
- True, A. E., Orville, A. M., Pearce, L. L., Lipscomb, J. D. and Que, L. (1990) An EXAFS study of the interaction of substrate with the ferric active site of protocatechuate 3,4-dioxygenase. *Biochemistry* 29, 10847-10854
- Van Berkel, W. J. H., Eppink, M. H. M. and Schreuder, H. A. (1994) Crystal structure of *p*-hydroxybenzoate hydroxylase reconstituted with the modified FAD present in alcohol oxidase from methylotrophic yeasts: Evidence for an arabinoflavin. *Protein Science* 3, 2245-2253

- Van Berkel, W. J. H. and Müller, F. (1989) The temperature and pH dependence of some properties of *p*-hydroxybenzoate hydroxylase from *Pseudomonas fluorescens*. *Eur. J. Biochem.* 179, 307-314
- Van Berkel, W. J. H. and Müller, F. (1991) Flavin-dependent monooxygenases with special reference to *p*-hydroxybenzoate hydroxylase. In: F. Müller (ed.) *Chemistry and biochemistry of flavoenzymes 2*. CRC Press, Boca Raton, Florida, pp. 1-29
- Van der Bolt, F. J. T., Van den Heuvel, R. H. H., Vervoort, J. and Van Berkel, W. J. H. (1997) ¹⁹F NMR study on the regiospecificity of hydroxylation of tetrafluoro-4-hydroxybenzoate by wild-type and Y385F *p*-hydroxybenzoate hydroxylase: evidence for a consecutive oxygenolytic dehalogenation mechanism. *Biochemistry* 36, 14192-14201
- Van der Bolt, F. J. T., Vervoort, J. and Van Berkel, W. J. H. (1996) Flavin motion in *p*-hydroxybenzoate hydroxylase: substrate and effector specificity of the Tyr222→Ala mutant. *Eur. J. Biochem.* 237, 592-600
- Van Haandel, M. J. H., Rietjens, I. M. C. M., Soffers, A. E. M. F., Veeger, C., Vervoort, J., Modi, S., Mondal, M. S., Patel, P. K. and Behere, D. V. (1996) Computer calculation-based quantitative structure-activity relationships for the oxidation of phenol derivatives by horseradish peroxidase compound II. *J. Bioinorg. Chem.* 1, 460 - 467
- Vervoort, J., De Jager, P. A., Steenbergen, J., Rietjens, I. M. C. M. (1992) Development of a ¹⁹F-n.m.r. method for studies on the in vivo and in vitro metabolism of 2-fluoroaniline. *Xenobiotica* 20,
- Vervoort, J., Müller, F., Lee, J., Moonen, C. T. W. and Van den Berg, W. A. M. (1986) Identifications of the true carbon-13 NMR spectrum of the stable intermediate II in bacterial luciferase. *Biochemistry* 25, 8062-8067
- Vervoort, J., Ridder, L., Van Berkel, W. J. H. and Rietjens, I. M. C. M. (1996) Flavoprotein monooxygenases: Mechanistic overview. In: Stevenson, K. J., Massey, V. and Williams, C. H. Jr. (eds.) *Flavins and flavoproteins*. University of Calgary Press, Calgary, pp.
- Vervoort, J. and Rietjens, I. M. C. M. (1996) Unifying concepts in flavin-dependent catalysis. *Biochem. Soc. Trans.* 24, 127-130
- Vervoort, J., Rietjens, I. M. C. M., Van Berkel, W. J. H. and Veeger, C. (1992) Frontier orbital study on the 4-hydroxybenzoate-3-hydroxylase-dependent activity with benzoate derivatives. *Eur. J. Biochem.* 206, 479-484
- VMD v1.2b (1996) Theoretical Biophysics Group, University of Illinois and Beckman Institute
- Walsh, T. A., Ballou, D. P., Mayer, R. and Que, L. (1983) Rapid reaction studies on the oxygenation reactions of catechol dioxygenase. *J. Biol. Chem.* 258, 14422-14427
- Walsh, T. A., Ballou, D.P. (1983) Halogenated protocatechuates as substrates for protocatechuate dioxygenase from *Pseudomonas cepacia*. *J. Biol. Chem.* 258, 14413-14421
- Warshel, A. (1978) Energetics of enzyme catalysis. *Proc. Natl. Acad. Sci. USA* 75, 5250-5254
- Warshel, A. (1992) Computer simulations of enzymatic reactions. *Curr. Opin. Struct. Biol.* 2, 230-236
- Warshel, A. and Levitt, M. (1976) Theoretical studies of enzymic reactions: dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* 103, 227-249
- Whittaker, J. W. and Lipscomb, J. D. (1984a) ¹⁷O-water and cyanide ligation by the active site iron of protocatechuate 3,4-dioxygenase. Evidence for displaceable ligands in the native enzyme and in complexes with inhibitors or transition state analogs. *J. Biol. Chem.* 259, 4487-4495

