Biochemical Pharmacology, Vol. 60, pp. 293–306, 2000.

ISSN 0006-2952/00/$–see front matter
PII S0006-2952(00)00335-X

COMMENTARY

On the Recognition of Mammalian Microsomal Cytochrome P450 Substrates and Their Characteristics

TOWARDS THE PREDICTION OF HUMAN P450 SUBSTRATE SPECIFICITY AND METABOLISM

*David F. V. Lewis**

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

ABSTRACT. The characteristics of mammalian microsomal P450 xenobiotic substrates are described, particularly with reference to the major P450 isoforms associated with drug metabolism in humans. It is further reported that a relatively small number of molecular, electronic, and physico-chemical properties are required to discriminate between chemicals that exhibit specificity for human P450 isoforms: CYP1A2, CYP2A6, CYP2B6, CYP2C9, CYP2C19, CYP2D6, CYP2E1, and CYP3A4. Molecular templates of superimposed substrates are shown to be complementary with the putative active sites of the relevant enzymes, thus enabling a possible prediction of P450 specificity from structure. Factors contributing to metabolic clearance and binding affinity are also discussed, and methods for their calculation are described. BIOCHEM PHARMACOL **60**;3:293–306, 2000.

KEY WORDS. cytochrome P450; substrate selectivity; human P450 isoforms; metabolic clearance; binding affinity

It is well established that the majority of Phase I metabolism of xenobiotics is mediated by CYP† isoforms in both Mammalia and other animal species [1]. Hepatic microsomal P450s of families CYP1, CYP2, and CYP3 are associated with the oxidative metabolism of most xenobiotic compounds, and it has been estimated that the total number of exogenous P450 substrates may exceed 200,000 chemicals [2]. Of the 750 P450s sequenced to date, 36 human isoforms are known, and it is thought that the total may exceed 50, whereas similar numbers have been identified in other mammalian species; for example, the rat has 47 different P450 isoforms [3]. Table 1 summarizes the known substrates, inducers, and inhibitors of the major subfamilies of P450 families (CYP1, CYP2, CYP3, and CYP4) from which it can be appreciated that different P450 isoforms tend to exhibit distinct substrate specificities.

For example, CYP1A isoforms show a preference for planar polyaromatic substrates such as benzo[*a*]pyrene [4], whereas CYP2E substrates are characterized by low molecular weight compounds, such as ethanol and acetone [5]. It is likely that these different substrate structural preferences result from the particular active site characteristics of the P450 isoforms themselves [6, 7], and, from extensive molecular modelling studies, we have demonstrated that known substrates of CYP1A, CYP2A, CYP2B, CYP2C, CYP2D, CYP2E, and CYP3A are able to fit the putative active sites of these enzymes due to complementarity with key amino acid residues in each case [8–14]. It is known, however, that some compounds are able to act as substrates for more than one P450 isoform, and it is common for different P450s to metabolize such chemicals at different positions in the molecule. Omeprazole is one example of such a compound, and we have shown that its metabolism in humans can also be explained via molecular modelling studies [15].

As far as human drug-metabolizing P450s are concerned, the development of substrate structure–activity relationships can lead to the possibility of predicting the likely metabolic fate of novel compounds [6, 7]. In addition to P450 specificity, it is also important to evaluate the probable extent of metabolism, and this involves a calculation of both binding affinity and rate of P450-mediated metabolism [16]. To achieve this aim, we have employed an expression originally formulated by Williams *et al.* [17] to estimate the binding free energies of various P450 substrates [18, 19] and have also noted that QSAR evaluations on several series of structurally related chemicals suggest that rates of metabolism frequently correlate with substrate ionization potentials [18, 20]. The present work involves an analysis of the characteristics of human P450 substrates from families CYP1, CYP2, and CYP3 to define more precisely the specific determinants of human P450 isoform selectivity.

* Correspondence. Tel. (44) 1483 300800; FAX (44) 1483 300803; E-mail: d.lewis@surrey.ac.uk

† *Abbreviations*: CYP, cytochrome P450; QSAR, quantitative structure–activity relationship; and MO, molecular orbital.

TABLE 1. Substrates, inhibitors, and inducers of mammalian P450 isoforms from families CYP1–4*

CYP	Substrates	Inhibitors	Inducers	Receptor involvement
1A	7-Ethoxyresorufin	9-Hydroxyellipticine	PAHs	AhR
2A	Coumarin	Pilocarpine	PB	GR?†
2B	Phenobarbital	Orphenadrine	PB	CAR, GR?
2C	Mephenytoin	Sulfaphenazole	PB	RAR, GR
2D	Debrisoquine	Quinidine	Non-inducible	None
2E	*p*-Nitrophenol	Disulfiram	Ethanol	Unknown
3A	Dexamethasone	Ketoconazole	Dexamethasone	GR, PXR
4A	Lauric acid	11-Imidazolyl lauric acid	Clofibrate	PPAR

References: [21, 22, 65–68]

*The majority of the compounds listed can be applicable to several mammalian species, including *Homo sapiens*. Abbreviations: PAHs, polycyclic aromatic hydrocarbons; PB, phenobarbital; AhR, aryl hydrocarbon receptor; GR, glucocorticoid receptor; PXR, pregnane X receptor; PPAR, peroxisome proliferator-activated receptor; CAR, constitutive androstane receptor; and RAR, retinoic acid receptor.

†The question mark (?) indicates that the involvement of this receptor could be inferred from the presence of response elements within the regulatory region of the relevant gene, although direct evidence has not been reported. Induction of the PB-inducible P450s may, in fact, occur via regulation of a labile repressor protein that can bind to the basal transcription element, although there is recently reported evidence for a nuclear receptor involvement in PB-associated induction of P450s.

METHODS

For each of the major known drug-metabolizing P450 isoforms, six relatively specific substrates (shown in Table 2) were selected from previously compiled studies on human P450 substrates [21–23]. Physico-chemical data in the form of lipophilicity (log P and log $D_{7.4}$) and ionization constant (pK_a) values were either collated from the literature [24–28] or calculated via the Pallas system (CompuDrug Limited). Molecular and electronic structural data for the 48 compounds were calculated using Sybyl 6.4 (Tripos Associates), and included MO calculations via the AM1 method [29]. Each compound was docked interactively within the putative active site of the human P450 isoform most closely associated with its metabolism, using molecular models of the relevant enzymes described previously [8–13, 19]. For each P450 isoform concerned, the six substrate molecules were superimposed within the putative active site such that maximum complementarity with amino acid residues likely to be involved in binding interactions was achieved, together with optimal positioning of the site of metabolism relative to the haem iron of the P450. This procedure was guided by the location of palmitoleic acid in the recently reported crystal structure of substrate-bound CYP102 [30] and from previous experience in the construction of substrate templates for various P450 isoforms [9–13, 19]. The various structural characteristics of the 48 P450 substrates were examined, following the generation of MolCAD (Tripos Associates) surfaces coded by molecular electrostatic potential energy on both individual compounds and substrate templates, in order to establish particular criteria governing selectivity for each human P450 isoform.

