First principles calculation of the activity of cytochrome P450

M. D. Segall and M. C. Payne
Cavendish Laboratory (TCM), University of Cambridge, Madingley Road, Cambridge CB3 0HE, United Kingdom

S. W. Ellis and G. T. Tucker
Department of Medicine and Pharmacology, University of Sheffield, Royal Hallamshire Hospital, Sheffield S10 2JF, United Kingdom

R. N. Boyes
ML Laboratories plc, 60 London Road, St. Albans AL1 1NG, United Kingdom
(Received 29 August 1997)

The cytochrome P450 superfamily of enzymes is of enormous interest in the biological sciences due to the wide range of endogenous and xenobiotic compounds which it metabolises, including many drugs. We describe the use of first principles quantum mechanical modeling techniques, based on density functional theory, to determine the outcome of interactions between an enzyme and a number of compounds. Specifically, we calculate the spin state of an Fe^{3+} ion present in a haem moiety at the active site of these enzymes. The spin state of this ion indicates if the catalytic reaction will proceed. The computational results obtained compare favorably with experimental data. Only the principle components of the active site of the enzyme are included in the computational models, demonstrating that only a small fragment of the protein needs to be included in the models in order to accurately reproduce this aspect of the enzymes' function. These results open the way for further investigation of this superfamily of enzymes using the methods detailed in this paper. [S1063-651X(98)09404-5]

PACS number(s): 87.22.Me, 87.15.−v

I. INTRODUCTION

The cytochromes P450 (P450's) [1] are a superfamily of enzymes which are found in all forms of living organisms. They are responsible for the metabolism of many endogenous compounds, participate in the activation or deactivation of many carcinogens, and detoxify many xenobiotics. In particular, in humans they metabolize many drugs and hence are of great interest to pharmacologists and toxicologists.

This paper describes the application of first principles, or *ab initio*, methods to the study of P450's. There are two principle goals for this investigation. *Ab initio* simulations may aid in the understanding of the detailed mechanisms of the interaction of ligand molecules with the active site of an enzyme. However, the ultimate goal is the development of techniques for the prediction of the outcome of such an interaction.

Ab initio methods solve the quantum mechanical equations which govern the behavior of a system. The only information which must be provided are the atomic numbers and positions of the atoms within the system. In contrast, conventional empirical or semiempirical approaches, widely used in biological modeling, require a model of the interactions between the atoms to be supplied. The parameters of these models are usually derived by fitting the outcome of simulations to experimental data. The problems with these techniques arise when you consider the question of the range of their applicability. If the parameters of the models were derived from system A, what guarantee is there that they apply to system B? First principles approaches are computationally more expensive than conventional techniques and only recently has the *ab initio* study of systems of biological interest become tractable.

An important question when investigating large biomolecular systems is, how much of the system must be included in a model to accurately reproduce the phenomena of interest? Because of the computational cost of *ab initio* methods, it is impossible to include an entire protein in the calculations. Therefore, if such techniques are to be useful in understanding biological systems, it is important to demonstrate that only a small fragment of the molecule need be considered. Previous studies (for example, that by Ghosh *et al.* [2]) have simulated the behavior of model systems which have similar structures to the P450 active site. In contrast, we include the full structure of the active site haem and cysteine in our calculations.

Section II describes features of the cytochromes P450 which relate to the investigation presented in this paper. The computational methods applied are detailed in Sec. III and the results obtained are given in Sec. IV. Finally, conclusions are drawn in Sec. V.

II. CYTOCHROMES P450

The cytochromes P450 are hemoproteins [3,4], made up of between 400 and ~500 amino acids and containing a single haem prosthetic group. The superfamily is named after a strong maximum in the UV spectrum at 450 nm which these enzymes exhibit when complexed with CO.

The principle physiological role of the P450 superfamily of enzymes is that of a monoxygenase. The catalytic reaction can be summarized:

$$RH + O_2 + 2H^+ + 2e^- \rightarrow ROH + H_2O \qquad (1)$$

where RH can be one of a large number of possible substrates.