- Whittaker, J. W. and Lipscomb, J. D. (1984b) Transition state analogs for protocatechuate 3,4-dioxygenase. Spectroscopic and kinetic studies of the binding reactions of ketonized substrate analogs. *J. Biol. Chem.* 259, 4476-4486
- Wierenga, R. K., De Jong, R. J., Kalk, K. H., Hol, W. G. J. and Drenth, J. (1979) Crystal structure of *p*-hydroxybenzoate hydroxylase. *J. Mol. Biol.* 131, 55-73
- Zheng, Y. and Ornstein, R. L. (1996) A theoretical study of the structures of flavin in different oxidation and protonation states. *J. Am. Chem. Soc.* 118, 9402-9408
- Zheng, Y. J. and Ornstein, R. L. (1997) Mechanism of nucleophilic aromatic substitution of 1-chloro-2,4-dinitrobenzene by glutathione in the gas phase and in solution. Implications for the mode of action of glutathione S-transferase. *J. Am. Chem. Soc.* 119, 648-655
- Zuilhof, H., Lodder, G. and Koch, H. F. (1997) Carbon-oxygen hydrogen-bonding in dehydrohalogenation reactions: PM3 calculations on polyhalogenated phenylethane derivatives. *J. Org. Chem.* 62, 7457-7463

Samenvatting voor niet-vakgenoten

Beste familieleden, vrienden en andere niet-vakgenoten,

In dit hoofdstuk wil ik proberen een aantal begrippen die in dit proefschrift veelvuldig voorkomen te verduidelijken en het onderzoek in wat minder specialistische termen samen te vatten.

Het onderzoek in dit proefschrift valt onder het vakgebied van de *bio-chemie*. Het voorvoegsel *bio* geeft aan dat dit vakgebied de levende materie betreft. Alles wat leeft is opgebouwd uit *cellen*. Levende cellen zijn klein, je kunt ze alleen met een microscoop zien. Ze bezitten in principe de fundamentele eigenschappen van het leven: ze kunnen voedingsstoffen opnemen en gebruiken om te groeien, te bewegen, zich te vermeerderen of andere taken te verrichten. De eenvoudigste organismen, bijvoorbeeld bacteriën bestaan uit één cel. Hogere organismen, planten, dieren en mensen, kunnen uit miljarden cellen bestaan. (In tegenstelling tot de bacteriecellen zijn de cellen van hogere organismen veelal gespecialiseerd in specifieke taken.) Vanwege hun (relatieve) eenvoud zijn de eencelligen geliefde studieobjecten voor biochemici.

Het tweede deel van het woord biochemie geeft aan dat dit vakgebied op een heel gedetailleerd niveau de levende materie onderzoekt, namelijk op het niveau van *moleculen* en *atomen*. Alle materie bestaat uit atomen. Deze atomen kunnen op verschillende manieren verbindingen aangaan en daardoor een zeer groot aantal verschillende moleculen vormen, ieder met andere eigenschappen. Wanneer bindingen tussen atomen worden gevormd of verbroken ontstaan nieuwe moleculen met gewijzigde eigenschappen. Men spreekt dan van *chemische reacties*.

Cellen zijn opgebouwd uit zeer veel verschillende moleculen. De grotere moleculen kunnen grofweg ingedeeld worden in *eiwitten*, *koolhydraten*, *vetten* en *nucleïnezuuren* (bijv. DNA). Deze worden opgebouwd en in elkaar omgezet via een ingewikkeld netwerk van chemische reacties, die veelal verlopen via kleinere moleculen. Deze chemische reacties, die nodig zijn voor het functioneren van een cel, verlopen niet

spontaan. Ze worden gekatalyseerd (d.w.z. versneld en gecontroleerd) door de zeer complexe eiwitmoleculen in de cel, die ook wel enzymen worden genoemd. Biochemisch onderzoek heeft tot doel te begrijpen hoe de vele chemische reacties, die nodig zijn voor het functioneren van cellen, verlopen en hoe deze reacties worden gekatalyseerd door de talloze enzymen.