In addition, the clearance data on 12 compounds comprising two series of structurally related chemicals (benzodiazepines and sodium channel blockers) were taken from a recent comparative study [31], and QSARs were generated using a combination of physico-chemical and structural properties (Table 3) to establish the relative importance of various compound properties in metabolic clearance mediated by P450 isoforms. In these QSAR studies, the same methods as those described above were employed for the calculation of structural descriptors and physico-chemical characteristics.

RESULTS AND DISCUSSION

(1) General Distinguishing Characteristics of P450 Substrates

Table 2 provides information on the physico-chemical properties and molecular structural characteristics of 48 substrates of CYP1A2, CYP2A6, CYP2B6, CYP2C9, CYP2C19, CYP2D6, CYP2E1, and CYP3A4. A number of clearly discriminating factors are apparent from an inspection of the material presented in Table 2. In particular, the overall molecular planarity (area/depth^2) is relatively marked in CYP1A2 substrates compared with the majority of other chemicals, although some exceptions exist.

For example, coumarin is a fairly specific substrate for CYP2A6 and shows a high planarity typical of CYP1A substrates. However, it is known that coumarin can be metabolized by other P450s, including CYP1A, and from inspection of Table 2 it is apparent that coumarin is somewhat atypical when compared with other CYP2A6 substrates, which generally exhibit substantially lower planarities. In addition to overall molecular planarity, the CYP1A2 substrates possess relatively low ΔE values, although this factor in itself is not discriminatory. A further finding from Table 2 concerns the variation in molecular size, as exemplified by both surface area and volume of the solvent-accessible surfaces. Those of the CYP1A2 substrates fall into a fairly narrow range, which can also be regarded as of medium magnitude when compared with, for example, those of CYP2E1 and CYP3A4 substrates that are themselves relatively small and large, respectively (Table 2). These findings are reflected in the relative molecular masses of the chemicals concerned, in addition to the

TABLE 2. Data for human P450 substrates

	log P	pK_a†	log $D_{7.4}$	a/d^2	l/w	SA	Vol	ΔE	E_{LUMO}	μ
1. CYP1A2										
Caffeine	0.08	Neutral	0.08	5.106	1.139	190.22	174.33	8.6021	−0.3448	3.4490
PhIP‡	2.23	8.69[b]	0.92	5.148	1.952	228.55	213.59	8.4878	−0.2199	4.3344
7-Methoxyresorufin	3.15	9.94[b]	0.64	4.609	1.883	221.88	203.51	7.4266	−1.6707	4.3911
Phenacetin	1.57	Neutral	1.57	5.118	2.017	208.79	183.91	8.7223	0.3635	3.0950
Tacrine	2.71	9.8[b]	0.36	3.183	1.459	212.84	203.69	8.1645	−0.2534	2.1640
IQ‡	1.84	9.18[b]	0.06	5.294	1.484	200.94	186.08	7.8889	−0.4215	3.4515
2. CYP2A6										
Coumarin	1.39	Neutral	1.39	6.738	2.929	152.67	133.08	8.5287	−0.9313	4.2692
Losigamone	1.46	Neutral	1.46	1.678	1.543	234.61	226.70	9.5150	−0.3150	4.0459
SM-12502	1.02	4.46[b]	1.02	1.595	1.489	209.34	204.27	8.5819	−0.4858	0.8560
Fadrozole	2.79	6.16[b]	2.77	1.835	1.480	232.79	227.83	8.2534	−0.7067	3.6289
Cotinine	0.07	4.44[b]	0.07	1.360	1.273	190.32	181.27	9.5498	−0.1467	1.6650
534U87	2.20	6.04[b]	2.18	1.863	1.707	237.70	226.17	8.0853	−0.4952	1.8609
3. CYP2B6										
4-Trifluoromethyl EC*‡	3.31	Neutral	3.31	3.083	1.379	229.86	212.00	8.1876	−1.3459	3.8658
Bupropion	2.54	8.35[b]	1.54	1.802	1.120	254.07	254.78	8.8666	−0.6235	2.1291
Cyclophosphamide	0.51	Neutral	0.51	1.808	1.029	235.79	233.18	11.2819	0.6327	3.5430
Ifosfamide	0.86	Neutral	0.86	2.620	1.003	238.76	232.48	10.9334	0.5199	3.1131
7-Pentoxyresorufin	5.14	10.06[b]	2.50	2.577	1.993	300.19	278.87	7.2014	−1.6870	4.5153
Antipyrine	0.23	Neutral	0.23	2.011	1.562	203.62	188.71	8.6238	−0.1603	3.9330
4. CYP2C9										
Tolbutamide	2.34	5.43[a]	0.37	2.573	2.523	280.37	260.13	9.3253	−0.8189	7.0221
Diclofenac	4.40	4.22[a]	1.22	1.466	1.190	257.17	262.62	8.7552	−0.2647	1.6864
Tienilic acid	3.15	4.8[a]	0.55	2.635	1.922	276.19	272.17	8.8443	−0.8578	4.4238
Ibuprofen	3.51	4.5[a]	1.07	2.092	1.535	239.49	231.98	9.5994	0.2013	1.8386
S-Warfarin	2.52	5.1[a]	0.12	1.734	1.383	285.05	291.99	8.4476	−0.8577	3.6629
Naproxen	3.18	4.15[a]	0.33	1.845	1.620	241.82	228.33	8.2505	−0.3931	2.2335
5. CYP2C19										
S-Mephenytoin	1.74	8.1[b]	0.96	1.865	1.129	223.45	218.90	9.8861	−0.0608	2.5168
Hexobarbital	1.49	8.2[b]	0.63	1.287	1.217	230.09	234.13	9.7495	−0.1303	2.0137
R-Mephobarbital	1.86	7.8[b]	1.31	1.306	1.141	233.52	236.94	9.7388	−0.1986	1.9466
Proguanil	2.53	10.4[b]	−0.47	2.784	1.384	263.33	246.67	8.8773	−0.1612	2.1785
Moclobemide	2.13	7.09[b]	1.96	3.002	2.343	275.86	262.03	8.9332	−0.4901	2.4491
Omeprazole	2.23	8.7[b]	0.91	1.841	1.035	340.19	330.37	8.2444	−0.3993	3.2240
6. CYP2D6										
Debrisoquine	0.75	13.01[b]	0.75	2.005	1.488	194.46	181.19	9.5405	0.4072	1.5615
Metoprolol	2.35	9.68[b]	0.07	1.703	1.446	310.45	303.15	9.3137	0.1657	1.1873
Bufuralol	3.50	9.0[b]	1.89	2.196	1.573	292.55	290.87	8.7460	0.0606	1.9560
Sparteine	2.13	11.8[b]	−2.27	1.649	1.268	250.06	271.25	11.5799	2.8428	0.3968
Codeine	1.07	8.2[b]	0.23	1.343	1.327	273.96	297.89	8.9460	0.3400	2.0116
Dextromethorphan	3.36	8.3[b]	0.91	1.490	1.155	273.14	299.54	9.2159	0.6788	1.1944
7. CYP2E1										
4-Nitrophenol	1.91	Neutral	1.91	6.009	1.363	139.82	119.01	9.0070	−1.0652	5.1168
Diethylnitrosamine	0.48	Neutral	0.48	1.522	1.327	134.55	115.33	10.8885	0.9370	2.8324
Aniline	0.90	4.7[b]	0.90	5.349	1.219	121.11	99.74	9.1636	0.6359	1.2790
Ethanol	−0.32	Neutral	−0.32	1.829	1.331	79.36	56.09	14.4412	3.5650	0.9200
Paracetamol	0.25	9.5[a]	0.25	3.957	1.603	169.03	146.50	8.7114	0.2508	2.3131
Chlorzoxazone	2.36	8.3[b]	1.41	5.711	1.635	155.70	134.95	8.8842	−0.5480	2.0617
8. CYP3A4										
Nifedipine	1.96	Neutral	1.96	1.735	1.199	311.02	325.29	8.3144	−0.5322	2.6233
Erythromycin	2.48	8.8[b]	1.06	2.641	1.330	501.08	634.04	10.0472	0.8507	4.9621
Cyclosporin A	2.92	7.97[b]	2.25	2.896	1.181	864.35	1318.37	10.7803	0.8079	11.0130
Lidocaine	2.26	7.85[b]	1.68	2.244	1.343	260.66	272.09	9.2559	0.4132	2.8288
Dapsone	0.97	Neutral	0.97	2.014	1.540	238.58	228.76	8.6327	−0.2536	5.1632
Midazolam	1.53	4.93[b]	1.53	1.988	1.076	290.11	293.91	8.1995	−0.7595	3.4458

