Pergamon

Biochemical Pharmacology, Vol. 50, No. 5, pp. 619–625, 1995.

0006-2952/95 $9.50 + 0.00

0006-2952(95)00188-3

A QUANTITATIVE STRUCTURE–ACTIVITY RELATIONSHIP STUDY ON A SERIES OF 10 *PARA*-SUBSTITUTED TOLUENES BINDING TO CYTOCHROME P4502B4 (CYP2B4), AND THEIR HYDROXYLATION RATES

DAVID F. V. LEWIS,* COSTAS IOANNIDES and DENNIS V. PARKE

Molecular Toxicology Group, School of Biological Sciences University of Surrey, Guildford, Surrey GU2 5XH, U.K.

(*Received* 26 *January* 1995; *accepted* 13 *April* 1995)

Abstract—Molecular structural and molecular orbital calculations (AM1 method) are reported on a series of 10 *para*-substituted toluene derivatives and this structural information has been used to rationalize the differences between both rates of hydroxylation catalysed by cytochrome P4502B4 and binding to the same cytochrome P450, via the generation of quantitative structure–activity relationships (QSARs). It was found that the rate constant for hydroxylation can be described by a two-variable expression involving the dipole moment and volume of the solvent-accessible molecular surface (r = 0.98), whereas binding free energies are well characterized by combinations of molecular volume and various electronic frontier orbital parameters (r = 0.98 and 0.99). This study represents an advance on a previous evaluation by White and McCarthy (*Arch Biochem Biophys* **246**: 19–32, 1986) who used empirical physico-chemical parameters to obtain similar results which were generally of lower statistical significance to those of the present work. The QSAR expressions suggest that both binding to P450 and metabolism for this series of compounds are dependent on the relative ability of the molecules to desolvate and occupy the heme binding site, together with electronic properties of the whole molecule and of the methyl group which undergoes hydroxylation.

Key words: quantitative structure–activity relationships; cytochromes P450; molecular orbital calculations; metabolism and substrate binding

The cytochromes P450 (CYP†) represent a superfamily of mixed-function oxidases/monooxygenases involved in the metabolism of a large number and variety of exogenous and endogenous chemicals [1]. As the P450 system appears to be the major phase I metabolizing enzyme complex for xenobiotics, which often induce their own metabolism, there is considerable interest and activity in this area, particularly regarding possible relationships between the structures of chemical substrates and various measures of biological activity such as catalytic turnover and enzyme binding affinity.

The CYP2B subfamily is one of the important inducible CYP2 subfamilies in mammalia, particularly in rodents and the rabbit, as these are the experimental species where most studies have been undertaken. In general, the CYP2B subfamily is involved in the detoxication of exogenous substrates such as phenobarbital, which is also the recognized chemical inducer of this isoform [2]. An investigation of the structural and electronic requirements for CYP2B substrate specificity, binding affinity and catalytic rate is important for the detailed understanding of the mode of action of the P450 enzymes, including the possible relationship between different thermodynamic and kinetic effects within cytochrome P450 biophysical chemistry.

The current study represents an investigation of possible QSARs within a series of 10 *para*-substituted toluene derivatives exhibiting differences in both binding affinity with, and hydroxylation rate mediated by, cytochrome P4502B4 (CYP2B4) which is the rabbit orthologue of this subfamily [1]. Experimental data, previously tabulated in an earlier publication [3] were also evaluated for QSARs in the same work using empirical substituent parameters such as π, σ, σ, E_R, E_S, (where π is the Hansch hydrophobic substituent parameter, σ is the Hammett electronic substituent parameter, σ is the Okamoto–Brown electronic substituent parameter, E_R is the Otsu homolytic substituent parameter and E_S is the Taft steric substituent parameter) together with experimental values for molar volume, dipole moment, refractive index and dielectric constant, although some of these values were either estimated or had to be corrected to 25° [3]. Although some good correlations were achieved using this data set, it was thought that improved QSARs, possibly with greater physical meaning, could be obtained from the employment of molecular orbital-calculated electronic structural parameters, together with more sophisticated determinations of molecular volumes, and by using experimental log P values.