Biochemisch onderzoek is in de meeste gevallen een kwestie van experimenten uitvoeren. Het begint meestal met het isoleren van één specifiek eiwit uit het mengsel van vele typen (eiwit-)moleculen in een *celextract* (= stukgemaakte cellen) waarna allerlei eigenschappen van dat specifieke eiwit getest kunnen worden. Met moderne meettechnieken kunnen veel verschillende eigenschappen van eiwitten bepaald worden, en daarmee kunnen biochemici inzicht krijgen in de functies en de werkingsmechanismen van eiwitten. Toch zijn er grenzen aan wat we met deze experimentele technieken te weten kunnen komen, niet alleen vanwege de complexiteit van eiwitmoleculen, maar ook omdat de chemische reacties die ze katalyseren erg snel verlopen, d.w.z. in de orde van microseconden (= miljoenste seconden). Daarom wordt het gebruik van computermodellen bij het verkrijgen van inzicht in de werking van enzymmoleculen steeds belangrijker. De snelheid van computers neemt voortdurend toe en ook de fysisch-chemische theorieën voor het berekenen van eigenschappen en reacties van moleculen zijn de laatste tientallen jaren sterk in ontwikkeling. (De Nobelprijs voor de chemie van 1998 is toegekend aan twee wetenschappers die hieraan een belangrijke bijdrage hebben geleverd.) Omdat eiwitmoleculen erg groot zijn, is het gebruik van computermodellen voor *enzymreacties* extra compliceerd en pas de laatste vijf jaar echt van de grond gekomen. (Bij het onderzoeken van *enzymstructuren* worden al veel langer computerberekeningen gebruikt.) Het onderzoek in dit proefschrift is een bijdrage aan deze recente ontwikkeling om computermodellen te gebruiken voor het rekenen aan enzymreacties.

Ik heb me gericht op drie verschillende enzymen die voorkomen in micro-organismen die zich in de bodem bevinden. Deze micro-organismen zijn in staat benzeenachtige moleculen te gebruiken als voedingsstof en af te breken tot kleinere moleculen zoals water en koolzuurgas. Dit afbraakproces heeft een belangrijke functie in de natuur. Het is een noodzakelijk onderdeel van de rottingsprocessen die ervoor zorgen dat hout wordt afgebroken, en zonder dat zouden belangrijke kringlooppromessen niet functioneren. Daarnaast zijn de betrokken enzymen ook in staat om door de mens geproduceerde milieuverontreinigende stoffen af te breken. In **hoofdstuk 1** wordt een kort overzicht gegeven van wat in de literatuur bekend is over het werkingsmechanisme van de drie enzymen die in dit proefschrift als modelsystemen zijn gebruikt.

In **hoofdstuk 2 en 3** worden de theoretische principes uiteengezet waar de computerberekeningen in dit onderzoek op gebaseerd zijn. Twee verschillende methoden om moleculen te modelleren zijn in het onderzoek gebruikt: *moleculaire mechanica* en *quantummechanica*. De moleculaire mechanica gebruikt een sterk gesimplificeerd model van moleculen: atomen worden hierbij beschreven als bolletjes met een bepaalde elektrische lading en de bindingen tussen de atomen (waardoor moleculen gevormd worden) worden opgevat als veren. Dit eenvoudige model van moleculen is goed bruikbaar voor het berekenen van de optimale geometrie (=vorm) van een molecuul. Voor het simuleren van chemische reacties is dit model echter ontoereikend. Het verbreken van een binding tussen atomen, bijvoorbeeld, verloopt in werkelijkheid anders dan het breken van een veer. De kracht voor het uittrekken van een veer neemt alsmaar toe (totdat hij breekt). In werkelijkheid neemt de kracht tussen atomen bij het verbreken van een binding op een gegeven moment weer af (denk bijvoorbeeld aan het uit elkaar trekken van twee magneten). Voor een meer realistische beschrijving van chemische reacties moet gebruik gemaakt worden van de theorie van de quantummechanica. In de quantummechanische beschrijving van een molecuul wordt rekening gehouden met het feit dat een atoom bestaat uit een atoomkern waar elektronen omheen bewegen en dat deze elektronen zo klein zijn dat ze niet alleen de eigenschappen van deeltjes vertonen (d.w.z. een bepaalde massa, lading, plaats en snelheid), maar ook die van golven (denk bijvoorbeeld aan licht-golven). Juist deze "golf-eigenschappen" zijn essentieel voor de vorming van bindingen tussen de atomen. Voor het berekenen van veel eigenschappen van moleculen is een quantummechanische beschrijving dan ook noodzakelijk.