*References to physico-chemical data: [69–75]. Key to column headings:
log P = logarithm of the octanol/water partition coefficient.
pK_a = negative logarithm of the acid/base dissociation constant.
log $D_{7.4}$ = logarithm of the octanol/water distribution coefficient at pH 7.4.
a/d^2 = ratio of molecular area and depth squared.
l/w = ratio of molecular length and width.
SA = surface area of the solvent-accessible surface.
Vol = volume of the solvent-accessible surface.
ΔE = difference between the frontier orbital energies.
E_{LUMO} = energy of lowest unoccupied molecular orbital.
μ = molecular dipole moment.
†a = acid, and b = basic.
‡Abbreviations: PhIP = 2-amino-1-methyl-6-phenylimidazo[4,5-*b*]pyridine; IQ = 2-amino-3-methylimidazo[4,5-*f*]quinoline; and EC = 7-ethoxycoumarin.

TABLE 3. Structural and clearance data for benzodiazepines and sodium channel blockers*

	Compound	IP (eV)	Energy	HBD	Hf	ΔE	log P	log $D_{7.4}$	log CL
(A) Benzodiazepines									
1.	Oxazepam	9.618291	10.691	2	16.883085	8.8180	3.27	2.2	1.2553
2.	Temazepam	9.566280	12.488	1	11.619133	8.8098	3.17	2.3	1.6232
3.	Diazepam	9.439194	9.942	0	59.573968	8.7911	3.21	2.9	1.5563
4.	Clonazepam	9.860432	0.979	1	72.407753	8.0925	2.51	2.4	1.0414
5.	Flurazepam	8.514453	9.884	0	31.750856	7.9376	4.20	2.3	3.4771
6.	Medazepam	8.628606	13.968	0	92.913936	8.3929	4.12	4.4	3.3979
(B) Na^+ channel blockers									
7.	Bupivacaine	8.860139	10.757	1	−21.810096	9.2740	4.47	2.5	2.1461
8.	Lidocaine	8.844312	3.989	1	−3.880583	9.2632	2.82	1.6	1.4914
9.	Tocainide	9.267788	5.184	3	−12.394998	9.5989	1.67	0.3	0.2304
10.	Flecainide	9.616386	2.944	2	−381.817818	8.8064	3.81	1.1	0.6021
11.	Procainamide	8.342090	−17.454	3	−12.316804	8.8802	2.06	−0.8	0.0792
12.	Mexilitene	9.358473	0.535	2	−21.078544	9.6353	2.46	0.35	1.0792

	Regression Equations	n	s	*R*	F
1.	$\log CL = 0.64 \log D_{7.4} - 0.98\ IP + 9.33$ (± 0.11) (± 0.30)	12	0.498	0.91	21.39
2.	$\log CL = 0.094\ Energy - 1.18\ IP - 0.74\ \Delta E + 18.65$ (± 0.015) (± 0.25) (± 0.23)	12	0.400	0.95	24.10
3.	$\log CL = 0.055\ Energy - 0.95\ IP - 0.53\ HBD + 10.63$ (± 0.020) (± 0.25) (± 0.15)	12	0.382	0.95	26.65
4.	$\log CL = 0.067\ Energy - 1.01\ IP - 0.34\ HBD - 0.43\ \Delta E + 14.66$ (± 0.019) (± 0.23) (± 0.18) (± 0.26)	12	0.346	0.97	25.01
5.	$\log CL = 0.65 \log P - 0.40\ IP - 0.37\ HBD + 0.0025\ Hf + 3.63$ (± 0.26) (± 0.28) (± 0.22) (± 0.0015)	12	0.418	0.95	16.57