METHODS

The molecular structures of all 10 toluene

* Corresponding author: David F. V. Lewis.

† Abbreviations: CYP, cytochrome P450; AM1, Austin method 1; QSAR, quantitative structure–activity relationship; r, correlation coefficient; P, octanol/water partition coefficient; MO, molecular orbital; π, hydrophobic substituent parameter; COMPACT, computer-optimized molecular parametric analysis of chemical toxicity.

Table 1. Physicochemical, structural and experimental data for 10 toluene derivatives*

p-substituent	cd†	log *P*‡	μ§	E(HOMO)¶	Q_H‖	Q_L‖	*V***	log k_{cat}††	log K_D‡‡	ΔG_{obs}§§	ΔG_{bind}¶¶
1. I	6.2	3.73 (e)	1.816	−9.4719	0.0161	0.0106	115.7042	1.2601	1.0374	−28.3	−31.5
2. CH_3	6.0	3.15	0.059	−9.1529	0.0188	0.0108	101.4387	1.1139	1.8645	−23.6	−27.0
3. Br	6.0	3.47 (e)	1.744	−9.3825	0.0175	0.0106	104.4625	1.0607	1.3502	−26.5	−29.6
4. H	5.7	2.73	0.279	−9.3827	0.0205	0.0113	85.4317	0.8751	2.3181	−21.0	−25.0
5. Cl	6.0	3.33	1.828	−9.4315	0.0183	0.0104	99.2778	0.7853	1.7760	−24.1	−27.7
6. F	5.8	2.75 (e)	1.804	−9.3229	0.0187	0.0101	89.2799	0.6128	1.9176	−23.3	−25.8
7. CN	6.0	2.04 (e)	3.714	−9.8007	0.0191	0.0093	101.8676	0.3160	1.6096	−25.1	−26.2
8. NO_2	6.2	2.42	5.997	−10.4529	0.0000	0.0064	105.1350	0.01284	1.8733	−23.6	−25.8
9. $CH(CH_3)_2$	6.5	4.10	0.050	−9.1679	0.0168	0.0098	135.1357	—	0.4800	−31.5	−36.1
10. $CH(CH_3)_3$	6.8	4.59 (e)	0.086	−9.1528	0.0113	0.0092	151.9256	—	−0.04096	−34.5	−38.3

* All biochemical data were obtained from ref. [3].

† cd, molecular collision diameter (Å).

‡ log P, experimental values of log *P* taken from the compilation of Sangster, ref. [10] 1989; (e) represents an estimated value based on the benzene derivative and an increment of 0.48 for a CH_2 group.

§ μ, dipole moment calculated by the AM1 procedure (D).

¶ E(HOMO), energy (eV) of the highest occupied molecular orbital (AM1).

‖ $Q_{H,L}$, greatest populations in the HOMO and LUMO, respectively, for methyl group hydrogen atoms.

** *V*, solvent-accessible surface molecular volume ($Å^3$) using a 1.4 Å radius probe sphere.

†† k_{cat}, $P450_{LM2}$ catalysed hydroxylation rate (nmol/min).

‡‡ K_D, dissociation constant between substrate and $P450_{LM2}$ (μM).

§§ ΔG_{obs}, observed free energy of binding to $P450_{LM2}$ (kJ/mol).