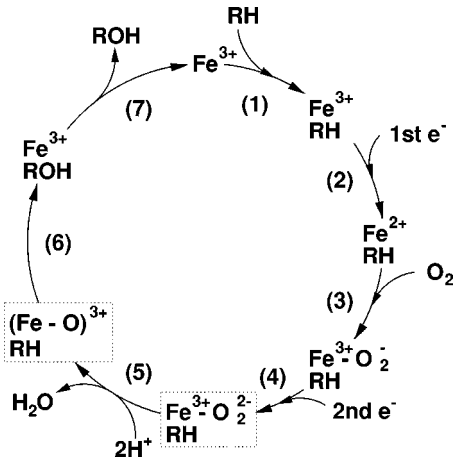

FIG. 1. The cytochrome P450 reaction cycle. R and ROH represent substrate and hydroxylated substrate, respectively. The states shown in a dotted box have not been directly observed.

The reaction takes place on the Fe^{3+} ion in the haem moiety found at the active site of all P450 enzymes. The overall catalytic cycle is outlined in Fig. 1. We focus on the state of the system between processes (1) and (2) after the substrate has a bound, but before the transfer of the first electron. Spectral evidence [5] indicates that in the substrate-free form of P450, the haem Fe^{3+} is in a low-spin (LS) state. However, on binding of a substrate the Fe^{3+} ion is found to change to a high-spin (HS) state [6]. This is accompanied by a change in redox potential of the haem Fe^{3+} by about 100 mV [7], which makes the reduction of the cytochrome P450 energetically favorable and permits the catalytic reaction to proceed.

The HS and LS states are not independent, but exist in equilibrium. Differences in absorption spectra allow the equilibrium constant between the spin states and hence the fraction of HS character to be determined. The equilibrium between HS and LS is only one of a number of microequilibria which must be considered in a full description of the substrate binding reaction. A detailed analysis of these may be found in Sligar and Murray *et al.* [8]. Furthermore, a linear free-energy relationship has been found between the spin state and the redox potential for transfer of the first electron [9].

On binding of a competitive inhibitor, changes in the absorption spectrum are observed which are associated with an increase in the LS character of Fe^{3+} [6]. The binding of substrate analogs, which are similar to true substrates but differ in their binding to the active site, cause changes in the spin character between those of the ligand-free and substrate-bound enzyme. However, the spin equilibria of substrate-analog-bound enzymes favor the LS state.

III. COMPUTATIONAL METHODS

The *ab initio* method we used was based upon the Kohn-Sham [10] formulation of density functional theory (DFT) [11]. The use of DFT avoids the high cost of treating the electron-electron interactions in conventional quantum chemistry techniques by formulating the total electronic energy as a functional of the electron density. A spin-dependent generalized gradient approximation (GGA) was used for the exchange-correlation potential [12]. A GGA parametrizes the exchange-correlation potential in terms of the magnitude and gradient of the electronic density. This contrasts with the more conventional local-density approximation which depends on the magnitude of the density alone. Although GGA's do not offer a consistent improvement on the LDA in all types of system, they have been shown to improve on the LDA for calculations of molecular structures and in representing weak intermolecular bonds [13,14]. The electronic wave functions were represented using a plane wave basis set with a cutoff of 600 eV. While confering many benefits, including the efficient calculation of atomic forces, a plane wave basis set requires the use of many more basis functions than a localized basis set. This disadvantage was mitigated by the use of *ab initio* pseudopotentials [15] to represent the potential due to the atomic nuclei and core electrons. In order to accurately model the spin state of Fe^{3+} nonlinear core corrections [16] were applied to the Fe pseudopotential. These explicitly treat the nonlinear interaction between the core and valence electronic densities when calculating the exchange-correlation potential. Another requirement of the use of a plane wave basis set is the imposition of periodic boundary conditions. This was achieved by the use of the supercell approach, whereby the system studied is repeated periodically in space separated by large vacuum regions. For the calculations described here a cubic supercell with a side of length 19 Å was used. A detailed description of these techniques may be found in [17].