*Molecular orbital calculations were conducted via the AM1 method [29]. Key to column headings:
IP = ionization potential (eV).
Energy = geometry-optimized minimum internal energy (kcal · mol^{-1}).
HBD = number of potential hydrogen bond donor atoms in the molecule.
Hf = enthalpy of formation (kcal · mol^{-1}).
$\Delta E = E_{LUMO} - E_{HOMO}$ (eV), where: E_{LUMO} and E_{HOMO} are the lowest unoccupied and highest occupied molecular orbitals, respectively.
log P = calculated logarithm of the octanol/water partition coefficient (Pallas system, CompuDrug Limited).
log $D_{7.4}$ = experimental logarithm of the octanol/water distribution coefficient [31].
log CL = logarithm of the free metabolic clearance (mL/min/kg) rates [31].

general range of log P values. For example, CYP3A4 substrates usually possess relatively high log P values indicative of fairly lipophilic molecules [32], whereas the values of the CYP2E1 substrates tend to be rather low, as these compounds are often polar, and all are small molecular structures [13]. However, the volume of the solvent-accessible surface probably represents the most satisfactory means of discriminating between these substrates, and it is also not affected by either molecular polarity or shape.

Other differences emerge when one considers the acid–base characteristics of the human hepatic microsomal P450 substrates listed in Table 2. For example, the CYP2C9 substrates are all weakly acidic, whereas those of CYP2D6 are basic in character [33]. In fact, it is also possible to differentiate between substrates of CYP2C9 and the related CYP2C19 on the grounds of their acid–base properties [14], although it is likely that there are other factors involved, such as the disposition of hydrogen bond donor/acceptors in the molecule, and this principle can be extended to other P450 substrates.

For CYP2A6 substrates, for example, there are two hydrogen bond acceptors situated a relatively short distance apart (2.5 Å) in virtually all compounds, and this may be due to the presence of two key hydrogen bond-donating amino acid residues located close to the heme moiety in the active site of the enzyme [34].

Consequently, although there are similarities between certain characteristics of CYP2A6 and CYP2B6 substrates, for example, the two types can be differentiated in terms of the variations in their disposition of hydrogen bond donor/acceptor atoms. However, leaving aside the issue of hydrogen bond donors and acceptors within the molecule, it is possible to achieve a fairly satisfactory differentiation between the 48 chemicals using only three properties, namely volume, planarity, and acid–base character.

Figure 1 sets out a decision tree approach that can be used to discriminate between these compounds, such that they may be classified in terms of their known P450 specificity. It should be emphasized, however, that there are some exceptions to this procedure, although, in most of

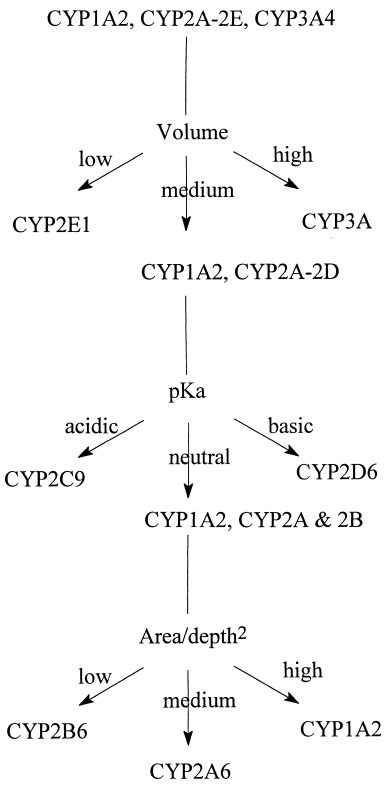


FIG. 1. A decision tree for human P450 substrates. Notes: (1) CYP2B6 substrates tend to be of higher volume than those of CYP2A6; (2) CYP2C19 substrates can be differentiated from those of CYP2C9 on the basis of their increased basicity; (3) the overlap between CYP2C19 specificity and CYP2B6 substrates can be separated via log P values; (4) CYP2C19 may be differentiated from CYP3A4 on the grounds of volume; and (5) the generally greater basicity of CYP2D6 substrates facilitates their discrimination from CYP2C19 substrates.

these cases, the chemicals concerned can act as substrates for more than one P450. Potential difficulties arise with the distinction between CYP2A6 and CYP2B6 substrates, but, by and large, the CYP2A6- specific compounds are generally more planar than those of CYP2B6 and, furthermore, CYP2B6 appears to exhibit a preference for higher volume substrates than does CYP2A6, as can be appreciated from an inspection of Table 2. There is, moreover, some degree of overlap between CYP2C19 substrates and those of CYP2D6 and CYP2B6, which can be surmounted by consideration of the greater basicity of CYP2D6 substrates, whereas log P values may be employed for differentiating between CYP2B6 and CYP2C19. It should be recognized, however, that omeprazole, for example, is a substrate of both CYP2C19 and CYP3A4, such that the scheme proposed in Fig. 1 may not necessarily apply for all chemicals, but rather should be used as an initial screen. It is felt, therefore, that active site modelling should be employed as a second stage in a screening programme.

However, it is possible to estimate the binding affinity, rate of oxidation, and metabolic clearance solely on the basis of substrate molecular and electronic properties, as follows.

(2) A Theoretical Treatment of Clearance and Binding Affinity

It is possible to derive expressions for metabolic clearance based on the following relationship governing systemic clearance (CL_S):

$$CL_S = \frac{Q\ CL_{int}}{(CL_{int} + Q)}$$

where Q is the organ blood flow, which exhibits an inverse correlation with body weight across several mammalian species, and CL_{int} is the intrinsic clearance [32].

As the organ blood flow term represents a means of extrapolating from one species to another based on an allometric scaling via body weight, it is important to focus on the intrinsic clearance, which can be formulated as follows:

$$CL_{int} = V_{max}/K_m$$

where V_{max} is the maximal velocity (or rate) of the metabolic process, and K_m is the apparent Michaelis constant of the enzyme and substrate concerned [32].

Taking the logarithm of both sides produces a linear relationship as shown below:

$$\log CL_{int} = \log V_{max} - \log K_m$$

where the catalytic turnover number, k_{cat}, for the substrate metabolism is related to V_{max} via the total enzyme concentration, $[E_T]$, as follows:

$$k_{cat} = \frac{V_{max}}{[E_T]}$$

and the catalytic efficiency (or specificity constant) for the enzyme–substrate pair is given by an expression related to that of intrinsic clearance, as follows:

$$\text{specificity constant} = k_{cat}/K_m$$

From this it can be appreciated that low K_m values will have an enhanced effect on both specificity and clearance, especially as most microsomal P450s exhibit relatively low turnover numbers.