In **hoofdstuk 4** worden quantummechanische methodes gebruikt om de *reactiviteit* van een serie catechol-moleculen, die omgezet worden door het enzym *catecholdioxygenase*, te berekenen. De hypothese was dat de langzaamste, en dus snelheidsbepalende, stap in het werkingsmechanisme van dit enzym, bestaat uit het vormen van een binding tussen het catecholmolecuul en een zuurstofmolecuul. De snelheid van deze stap zou worden bepaald door de energie van de elektronen in het catecholmolecuul die betrokken zijn bij het vormen van deze binding (dit zijn tevens de elektronen met de hoogste energie). Deze energie kon quantummechanisch berekend worden voor de verschillende catecholmoleculen en bleek inderdaad gerelateerd te zijn aan de snelheid waarmee de reactie verloopt. Deze gevonden correlatie tussen de gemeten snelheid en de berekende energie-parameter bevestigt de hypothese over het werkingsmechanisme van het enzym.

De berekeningen in hoofdstuk 4 betreffen alleen het reagerende (catechol-)molecuul. De berekeningen in hoofdstukken 5 tot en met 8 gaan een stap verder. Bij een enzymgekatalyseerde reactie wordt het reagerende molecuul altijd omgeven door het

veel grotere enzym-molecuul. De specifieke omgeving van het enzym zorgt ervoor dat de reactie sneller verloopt. Het is dus vanuit biochemisch oogpunt interessant om ook het enzymmolecuul in de berekening op te nemen. Echter, de wiskunde die nodig is om een molecuulstructuur quantummechanisch te berekenen is zeer complex en de benodigde computertijd neemt bovendien meer dan kwadratisch toe met het aantal atomen in het molecuul. Enzymmoleculen zijn te groot om volledig quantummechanisch berekend te kunnen worden, zelfs met de modernste computers. De modellen die in **hoofdstuk 5 t/m 8** van dit proefschrift beschreven staan maken gebruik van recent ontwikkelde computer-technieken die de quantummechanische en moleculair mechanische methodes combineren. De quantummechanica wordt alleen toegepast op het gedeelte waar de reactie plaats vindt (het zogenaamde katalytisch centrum) terwijl de rest van het enzymmolecuul wordt beschreven met behulp van de moleculaire mechanica. Op deze manier is het toch mogelijk om berekeningen voor het gehele complex van het enzym en het reagerende molecuul uit te voeren.

Deze berekeningen worden toegepast op twee enzymen die eenzelfde soort reactie katalyseren: *para-hydroxybenzoate hydroxylase* (hoofdstukken 5, 6 en 7) en *fenol-hydroxylase* (hoofdstuk 8). Ook hierbij wordt gekeken naar de langzaamste (en snelheidsbepalende) reactie in het gehele reactiemechanisme van beide enzymen. Het verloop van deze reactie wordt stap voor stap berekend waardoor een zeer gedetailleerd inzicht in deze reactie verkregen wordt. Ook geeft dit type berekening informatie over hoe gemakkelijk de reactie verloopt (in termen van energie). Deze informatie blijkt goed overeen te komen met de reactiesnelheden die experimenteel waargenomen zijn.

Een belangrijk voordeel van de bovengenoemde QM/MM berekeningen is dat het effect van de eiwit-omgeving op het verloop van de reactie kan worden bepaald. Dit laatste aspect van het type berekeningen dat in dit proefschrift is beschreven wordt in de nabije toekomst steeds belangrijker. Er komt namelijk met toenemende snelheid informatie beschikbaar over de precieze samenstelling van enzymen. Het *human-genome project* heeft er toe geleid dat het grootste deel van de erfelijke informatie van de mens bekend is geworden. Deze erfelijke informatie bepaalt hoe de enzymen, die alle essentiële processen in het menselijk lichaam sturen, zijn opgebouwd. De erfelijke informatie vertelt echter nog niet wat de functie van deze enzymen precies is. Ook is het niet direct uit de erfelijke informatie af te leiden wat het precieze effect is van kleine erfelijke verschillen tussen mensen, of van erfelijke afwijkingen, op het functioneren van enzymen. Het type berekeningen dat in dit proefschrift is uitgevoerd zou een belangrijke schakel kunnen zijn bij het maken van de benodigde vertaalslag van erfelijke informatie naar het functioneren van enzymen in alle mogelijke organismen, waaronder de mens.