¶¶ ΔG_{bind}, intrinsic binding free energy to $P450_{LM2}$ high-spin (kJ/mol) based on the equation:

$$\Delta G_{obs} = \Delta G_{bind} - RT \ln K_1$$

where K_1 is the equilibrium constant for the interconversion of substrate-free $P450_{LM2}$ low-spin and high-spin states.

derivatives were prepared using the Sybyl (Tripos Associates, St Louis, MO, U.S.A.) molecular modelling software system (version 6.03) and, in each case, the molecular geometries were energy minimized via the Tripos force field [4]. Solvent-accessible molecular surfaces were generated for a 1.4 Å radius probe sphere, using the Connolly surface [5] implementation in Sybyl 6.03, for each geometry optimized molecular structure. Molecular orbital calculations via the AM1 method [6, 7] were executed for all 10 energy-minimized toluene derivatives from within the MOPAC module in Sybyl using the normal covergence option. However, the COSMIC molecular modelling framework [8] was employed for the generation of frontier orbital populations under MOPAC [9]. In the latter case, molecular geometries were optimized using the COSMIC force field prior to electronic structural calculations.

The experimental values for log P (where P is the octanol–water partition coefficient) were taken from a published compilation [10] but, in some cases, the values were estimated from the corresponding benzene derivative using an incremental value of 0.48 for a methylene group. Spectral binding data and hydroxylation rates for the 10 toluene derivatives were obtained from the literature [3].

Statistical calculations of correlations between structural descriptors and biochemical data were carried out using stepwise multiple linear regression analysis with the elimination of cross-correlations between independent variables. Regressions containing more than two independent variables were also excluded from the analysis.

Molecular modelling was performed on a Silicon Graphics Personal Iris workstation (Sybyl) and on Sigmex 6130/Micro VAX II combination (COSMIC) whereas the statistical calculations were run on a Hewlett-Packard mainframe, operating under the Unix system.

RESULTS AND DISCUSSION

The data used for the generation of QSARs for the 10 toluene derivatives is shown in Table 1, and this represents the results of biochemical measurements [3] together with those molecular and electronic structural features which gave rise to significant correlations with the experimental values. The tabulated log P data contains estimated values for five of the *para*-substituted toluenes and, therefore, these cannot be considered as entirely reliable. However, the log P values for the 10 compounds gave an extremely good correlation (r = 0.999) with the π substituent parameter data listed by White and McCarthy [3]. Although the dipole moments, presented in Table 1, were calculated using the AM1 MO procedure, they exhibit a satisfactory concordance (r = 0.958) with the experimental compilation of White and McCarthy [3], even though these were estimated for four compounds.

The most statistically significant correlations between one or more of the various structural descriptors and the experimental kinetic or thermodynamic data for 8 and 10 toluene derivatives, respectively, are shown in Table 2. Tables 3–6 summarize the correlations with individual variables for catalytic activity and binding affinity, such that the relative performance of each parameter can be compared.

Rate of hydroxylation

For the hydroxylation rates (log k_{cat}) biochemical

Table 2. QSARs for 10 toluene derivatives*

Regression	equation	n†	s‡	R§	F¶
1. log k_{cat} =	495.99 Q_L − 4.35 (± 125.57)	8	0.2412	0.85	15.6
2. log k_{cat} =	−0.227 μ + 0.024 V − 1.143 (± 0.022) (± 0.005)	8	0.1067	0.98	52.7
3. log k_D =	−0.035 V + 5.206 (± 0.003)	10	0.1789	0.97	141.7
4. log k_D =	−0.036 V − 18.201 Q_H + 5.677 (± 0.003) (± 8.689)	10	0.1499	0.98	103.0
5. ΔG_{obs} =	−0.199 V − 4.513 (± 0.016)	10	1.0046	0.97	146.6
6. ΔG_{obs} =	−0.208 V − 101.12 Q_H − 1.895 (± 0.015) (± 49.10)	10	0.8472	0.98	105.2
7. ΔG_{bind} =	−0.218 V − 5.597 (± 0.022)	10	1.3708	0.96	94.5
8. ΔG_{bind} =	−0.204 V − 2.658 E(HOMO) − 32.28 (± 0.015) (± 0.798)	10	0.9117	0.98	112.4
9. ΔG_{bind} =	−0.231 V − 792.95 Q_L + 3.640 (± 0.014) (± 208.69)	10	0.8374	0.99	133.8

* Correlation between V (volume) and log P (expt.) = 0.81 for 10 compounds.
† n, numbers of observations.
‡ s, standard error.
§ R, correlation coefficient.
¶ F, variance ratio (Fisher F test).