The computational cost of this approach scales as the third power of the size of system studied. This is significantly better than the scaling of quantum chemistry methods of similar accuracy. Combined with the use of parallel supercomputers, this permits biological systems containing up to a few hundred atoms to be modeled within a reasonable time.

A typical P450 enzyme, containing approximately 3500 atoms, is prohibitively large to be modeled in its entirety. Therefore, attention must be focused on a smaller region of the enzyme around the active site. In order to limit the computational resources required and to address the properties of P450's in general, none of the contact residues which define the substrate binding pocket were included in the simulations. The model systems consisted of the ligand molecule, haem moiety, Fe^{3+}-ligated cysteine residue, and Fe^{3+}-ligated water (present in ligand-free and substrate-analog-bound active sites), and contained between 90 and 118 atoms. An example of such a system is shown in Fig. 2. On removing the active site, including the cysteine residue, from the remaining structure of the enzyme, dangling bonds were created on the cysteine. Hydrogen atoms were added at the locations of these dangling bonds. This is a good approximation as saturated bonds to the protein backbone are replaced by saturated bonds to hydrogen. As the electronegativities of hydrogen and carbon are similar, any change in charge transfer to the active site is minimized. In addition, as the sites of these altered bonds on the cysteine residue are relatively remote from the Fe^{3+} ion of interest, we would expect the effects of these alterations on the chemistry of the iron to be small.

No questions of substrate specificity were addressed as this is defined by the geometry of the substrate binding

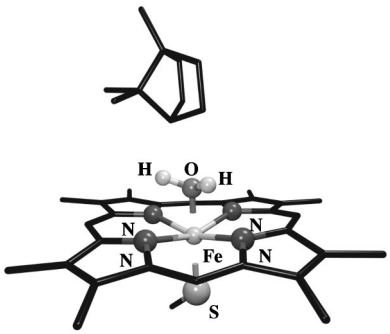

FIG. 2. A fragment of the camphane-bound system modeled. Camphane, a substrate analog, and fragments of the haem moiety and iron-ligated cysteine residue are shown. Note the water molecule that remains coordinated with the iron atom. The full haem moiety and cysteine residue were included in the computational models. Carbon atoms are shown as bonds only; other atom types are labeled. The hydrogen atoms have been omitted except in the case of the iron-ligated water molecule.

pocket. In the future, questions regarding the nature of the O_2-bound system and also specific P450 enzyme species may be addressed. This will require the inclusion of additional contact residues.

Because of the limited size of the system modeled and the uncertainty in the structure of the active sites of most P450's, cytochrome $P450_{cam}$ (CYP101) was chosen as the subject of this investigation because accurate crystal structures of this enzyme have been obtained. These structures include the enzyme complexed with substrates, substrate analogs, and inhibitors. The systems included in this study are listed in Table I. The use of crystal structures allows the determination of the position of the ligand relative to the active site and the changes in the geometry of the active site caused by ligand binding. This restricts one source of uncertainty in the results obtained from the calculations. At this early stage it is important to limit any possible external sources of error so that any deviations from experimental data can be attributed solely to the computational modeling approach taken in the simulation.

The main spin-dependent calculations were performed on 64 processors of a Hitachi SR2201 parallel supercomputer. Preliminary calculations were performed on 64 nodes of a Hitachi SR2001 supercomputer and 16 nodes of an SR2201.

TABLE I. The $P450_{cam}$ structures used in the *ab initio* investigation.

Ligand	Nature of ligand	PDB[a] reference	Reference
Substrate-free	N/A	1PHC	[18]
Camphor	Substrate	4CP4	[19]
Adamantanone	Substrate	5CPP	[20]
Camphane	Substrate analog	6CPP	[19]
Norcamphor	Substrate analog	7CPP	[20]
Metyrapone	Inhibitor	1PHG	[21]

[a]The Brookhaven Protein Database.

These preliminary calculations primarily involved the relaxation of the hydrogen atoms in the systems into their equilibrium positions as their atomic coordinates were not specified in the crystal structure data sets.