(a) THE RATE CONSTANT. According to the Eyring equation, it is possible to describe the rate constant, k, for a given reaction as follows:

$$k = \frac{RT}{Nh} \exp(-\Delta H^{\ddagger}/RT) \exp(\Delta S^{\ddagger}/R)$$

where R is the gas constant, T is the absolute temperature, N is the Avogadro number, h is Planck's constant, and $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ are enthalpy and entropy changes, respectively, for

the formation of an activated complex between enzyme and substrate.

Taking logs of both sides simplifies this expression to yield a linear equation of the form:

$$\log k = \log\left(\frac{RT}{Nh}\right) - \frac{\Delta H^{\ddagger}}{2.3RT} + \frac{\Delta S^{\ddagger}}{2.3R}$$

so that for a reaction at 37°, the relationship can be formulated as:

$$\log k = 12.81 - \frac{\Delta H^{\ddagger}}{1.418} + \frac{\Delta S^{\ddagger}}{4.576}$$

where the terms in $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ contain factors calculated in kcal · mol^{-1} and cal · mol^{-1}, respectively.

The problem arises, however, when one considers evaluation of the thermodynamics of activated complex formation and, hence, estimation of the $\Delta H^{\ddagger}$ and $\Delta S^{\ddagger}$ values. The usual procedure is to assume that these quantities, especially $\Delta S^{\ddagger}$, are approximately equivalent to that of product formation, although this may be an unrealistic assumption in the case of P450-mediated reactions. However, it is now possible to calculate $\Delta H^{\ddagger}$ directly using MO procedures such as AM1, which enables one to derive values for the enthalpies of formation of the reactant molecule and its activated intermediate, giving:

$$\Delta H^{\ddagger} = Hf_{\text{intermediate}} - Hf_{\text{reactant}}$$

where *Hf* is the MO-calculated heat of formation for the reactive intermediate and original reactant (substrate) molecule in kcal·mol^{-1}.

For a series of 4-substituted toluenes, it can be shown that the log rate constant for P450-mediated hydroxylation correlates closely ($r = 0.99$) with the AM1-calculated $\Delta H^{\ddagger}$ for a hydrogen-abstracted intermediate [20]. However, the ionization potential values for the reactive intermediate were combined with the enthalpies of formation of this species to give rise to such a good agreement between theory and experiment. In fact, several other independent studies conducted on a different series of P450 substrates (reviewed in Lewis *et al.* [18]) report that ionization potential (or the related quantity E_{HOMO}) gives good correlations with log rate constant, thus indicating the importance of this quantity as a determinant of the relative rate of P450-mediated reactions. A possible rationale for the appearance of substrate ionization potential in correlations with rate could center around the likely mechanistic scheme for P450 oxygenations, where an electrophilic oxygenating species has been postulated (as reviewed by Lewis and Pratt [20]). Additionally, a more complete analysis involves consideration of the $\Delta S^{\ddagger}$ term, which may be estimated from the likely desolvation component of the binding interaction between substrate and P450 isoform. For example, in the series of toluenes mentioned above, it is found that the substrate solvent-accessible surface area will improve the aforementioned correlation with log rate constant from $r = 0.987$ to $r = 0.993$ [18].

(b) BINDING AFFINITY. The binding affinity between substrate and enzyme as measured by enzyme kinetics determinations is an equilibrium constant, namely, the Michaelis constant, K_m. Consequently, it can be expressed in terms of the free energy change accompanying substrate binding, ΔG_{bind}, as follows:

$$\Delta G_{bind} = RT \ln K_m$$

where R is the gas constant and T is the absolute temperature. For enzymic reactions where the rate constant for the formation of the enzyme–substrate complex is large compared with that of product formation, the K_m value is approximately equivalent to the dissociation constant, K_D, and, in most P450-mediated reactions, one finds that $K_m \cong K_D$, although variations do exist.

There are several factors that contribute to the enzyme–substrate binding free energy, ΔG_{bind}, and various theoretical treatments have been reported for ligand–protein binding, which show satisfactory agreement with experimental findings [17, 35, 36]. Broadly, the various components to the binding energy can be classified in terms of ionic, hydrogen bond, lipophilic or desolvation, charge-transfer/π-π stacking, and conformational and internal energy contributions, together with the loss of translational and rotational freedom that occurs on binding. Essentially, the favourable contributions arising from the formation of intermolecular attractions between substrate and enzyme are partially offset by the unfavourable reduction in degrees of freedom experienced by the substrate upon binding. Furthermore, there is usually a substantial entropic term, which accompanies substrate binding to the P450 active site, stemming from the expulsion of enzyme-bound water molecules. This so-called desolvation term is generally regarded as representing the major component of the enzyme–substrate binding affinity in most P450-mediated reactions.

(i) Desolvation or Lipophilic Term. One means of evaluating the desolvation term in the binding energy expression is to compare log P values for various hydrocarbons with their solvent-accessible surface areas and/or volumes. In Table 4, for example, a small number of polyaromatic hydrocarbons have been presented together with their experimental log P values and solvent-accessible surface data. Equations 1 and 2 in this table (i.e. Table 4) show that both surface area and volume, respectively, give very good correlations with log P_{oct}, where the latter appears to be slightly better than the former. According to Simon *et al.* [35], the partitioning free energy, ΔG_{part}, for a solute distributed between *n*-octanol and water can be expressed as follows:

$$\Delta G_{part} = -RT \ln P_{oct}$$

TABLE 4. Physico-chemical data for eight aromatic hydrocarbons

	Compound	Surface area (Å^2)*	Volume (Å^3)†	log P_{oct}‡
1.	Benzene	96.4379	69.7669	2.13
2.	Naphthalene	136.5117	108.1560	3.37
3.	Anthracene	177.1206	146.7754	4.45
4.	Phenanthrene	175.5002	146.5591	4.46
5.	Pyrene	187.6484	160.98145	5.08
6.	Chrysene	214.6294	185.2215	5.91
7.	Benz[*a*]anthracene	215.9908	185.0756	5.69
8.	Dibenz[*a*,*h*]anthracene	254.9842	223.5056	6.93

	Regression equations§	n	s	*R*	F
1.	log P = 0.0263 SA (± 0.0019)	8	0.246	0.989	302.7
2.	log P = 0.031 Vol (± 0.0007)	8	0.090	0.999	2326.3

*Surface area (SA): solvent-accessible surface area (Å^2) based on a sphere radius of 1.4 Å.

†Volume (Vol): volume of the solvent-accessible surface (Å^3) based on a sphere radius of 1.4 Å.