Table 3. Correlations with log k_{cat}

Parameters	R	F
log P	0.85	15.5
μ	0.84	14.5
E(HOMO)	0.83	13.4
Q_H	0.62	3.8
Q_L	0.85	15.6
V	0.18	0.2
μ, V	0.98	52.7

Table 4. Correlations with log K_D

Parameters	R	F
log P	0.84	18.8
μ	0.34	1.1
E(HOMO)	0.37	1.3
Q_H	0.17	0.2
Q_L	0.11	0.1
V	0.97	141.7
Q_H, V	0.98	103.0

Table 5. Correlations with log ΔG_{obs}

Parameters	R	F
log P	0.84	18.4
μ	0.34	1.0
E(HOMO)	0.36	1.2
Q_H	0.17	0.2
Q_L	0.12	0.1
V	0.97	146.6
Q_H, V	0.98	105.2

Table 6. Correlations with log ΔG_{bind}

Parameters	R	F
log P	0.91	40.6
μ	0.47	2.3
E(HOMO)	0.47	2.3
Q_H	0.11	0.1
Q_L	0.01	0.0
V	0.96	94.5
E(HOMO), V	0.98	112.4
Q_H, V	0.99	133.8

data were only available for 8 out of the 10 compounds, and Table 3 shows that the hydrophobic parameter, log P, correlates quite well with the experimental values ($r = 0.85$). Only the LUMO population on the methyl hydrogens is marginally more significant in the single parameter regressions, and this equation (eqn 1) is presented in Table 2. As the hydroxylation of toluenes primarily occurs on the methyl group [3], it was expected that electronic features on this moiety would correlate with the catalytic rate. However, it is interesting that the LUMO, rather than the HOMO, is involved. The correlation with Q_H is substantially less significant than that with Q_L (Table 3), but the correlation with E(HOMO), which is for the whole molecule, is quite significant.

Although the molecular volume does not correlate with log k_{cat}, its combination with dipole moment produces a highly significant regression ($r = 0.98$) indicating that the polarity of the substrate is an important contributory factor in explaining the hydroxylation data. This double regression equation (eqn 2) is given in Table 2, where it can be seen that the variance ratio (F-value) is increased considerably and the cross-correlation between the two independent variables is low ($r = 0.29$). Equation (2) (Table 2) can be compared with the two variable expressions derived in the White and McCarthy study, where these co-workers obtained two analogous regressions involving π and either σ or E_R [3]. The molecular volume correlates fairly well with the π parameter ($r = 0.81$) for the compounds under investigation, whereas σ and E_R are electronic substituents parameters that probably estimate, to some degree, the effect of *para*-substitution on the reactivity of the toluene methyl group. However, eqn (2) is of higher statistical significance than those produced in the earlier study, and eqn (1) is of equal significance to the highest one-variable correlation (with π) reported by White and McCarthy [3]. A plot of eqn (2)—calculated log k_{cat} values against observed log k_{cat} data is shown as Fig. 1.

Clearly, the relative electron-withdrawing power of the para substituent affects the rate of methyl hydroxylation, and this lowers the LUMO population of the methyl hydrogens, thus increasing its electron-accepting ability and, hence, suggesting that the rate-determining step in the hydroxylation may be proton abstraction from the methyl group. The electron-withdrawing nature of the para substituent manifests itself in the relative polarity of the methyl toluenes, but the size of that substituent is also of some importance in the explanation of the rate data.