IV. RESULTS

The electronic structure calculations performed on each of the systems listed in Table I yield charge- and spin-density distributions for the ground states. Population analysis of the results of the electronic structure calculations, using the method described in [22], gives a quantitative measurement of the spin and charge on iron for each system studied. The results of the population analyses are presented in Table II. It should be noted that these results relate only to the ground state of each system and do not include any thermodynamic or entropic effects. It should also be noted that the charge on an Fe^{3+} ion embedded in a porphyrin ring is expected to be approximately $1e$. This compares favorably with the charge of 1.10 e calculated for the ligand-free system.

Figure 3 shows a graph of the fraction of HS component against the calculated ground-state spin on the Fe^{3+} ion for the ligand-free, substrate-bound, and substrate-analog-bound systems. This shows an excellent correlation between the calculated ground-state iron spin and the spin equilibrium. This indicates that the *ab initio* approach accurately predicts the spin state of the iron in the P450 system. The results also demonstrate that the majority of the difference in the spin state between complexes can be accounted for by the difference in the ground-state spin distributions.

The spin on haem iron in the inhibitor-bound system is

TABLE II. Calculated ground-state spins and charges, and high-spin fractions for P450 systems studied. In some cases multiple values have been reported for the spin equilibrium constants, resulting in a range of possible HS fractions. No spin equilibrium data were available for the metyrapone-bound system.

Ligand	Total spin (units of $\hbar$)	Fe spin (units of $\hbar$)	Fe charge (units of e)	HS fraction (%)
None	0	0.11	1.10	8
Camphor	1	1.14	1.11	94–97[a]
Adamantanone	1	1.06	1.11	96–98[b]
Camphane	0	0.34	1.24	46[c]
Norcamphor	0	0.27	1.24	46[d]
Metyrapone	0	0.28	1.19	N/A

[a]Refs. [23,24,9].
[b]Refs. [9,24].
[c]Ref. [25].
[d]Ref. [9].

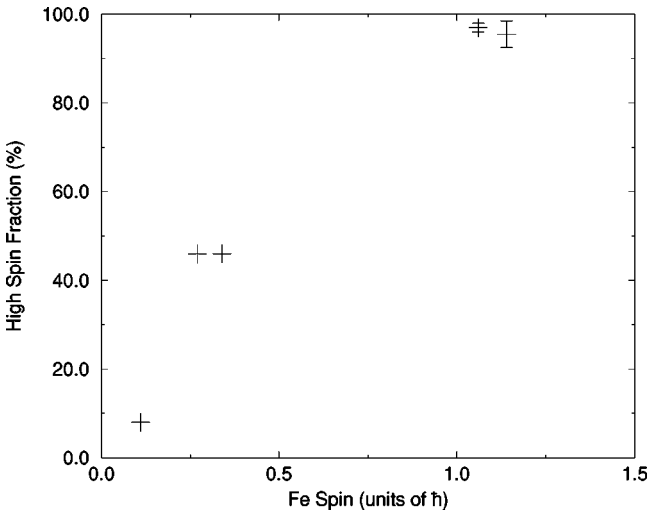

FIG. 3. Graph of fraction of HS character found experimentally against calculated Fe^{3+} spin. Where more than one value for the fraction of HS has been reported, a range of possible values has been plotted.

expected to be less than that for the ligand-free case [26]. While the results show that the ground state is LS in character, the spin on the iron is found to be higher than that in the ligand-free system and of a similar magnitude to that found in the substrate-analog-bound complexes. The difference in the HS fraction seen in experiments may be due to thermodynamic mixing of high-spin states which are accessible to the ligand-free and substrate-analog-bound systems, but not in the inhibitor-bound case. For example, the dissociation of the water molecule or inhibitor from the Fe^{3+} ion would make the HS state more favorable. However, the dissociation of water is probably much more likely than the dissociation of the inhibitor. This would lead to a larger proportion of high-spin states in thermodynamic equilibrium for the ligand-free and substrate-analog-bound cases than in the inhibitor-bound case, even if the ground-state spins were similar.

V. CONCLUSION

The results in this paper indicate that the *ab initio* approach described accurately reproduces the ground-state spin of the Fe^{3+} at the active site of a cytochrome P450 enzyme. In particular the low- and high-spin cases are clearly differentiated, allowing substrate ligands to be clearly identified. This is the first step along the route to develop techniques for predicting the action of a P450 enzyme on a compound.