‡log P_{oct}: experimental octanol/water partition coefficients, expressed as logarithmic values [24].

§Both of the above equations have been forced through the origin for the purposes of comparison and evaluation of the factors governing free energies of solvation and partition (see text for details).

which becomes:

$$\Delta G_{part} = -1.36356 \log P_{oct}$$

based on a temperature of 298K (25°).

If one makes the assumption that, for purely hydrophobic molecules, the free energy of desolvation (ΔG_{desol}) is equivalent to the partitioning energy, then the relationship between surface area and ΔG_{desol} should be derivable from the above expression. It is generally accepted [37] that the desolvation free energy for hydrocarbons is related to their solvent-accessible surface areas, as follows:

$$\Delta G_{desol} = -0.025 \text{ SA}$$

where the factor is designed to give ΔG in kcal·mol^{-1} for surface areas measured in square Ångstroms. However, Sharp and coworkers [37] have argued that the factor should be increased to −0.047. Taking an average of these two values would give rise to an expression of the form:

$$\Delta G_{desol} = -0.036 \text{ SA}$$

which can also be referred to as the hydrophobic component, or entropy term, of the overall binding energy [17].

Comparing the two equations for ΔG_{part} and ΔG_{desol} yields the following expression for the relationship between log P_{oct} and SA, assuming an equivalence for the two free energy changes:

$$1.36356 \log P_{oct} = 0.036 \text{ SA}$$

which gives rise to the equation:

$$\log P_{oct} = 0.0264 \text{ SA}$$

such that a factor of 0.0264 should relate the surface areas of hydrocarbon solutes to their log P_{oct} values. In fact, this finding is in striking agreement with the value of 0.0263 produced by statistical correlation between experimental log P_{oct} and calculated surface areas of the polyaromatic hydrocarbons presented in Table 4, equation 1. If one considers the free energy of partitioning at 310K, however, then a factor of 0.0263 is obtained, provided that the expression for the desolvation energy dependence on hydrocarbon surface area is as follows:

$$\Delta G_{desol} = -0.037 \text{ SA}$$

which represents only a modest increase over the previously employed average value of 0.036 kcal · mol^{-1} Å^{-2} for relating free energy of desolvation to solvent-accessible surface area. The correlation ($R = 0.999$) between log P_{oct} and volume of the solvent-accessible surface is significantly better than that obtained for surface area (compare equations 1 and 2, Table 4), and may reflect the cavity term component to the partitioning or solvation energy [38]. Moreover, consideration of the factor of 0.031 describing the dependence of volume of log P_{oct} suggests that, to relate molecular volumes with desolvation energy, a value of 0.0425 should accord with the QSAR (or, strictly, QSPR) equation based on a temperature of 298K. For the higher temperature of 300K, however, the factor would be 0.044 kcal·mol^{-1} Å^{-3} for evaluating desolvation free energies from the molecular volumes of solvent-accessible surfaces. It should be recognized, however, that this expression may only apply rigorously to polyaromatic hydrocarbons, and, for polyaromatic hydrocarbon derivatives, the presence of polar atoms could affect the relationship.

(ii) Electrostatic or Ionic Term. Based on Coulomb's Law, the energy of attraction between two oppositely charged species, A and B, can be expressed as follows:

$$\Delta G_{ionic} = \frac{-q_A q_B}{4\pi\varepsilon_O \varepsilon_r d}$$

where q_A and q_B are the charges on the two interacting species separated at a distance, d, in a medium of relative permittivity (dielectric constant), ε_r, and ε_o is the permittivity of free space.

By evaluating the constant terms and converting to kcal·mol^{-1} [35] for interionic distances in Ångstrom units, this expression becomes:

$$\Delta G_{ionic} = \frac{-331.9 q_A q_B}{\varepsilon_r d}$$

where the dielectric constant, ε_r, will vary depending on the local environment and can have a marked effect on the magnitude of the electrostatic energy contribution. Pure water at 298K has a dielectric constant of 78.54 [39], which exhibits a negative dependence with temperature of about −0.36 per degree [40]. In the essentially hydrophobic environment encountered for a globular protein, however, the local dielectric constant is much diminished, being about 4 at the protein interior and around 28 at the surface of the protein [41]. It is common, therefore, for protein modellers to employ a value of 4 for the local dielectric constant of an enzyme–substrate complex, including those involving heme enzymes [42], and a distance-dependent function may also be used [43] to describe the increase in dielectric constant from protein interior to surface regions. For more precise calculations, however, the above function should be corrected for repulsive forces between interacting species, such that the expression then becomes as follows:

$$\Delta G_{ionic} = \frac{-304.2 q_A q_B}{\varepsilon_r d}$$

In the presence of electrolytes, an additional term should be included to account for the shielding effect on the overall electrostatic energy [44] according to the Debye–Huckel theory, as follows:

$$\Delta G_{ionic} = \frac{-304.2 q_A q_B}{\varepsilon_r d} \exp(-Kr)$$

where K is the Debye–Huckel constant and has a value of 1.25 nm^{-1} for 0.15 M NaCl [44].

Due to the diminished effect of the dielectric constant at protein interiors, ionic interactions usually have a profound influence on the enzyme–substrate binding affinity. As far as P450–substrate interactions are concerned, such considerations are of major importance in the case of CYP2D6, where it has been shown [45] that a key electrostatic interaction between aspartate-301 and positively charged basic substrates is the main contributor to the overall binding affinity. In this instance, it would appear that the ion-paired contact between enzyme and substrate contributes about −4.5 kcal · mol^{-1} to the total binding energy, based on the effect of site-directed mutagenesis on aspartate-301 [45].

(iii) Dipolar and Hydrogen Bond Terms. Hydrogen bonding can be regarded as a special case of dipole–dipole interaction, and it is possible to ascribe a particular energy function for a hydrogen-bonded interaction based on a variant of the Lennard–Jones [6–12] potential, as follows:

$$E_{Hbond} = \frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}$$

where C_{ij} and D_{ij} are functions of the depth of the potential well and equilibrium distance of the hydrogen bond donor–acceptor interaction between atoms i,j separated at a distance r_{ij} [46]. However, the expression for a typical dipolar interaction is shown below, where it can be seen that the energy is inversely proportional to the cube of the distance between the two dipoles:

$$E_{dipole} = \frac{2\mu_1\mu_2 \cos\theta}{4\pi\varepsilon_o d^3 \varepsilon_r}$$

where μ_1 and μ_2 are the dipole moments of the two interacting species aligned at an angle θ and separated by a distance d apart [47].