Binding affinity

The correlations between molecular volume and the dissociation constant, K_D, both singly and combined with HOMO populations on the methyl hydrogen atoms, are presented in Table 2 as eqns (3) and (4), respectively. Table 4 lists correlations with other structural descriptors, showing that only log P and molecular volume produced significant relationships with log K_D, of which the latter gave a very good correlation ($r = 0.97$) with dissociation constant. These regressions demonstrate clearly that parameters primarily associated with molecular size and hydrophobicity are associated with binding affinity for the 10 toluene derivatives, whereas electronic factors do not appear to be very important in this process. These findings are essentially in agreement with those reported by White and

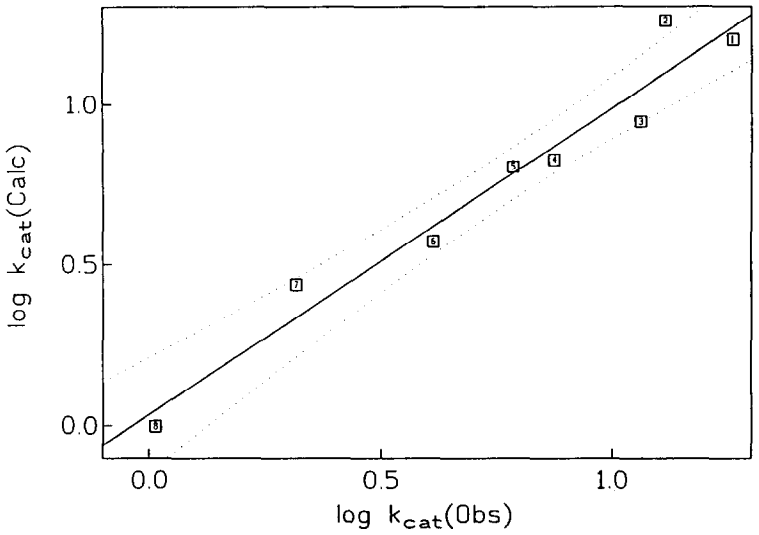

Fig. 1. A plot of calculated log k_{cat} values, using eqn (2), versus observed log k_{cat} data, presented in Table 1. The 95% confidence interval is also shown about the correspondence line.

McCarthy [3] and are of the same degree of statistical significance. However, the regression involving π and molar volume presented by the aforementioned co-workers should be discounted as there is considerable autocorrelation between the two 'independent' variables ($r = 0.81$).

Consequently, the relative degree of binding affinity for the 10 toluenes, as measured by log K_D, can be adequately explained in terms of the volume of the solvent-accessible molecular surface, V, which can be regarded as essentially describing the desolvation contribution to substrate binding and is proportional to the number of water molecules removed from the active site on binding, which is itself a major component of the positive entropy change, ΔS. It is well known that substrate binding to cytochromes P450 is essentially an entropy-driven process. At 21° the thermodynamics of binding the endogenous substrate, camphor, to cytochrome $P450_{cam}$ (CYP101) show that the entropy change is the major component of the free energy change, ΔG [11]. A similar situation has been established for benzphetamine binding to CYP2B4 [12] and the ratios of the entropic components for benzphetamine and camphor are equivalent to those of the molecular volumes for the two compounds. However, both

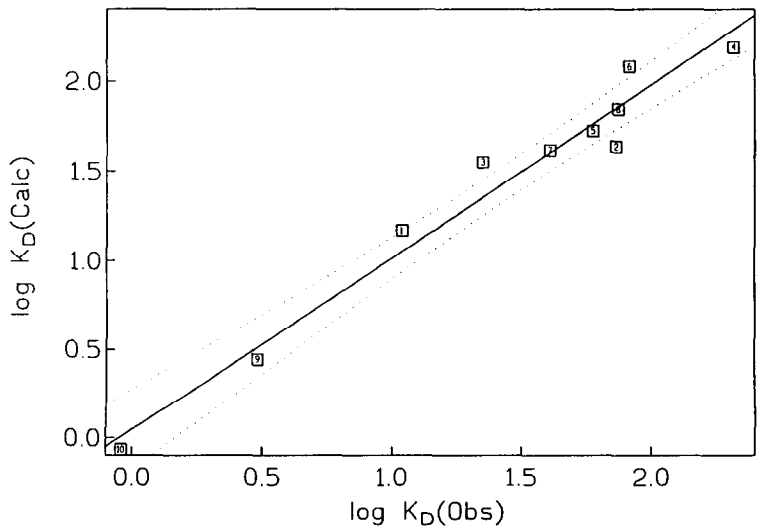

Fig. 2. A plot of calculated log K_D values, using eqn (4), versus observed log K_D data, presented in Table 1. The 95% confidence interval is also shown about the correspondence line.