Dankwoord

Dit proefschrift is het resultaat van ruim vier jaar onderzoek, waarbij een groot aantal mensen betrokken is geweest.

Het onderzoek is begonnen met een projectvoorstel geschreven door begeleider en promotor, Ivonne Rietjens. Hierin beschrijft zij hoe de toepassing van quantumchemische berekeningen in het biochemisch onderzoek verdergaande inzichten kan geven in de mechanismen van enzymkatalyse. Deze ideeën vormden het uitgangspunt en een belangrijke inspiratiebron voor het promotieonderzoek waarvan dit proefschrift verslag doet.

Ivonne wil ik, behalve voor het initiëren van het onderzoek, ook bedanken voor haar supervisie. Ze heeft daarbij enerzijds vrijheid gegeven voor het uitwerken van nieuwe ideeën en anderzijds de rode draad op cruciale momenten bewaakt. Ook heb ik grote waardering voor de intensieve wijze waarop Ivonne vele versies van de manuscripten onder de loep heeft genomen.

Copromotor Jacques Vervoort heeft op ieder moment klaar gestaan voor het bediscussiëren en genereren van nieuwe ideeën, het geven van advies en het delen van zijn kennis op het gebied van flavinekatalyse. Daarnaast was Jacques de drijvende kracht achter het uitbreiden en beschikbaar maken van de uitstekende computerfaciliteiten en programmatuur, die essentieel waren bij dit onderzoek. Jacques was ook betrokken bij het initiëren van samenwerkingsverbanden, o.a. met Adrian Mulholland, Han Zuilhof en Bruce Palfey.

I want to thank Adrian Mulholland for our cooperation, which, despite the fact that it depended very much on the e-mail facility, has been very fruitful. Adrian has been a tremendous help in sharing his expertise on combined quantum mechanical/molecular mechanical calculations. Furthermore, his active role as second author has been indispensable in the writing of three chapters of this thesis.

Han Zuilhof heeft een belangrijke rol gespeeld in het tot stand komen van hoofdstuk 3. Ook wil ik hem bedanken voor de leerzame discussies die we daarnaast gevoerd hebben.

The times we've met, Bruce Palfey came up with critical and therefore useful comments on the PHBH work. I would also like to thank Bruce for his cooperation on chapter 7.

Voor hun hulp bij hoofdstuk 4, gaat mijn dank uit naar de mede-auteurs Eric Vis, Marelle Boersma, Sjef Boeren, Fabrizio Briganti, Andrea Scozzafava en Cees Veeger. Eric heeft in het kader van zijn doctoraal onderzoek belangrijk voorbereidend werk voor dit hoofdstuk verricht o.a. in samenwerking met Fabrizio Briganti en Andrea Scozzafava. Marelle en Sjef waren betrokken bij de kinetische experimenten en eiwitzuiveringen, en Cees Veeger heeft een grote bijdrage geleverd aan de kinetische analyse.

Ook vele mensen, die niet direct als mede-auteur bij het onderzoek zijn betrokken, hebben een belangrijke bijdrage aan mijn onderzoeksproject geleverd. Een aantal van hen wil ik hier graag noemen. Willem van Berkel was altijd te vinden voor een goed gesprek over flavine monooxygenases, waar ik veel van geleerd heb. Ik wil hem graag bedanken voor zijn interesse en betrokkenheid bij mijn werk. Frank Vergeldt stond, als beheerder van het UNIX computersysteem, altijd klaar als dat nodig was. Michel Eppink heeft de nodige hulp geboden, zowel bij experimenteel- als computer-werk. Ans Soffers en Puck Knipscheer zijn betrokken geweest bij een deel van het onderzoek dat niet is opgenomen in dit proefschrift, maar dat ik nog zal voortzetten in een post-doc project.

Kamergenoten zijn vaak de eersten waarmee je successen viert of frustraties deelt. Jean-Louis, Marjon, Hanem, Kees-Jan en verscheidene gast-medewerkers, die tijdelijk onze kamer gedeeld hebben, ben ik dan ook erkentelijk voor de uitstekende sfeer waarin we al die tijd gewerkt hebben.

Tenslotte is promoveren soms ook wel eens een kwestie van niet de moed opgeven en doorzetten. Daarbij is het thuisfront van cruciale betekenis. Bij het gereed komen van dit proefschrift wil ik daarom ook mijn ouders en, bovenal Carla, bedanken, voor hun onvoorwaardelijke steun, geduld en betrokkenheid.