Simplification of this expression and evaluation of the constant terms leads to the following equation for a dipole–dipole interaction:

$$E_{dipole} = \frac{14.4\ \mu_1\mu_2 \cos\theta}{\varepsilon_r d^3}$$

which gives energy values in kcal · mol^{-1} for distances in Ångstrom units.

As a hydrogen bond is essentially electrostatic in nature, however, it is possible to derive hydrogen bond energy values using the standard expression for an electrostatic interaction given in the previous section. In addition, quantum-mechanical analysis of hydrogen bonding in the liquid state has yielded the following empirical expression for determining hydrogen bond energies:

$$E_{Hbond} = \frac{-6.86\ \mu_{XH} \Delta I_y}{\varepsilon_r d_{xy}}$$

where the hydrogen bond energy in kJ · mol^{-1} can be evaluated for an X-H···Y system of XH dipole moment, μ_{XH}, and XY distance of d_{xy} in a medium of dielectric constant ε_r. The difference in first ionization potential between the element Y and the corresponding inert gas of the same period is represented by ΔI_Y [47].

Hydrogen bonds vary in energy from about −1 to −5 kcal · mol^{-1} in biological systems [41], with −2 kcal · mol^{-1} representing a reasonable average value, although the hydrogen bond energy is susceptible to changes in local dielectric constant and will alter markedly depending on the electronegativity of the atoms involved. For the substrate-bound cytochrome $P450_{cam}$ system, it is possible

to calculate the dipole–dipole or hydrogen bond energy between the camphor substrate and an active site tyrosine residue (Tyr-96), which lies at an optimal distance for hydrogen bond formation, as evidenced by the 1.63 Å resolution crystal structure [48]. For example, using empirically calculated point charges of +0.217 on the tyrosine hydroxyl hydrogen and −0.297 for the camphor carbonyl oxygen at a distance of 1.711 Å, the electrostatic energy is -3.125 kcal · mol^{-1} assuming a local dielectric constant of 4. However, based on a purely dipolar interaction, the energy evaluates to either -1.714 or -2.202 kcal · mol^{-1} depending on whether experimental or calculated dipole moments are used. In this case, the angle between the dipoles is 120.86° and the experimental dipole moments are 1.5 and 3.1 for tyrosine and camphor, respectively [49], whereas the calculated dipole moments of tyrosine and camphor are, respectively, 2.195 and 2.721 based on an empirical charge calculation. Employment of the specific expression for a hydrogen bond energy based on the hydroxyl bond dipole moment of 1.51 and a difference in ionization potential between oxygen and neon of 770 kJ · mol^{-1}, the camphor–tyrosine hydrogen bond energy can be calculated as -1.862 kcal · mol^{-1}, again assuming a dielectric constant of 4. The average calculated hydrogen bond energy for this system is, therefore, close to the anticipated value of -2 kcal · mol^{-1} mentioned previously, and, when combined with the estimated desolvation energy component of -5.7 kcal · mol^{-1} obtained (as described in section a) from the solvent-accessible surface area of camphor, this gives a total binding energy of -7.7 kcal · mol^{-1}, which is in full agreement with the experimental value [50].

(iv) Loss in Translational and Rotational Energy. It is accepted that there will be a reduction in degrees of freedom experienced by the substrate molecule when it becomes enzyme-bound. Difficulties arise, however, in estimating the magnitude of this effect because there is likely to be some degree of residual motion, both translational and rotational, for the substrate following enzyme binding. Most evaluations of the loss in translational/rotational freedom that occurs when a small molecule binds to a protein assume that there is virtually no apparent residual motion of the substrate or ligand molecule [17, 51]. Essentially, one can consider that the loss of translation and rotation on binding is analogous to the situation experienced by a molecule undergoing a change of state, i.e. gas → liquid or liquid → solid, as these involve a reduction in molecular motion. According to statistical thermodynamics, the mean translational energy obtained from evaluation of the relevant partition function [52] is $^3/_2$ k_BT and, by a similar argument, the mean rotational energy is also given by $^3/_2$ k_BT, such that the overall energy, ΔG_{T+R}, per mole is:

$$\Delta G_{T+R} = 3RT$$

where R is the gas constant ($R = N_A k_B$) and T is the absolute temperature.

Assuming that a compound in dilute solution can be approximated by the situation in the gaseous phase, then the entropy loss experienced due to its binding with a biological macromolecule (such as an enzyme or protein) could be roughly equivalent to $3RT_B$, where T_B is the boiling point. In practice, this appears to be an overestimate, and a more realistic expression for ΔG_{T+R} would be as follows:

$$\Delta G_{T+R} = 3RT_m$$

where T_m is the melting point in degrees absolute.

However, the rationale for using T_B rather than T_m relates to the fact that there will be some residual rotation and, also, a certain degree of translation even for the substrate-bound complex. It should also be recognized that the enzyme itself will have some rotational and translational freedom, and this includes membrane-bound microsomal P450s. Indeed, it would appear that substrate binding lowers the rotational correlation time when benzphetamine binds to $P450_{LM2}$ (CYP2B4) as reported by Finch and Stier [53]. Williams and coworkers [17, 54–57] have demonstrated that ΔG_{T+R} exhibits a linear correlation with the relative molecular mass, M_r, of the ligand molecule, provided that a logarithmic scale is employed. Consequently, it is expected that plotting the values of RT_B for various compounds against their log M_r should give a straight line. In fact, for a homologous series of 30 straight-chain alkanes (methane to triacontane), there is an excellent correlation ($r = 0.989$) between RT_B and log M_r, although there is a deviation from linearity at low molecular mass (equation 1, Table 5). In fact, the data are better fitted via a quadratic expression in log M_r, which improves the correlation coefficient to 0.999, as shown in Table 5 (equation 2). Based on the values presented in Table 5, the high correlation with log M_r provides a means of estimating ΔG_{T+R} for other molecules. Despite the assumptions employed in this method, there is, nevertheless, a good agreement with experimental findings. For example, in the case of benzphetamine binding to CYP2B4, the experimental estimate of 1.4 kcal · mol^{-1} is satisfactorily reproduced by a calculated value of 1.2 kcal · mol^{-1} for ΔG_{T+R} using the relationship presented in Table 5, equation 2.