An important conclusion of this study is that the ligand-P450 interactions can be modeled by considering only a small fragment of the P450 molecule, comprising the haem moiety and haem-ligated cysteine residue. These calculations open the possibility of many other avenues of enquiry including first principles calculations of redox potentials and the calculation of the structures and energies of the oxygen-bound intermediary states which are not directly accessible to conventional experiments (see Fig. 1). These calculations will require the modeling of larger systems, including additional residues in the active site. However, these calculations should be tractable given the computing power that is now available from high-performance parallel computers.

ACKNOWLEDGMENTS

M.D.S. would like to acknowledge the financial support of ML Laboratories plc during the course of this work. The calculations presented in this article were performed on the Hitachi SR2201 located at the University of Cambridge High Performance Computing Facility and the Hitachi SR2001 located at Hitachi Europe's Maidenhead (United Kingdom) headquarters.

[1] D. F. V. Levis, *Cytochromes P450: Structure, Function and Mechanism* (Taylor & Francis, London, 1996).
[2] A. Ghosh, J. Almlöf, and L. Que, Jr., J. Phys. Chem. **98**, 5576 (1994).
[3] T. Omura and R. Sato, J. Biol. Chem. **239**, 2370 (1964).
[4] T. Omura and R. Sato, J. Biol. Chem. **239**, 2378 (1964).
[5] K. Kobayashi, M. Amano, Y. Kanbara, and K. Hayashi, J. Biol. Chem. **262**, 5445 (1987).
[6] J. B. Schenkman, S. G. Sligar, and D. L. Cinti, Pharmacol. Therapeut. **12**, 43 (1981).
[7] K. Ruckpaul, H. Rein, and J. Blank, Frontiers Biotransform. **1**, 1 (1989).
[8] S. G. Sligar and R. I. Murray, in *Cytochrome P-450*, edited by P. R. Ortiz de Montellano (Plenum, New York, 1986), pp. 429–503.
[9] M. Fisher and S. Sligar, J. Am. Chem. Soc. **107**, 5018 (1985).
[10] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).
[11] P. Honenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).
[12] J. P. Perdew, in *Electronic Structure of Solids*, edited by P. Ziesche and H. Eschrig (Akademie-Verlag, Berlin, 1991).
[13] Y.-M. Juan and E. Kaxiras, Phys. Rev. B **48**, 14 944 (1993).
[14] J. P. Perdew *et al.*, Phys. Rev. B **46**, 6671 (1992).
[15] M. T. Yin and M. L. Cohen, Phys. Rev. B **25**, 7403 (1982).
[16] S. G. Louie, S. Froyen, and M. L. Cohen, Phys. Rev. B **26**, 1738 (1982).
[17] M. C. Payne *et al.*, Rev. Mod. Phys. **64**, 1045 (1992).
[18] T. L. Poulos, B. C. Finzel, and A. J. Howard, Biochemistry **25**, 5314 (1986).
[19] R. Raag and T. L. Poulos, Biochemistry **30**, 2674 (1991).
[20] R. Raag and T. L. Poulos, Biochemistry **28**, 917 (1989).
[21] T. L. Poulos and A. J. Howard, Biochemistry **26**, 8165 (1987).
[22] M. D. Segall, C. J. Pickard, R. Shah, and M. C. Payne, Mol. Phys. **89**, 571 (1996).
[23] G. G. Gibson, in *Development of Drugs and Modern Medicines*, edited by J. Gorrod (Ellis Horwood, Chichester, 1986), pp. 253–266.
[24] R. E. White, M. B. McCarthy, K. D. Egeberg, and S. G. Sligar, Arch. Biochem. Biophys. **228**, 493 (1984).
[25] W. M. Atkins and S. G. Sligar, J. Biol. Chem. **263**, 18 842 (1988).
[26] J. H. Dawson, L. A. Andersson, and M. Sono, J. Biol. Chem. **257**, 3606 (1982).