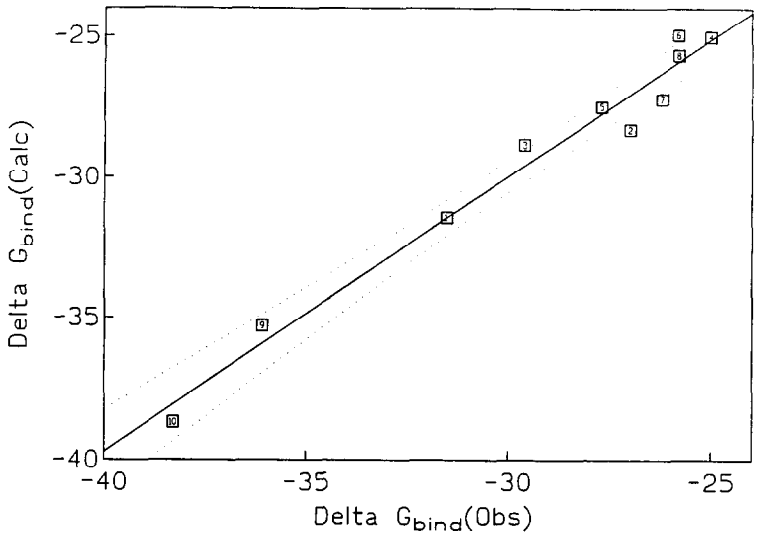

Fig. 3. A plot of calculated ΔG_{bind} values, using eqn (9), versus observed ΔG_{bind} data, presented in Table 1. The 95% confidence interval is also shown about the correspondence line.

eqns (3) and (4) (Table 2) show a negative dependence of molecular volume with log K_D, implying that a bulky substituent reduces the binding affinity, but also there is probably a small electronic contribution, as the HOMO populations of methyl hydrogens (Q_H) slightly improve the correlation (r = 0.98) although the significance is somewhat lower (compare F-values in Table 4). A graphical plot of calculated versus observed values of log K_D, using eqn (4), is presented as Fig. 2.

A similar picture emerges for the correlations with the observed free energy change, ΔG_{obs}, as shown in Table 5 with the most significant regressions presented as eqns (5) and (6) in Table 2. This is not surprising as the two quantities of log K_D and ΔG are linearly related by the Arrhenius equation. However, the high correlation with molecular volume (eqn 5, Table 2) is improved (r = 0.98) by combination with the HOMO population of the methyl hydrogen atoms (Q_H) suggesting that electronic features of the methyl group make some contribution to the binding process.

White and McCarthy [3] were able to estimate the intrinsic binding free energies (ΔG_{bind}) of the 10 toluenes by subtracting the spin-state energy change from the observed overall free energy. Table 6 shows that, although molecular volume is primarily the main descriptor variable of importance, the correlation is significantly improved (r = 0.98) by a combination with either E(HOMO) or Q_L. The relevant regressions are presented as eqns (8) and (9), Table 2, and Table 6 shows that log P also gives a reasonably good correlation (r = 0.91) with ΔG_{bind}. Both eqns (7) and (8) and eqn (9) (Table 2) are of higher significance than those reported by White and McCarthy [3] who found that the reciprocal of dielectric constant improved the correlation with π (r = 0.865). However, dielectric constant is a property of a bulk material rather than one relating to the constituent molecules themselves. Equations (8) and (9) (Table 2) are particularly significant as there is an increase in the variance ratio (F-value) and the better of these is presented graphically as a plot of calculated ΔG_{bind} values against observed ΔG_{bind} data in Fig. 3.