List of publications

- Ridder, L. Vervoort, J., Van Berkel, W. J. H., Veeger, C. and Rietjens, I. M. C. M. (1996) Molecular orbital analysis of the reaction pathway for hydroxylation of *p*-hydroxybenzoate by the flavin (C_{4a})-hydroperoxide intermediate of *p*-hydroxybenzoate hydroxylase. In: Stevenson, K. J., Massey, V., Williams, C. H. Jr. *Flavins and flavoproteins 1996*. University Press, Calgary, pp. 327-330
- Vervoort, J. Ridder, L. Van Berkel, W. J. H., and Rietjens, I. M. C. M. (1996) Flavoprotein monooxygenases: Mechanistic overview. In: Stevenson, K. J., Massey, V., Williams, C. H. Jr. *Flavins and flavoproteins 1996*. University Press, Calgary, pp. 327-330
- Ridder, L., Briganti, F., Boersma, M. G., Boeren, J., Vis, E. H., Scozzafava, A., Veeger, C. and Rietjens, I. M. C. M. (1998) Quantitative structure activity relationship for the rate of conversion of C4-substituted catechols by catechol-1,2-dioxygenase from *Pseudomonas putida* (arvilla) C1. *Eur. J. Biochem.* 257, 92-100
- Ridder, L., Mulholland, A. J., Vervoort, J. and Rietjens, I. M. C. M. (1998) Correlation of calculated activation energies with experimental rate constants for an enzyme catalyzed aromatic hydroxylation. *J. Am. Chem. Soc.* 120, 7641-7642
- Ridder, L., Mulholland, A. J., Rietjens, I. M. C. M. and Vervoort, J. (1999) Combined quantum mechanical and molecular mechanical reaction pathway calculation for aromatic hydroxylation by *p*-hydroxybenzoate-3-hydroxylase. *J. Mol. Graphics Mod.* 17, 163-175
- Ridder, L., Zuilhof, H., Vervoort, J. and Rietjens, I. M. C. M. (1999) Computational methods in flavin research. In: S. K. Chapman and G. A. Reid (eds.) *Meth. Mol. Biol. Flavoprotein Protocols*. Humana Press Inc., NJ, pp. 207-228
- Ridder, L., Vervoort, J., Rietjens, I. M. C. M. (1999) The application of QM/MM techniques to flavoproteins; *p*-hydroxybenzoate hydroxylase (PHBH). In: Ghisla, S., Kroneck, P. M. H., Macheroux, P. and Sund, H. (eds.) *Flavins and flavoproteins 1999*. Agency for Scientific Publ., Berlin, pp. 719-727
- Ridder, L., Palfey, B. A., Vervoort, J. and Rietjens, I. M. C. M. (2000) Modelling flavin- and substrate substituent effects on the activation barrier and rate of oxygen transfer by *p*-hydroxybenzoate hydroxylase. *FEBS Letters* In press.
- Ridder, L., Adrian J. Mulholland, Rietjens, I. M. C. M. and Vervoort, J. (2000) A quantum mechanical/molecular mechanical study of the hydroxylation of phenol and halogenated derivatives by phenol hydroxylase. *J. Am. Chem. Soc.* In press.

Curriculum Vitae

Lars-Olaf Ridder werd op 19 december 1970 geboren te Utrecht. In 1989 behaalde hij het VWO diploma aan de Rijksscholengemeenschap Broeklede te Breukelen. In september 1990 begon hij met de studie Moleculaire Wetenschappen aan de Landbouwniversiteit te Wageningen. Tijdens de doctoraalfase deed hij afstudeervakken bij de vakgroepen Microbiologie (dr. S.W.M. Kengen, dr. A.J.M. Stams) en Biochemie (dr. I.M.C.M. Rietjens, prof. dr. C. Veeger). Tevens voerde hij een stage uit bij Organon, sectie Computational Medicinal Chemistry (dr. P.J.M. van Galen, dr. P.D.J. Grootenhuis) te Oss. In januari 1996 behaalde hij het doctoraal diploma met lof.

Van februari 1996 tot en met mei 2000 is hij als onderzoeker in opleiding verbonden geweest aan het Laboratorium voor Biochemie van de Wageningen Universiteit, waar het in dit proefschrift beschreven onderzoek werd uitgevoerd onder supervisie van prof. dr. I.M.C.M. Rietjens en dr. J. Vervoort. Dit onderzoek werd gefinancierd door de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO).

Vanaf juni 2000 is hij als post-doc onderzoeker verbonden aan de University of Bristol, in het kader van een Marie-Curie fellowship.