In summary, the results of the present QSAR analysis show, therefore, that substrate binding to CYP2B4 can be well correlated with molecular size, that certain electronic structural parameters improve such correlations, as has been reported in other QSAR analyses of P450-mediated reactions [13] and that catalytic rates may also be similarly rationalized by molecular parameters.

As the molecular volume exhibits a strong correlation (r = 0.986) with the COMPACT parameter, collision diameter (Table 1) for the 10 toluene derivatives in this study, the current QSAR investigation provides additional support for the use of this descriptor in COMPACT evaluations [14–16], which could be discriminatory in assessing differing P450 substrate specificities, whereas molecular volume or surface area may describe the desolvation component of the P450-binding interaction. It is hoped that further studies on the P450 system will elucidate the mechanisms of enzyme–substrate interactions via a structural rationale.

Acknowledgement—The financial support of Glaxo Group Research is gratefully acknowledged.

REFERENCES

1. Nelson DR, Kamataki T, Waxman DJ, Guengerich FP, Estabrook RW, Feyereisen R, Gonzalez FJ, Coon MJ, Gunsalus IC, Gotoh O, Okuda K and Nebert DW, The P450 superfamily: update on new sequences, gene mapping, accession numbers, early trivial names of enzymes, and nomenclature. *DNA Cell Biol* **12**: 1–51, 1993.
2. Waxman DJ and Azaroff L, Phenobarbital induction

of cytochrome P-450 gene expression. *Biochem J* **281**: 577–592, 1992.
3. White RE and McCarthy M-B, Active site mechanics of liver microsomal cytochrome P-450. *Arch Biochem Biophys* **246**: 19–32, 1986.
4. Clark M, Cramer RD and Van Opdenbosch N, Validation of the general purpose Tripos 5.2 force field. *J Computational Chem* **10**: 982–1012, 1989.
5. Connolly ML, Solvent-accessible surfaces of proteins and nucleic acids. *Science* **221**: 709–713, 1983.
6. Dewar MJS and Storch DM, Comparative tests of theoretical procedures for studying chemical reactions. *J Am Chem Soc* **107**: 3898–3902, 1985.
7. Dewar MJS, Zoebisch EG, Healy EF and Stewart JJP, AM1: A new general purpose quantum mechanical molecular model. *J Am Chem Soc* **107**: 3902–3909, 1985.
8. Vinter JG, Davis A and Saunders MR, Strategic approaches to drug design I. An integrated software framework for molecular modelling. *J Computer-Aided Molec Design* **1**: 31–51, 1987.
9. Stewart JJP, MOPAC: A semi empirical molecular orbital program. *J Computer-Aided Molec Design* **4**: 1–105, 1990.
10. Sangster J, Octanol–water partition coefficients of simple organic compounds. *J Phys Chem Ref Data* **18**: 1111–1229, 1989.
11. Griffin BW and Peterson JA, Camphor binding by *Pseudomonas putida* cytochrome P-450. Kinetics and thermodynamics of the reaction. *Biochemistry* **11**: 4740–4746, 1972.
12. Ruckpaul K, Rein H and Blanck J, Regulation mechanisms of the activity of the hepatic endoplasmic cytochrome P-450. *Frontiers Biotransformation* **1**: 1–65, 1989.
13. Hansch C and Zhang L, Quantitative structure–activity relationships of cytochrome P-450. *Drug Metab Rev* **25**: 1–48, 1993.
14. Lewis DFV, Computer-assisted methods in the evaluation of chemical toxicity. *Rev Computational Chem* **3**: 173–222, 1992.
15. Lewis DFV, Computer modelling of cytochromes P-450 and their substrates: A rational approach to the prediction of carcinogenicity. *Frontiers Biotransformation* **7**: 90–136, 1992.
16. Lewis DFV, Moereels H, Lake BG, Ioannides C and Parke DV, Molecular modelling of enzymes and receptors involved in carcinogenesis: QSARs and COMPACT-3D. *Drug Metab Rev* **26**: 261–285, 1994.