(v) Bond Rotational Term. In addition to the loss in translation and rotation energy, the binding of a substrate will also give rise to a lowering in energy due to restriction of the freedom of rotatable bonds in the molecule. The problem with estimating the extent of this energy difference is due to the fact that some bonds may be relatively free to rotate when enzyme-bound, whereas others will not. However, a value of 0.6 kcal · mol^{-1} per rotatable bond has been suggested [51] as a measure of the ΔG_{rotors} term, whereas Williams and coworkers [17] have reported values of about 1.2 kcal · mol^{-1} for this quantity, and Bohm [36] has suggested that 0.3 kcal · mol^{-1} accords well with the rotatable bond energy contribution. As this latter study was

TABLE 5. Melting points and boiling points of alkanes

	Alkane	M_r (Da)	log M_r	T_B (K)	RT_B (kcal · mol^{-1})	T_M (K)	RT_M (kcal · mol^{-1})
1.	CH_4	16.04	1.2052	109.0	0.2166	91.0*	0.1808
2.	C_2H_6	30.07	1.4781	184.4	0.3664	89.7	0.1783
3.	C_3H_8	44.10	1.6444	230.9	0.4588	83.3*	0.1655
4.	C_4H_{10}	58.12	1.7643	272.5	0.5415	134.6	0.2675
5.	C_5H_{12}	72.15	1.8582	309.1	0.6142	143.0	0.2842
6.	C_6H_{14}	86.18	1.9354	342.0	0.6796	178.0	0.3537
7.	C_7H_{16}	100.20	2.0009	371.4	0.7380	182.4	0.3625
8.	C_8H_{18}	114.23	2.0578	398.7	0.7923	216.2	0.4296
9.	C_9H_{20}	128.26	2.1081	423.8	0.8422	222.0	0.4412
10.	$C_{10}H_{22}$	142.28	2.1531	447.1	0.8885	243.3	0.4835
11.	$C_{11}H_{24}$	156.31	2.1940	469.0	0.9320	247.4	0.4916
12.	$C_{12}H_{26}$	170.34	2.2313	489.3	0.9723	263.4	0.5234
13.	$C_{13}H_{28}$	184.37	2.2657	508.4	1.0103	267.5	0.5316
14.	$C_{14}H_{30}$	198.39	2.2975	526.7	1.0467	278.9	0.5542
15.	$C_{15}H_{32}$	212.42	2.3272	543.6	1.0802	283.0	0.5624
16.	$C_{16}H_{34}$	226.45	2.3550	560.0	1.1128	291.2	0.5787
17.	$C_{17}H_{36}$	240.47	2.3811	574.8	1.1422	295.0	0.5862
18.	$C_{18}H_{38}$	254.50	2.4057	589.1	1.1707	301.2	0.5985
19.	$C_{19}H_{40}$	268.53	2.4290	602.7	1.1977	305.1	0.6063
20.	$C_{20}H_{42}$	282.55	2.4511	616.0	1.2241	309.8	0.6156
21.	$C_{21}H_{44}$	296.58	2.4721	629.5	1.2509	313.5	0.6230
22.	$C_{22}H_{46}$	310.61	2.4922	641.6	1.2750	317.4	0.6307
23.	$C_{23}H_{48}$	324.63	2.5114	653.0	1.2976	320.6	0.6371
24.	$C_{24}H_{50}$	338.66	2.5298	664.3	1.3201	327.0	0.6498
25.	$C_{25}H_{52}$	352.69	2.5474	674.9	1.3412	328.0	0.6518
26.	$C_{26}H_{54}$	366.71	2.5643	685.2	1.3616	329.4	0.6546
27.	$C_{27}H_{56}$	380.74	2.5806	695.0	1.3811	332.5	0.6607
28.	$C_{28}H_{58}$	394.77	2.5963	704.6	1.4002	337.5	0.6707
29.	$C_{29}H_{60}$	408.80	2.6115	713.8	1.4185	336.7	0.6691
30.	$C_{30}H_{62}$	422.82	2.6262	722.7	1.4361	338.8	0.6733

	Regression equations	n	s	R	F
1.	$RT_B = 0.932 \log M_r - 1.067$ (± 0.026)	30	0.0499	0.989	1312.4
2.	$RT_B = 0.274 \log M_r^2 - 0.173 \log M_r$ (± 0.003) (± 0.006)	30	0.0091	0.999	13479.8
3.	$2RT_M = 0.948 \log M_r - 1.105$ (± 0.022)	28	0.0328	0.993	1871.8

Reference to data: [39].
*Compounds 1 and 3 were omitted from equation 3.

conducted on 45 protein–ligand complexes and involved a multivariate analysis of various components for the binding energy, one can have a fair measure of confidence in the results obtained. However, because the ΔG_{T+R} term was not explicitly calculated in this particular exercise, it is possible that a somewhat higher value for ΔG_{rotors} (i.e. 0.6 kcal·mol^{-1}) may be more widely applicable, particularly for enzyme–substrate complexes.

(vi) Aromatic Ring Stacking Term. There is a particular energy associated with the parallel stacking of aromatic rings which arises from an electrostatic attraction between the delocalized π-system of one ring and the hydrogen atoms on the other. Consequently, this attractive energy term is sometimes referred to as the π-π stacking interaction energy. Additionally, there is the possibility of a charge-transfer interaction between two aromatic rings (depending on the substituents), which can affect the electron donor–acceptor properties of the two ring systems, and the stacking of DNA base-pairs is an example of this type of interaction, although there is also a strong dipole–dipole component to the overall energy [49]. The energy of a typical π-π stacking interaction is around −1 kcal · mol^{-1}, although there is a range of −0.5 to −5.5 kcal · mol^{-1}, and the strength of the π-π stacking term appears to be related to the area of overlap between the aromatic ring systems involved in the interaction [35]. For example, 6-membered rings such as benzene and pyridine exhibit a π-π stacking energy of −1 kcal·mol^{-1}, whereas 5-membered rings such as imidazole and pyrrole show a lowered value of −1.0 kcal · mol^{-1} [35]. A comparison of the ring areas of benzene (54.8 Å^2) and imidazole (39.6 Å^2) indicates that these are indeed approximately in proportion to their π-π stacking terms, thus leading to a straightfor-

ward method of estimating this component to the overall binding energy, as areas of overlap are readily calculated from enzyme–substrate complexes. Furthermore, the value of -1.0 kcal $\cdot$ mol^{-1} as typical for π-π stacking energies accords well with evidence from site-directed mutagenesis studies. For the interaction between coumarin and CYP2A5, for example, a change of phenylalanine to leucine at position 209 gives rise to a 0.9 kcal $\cdot$ mol^{-1} reduction in the binding affinity [58], which is in good agreement with the above value.

(vii) Overall Expression of Binding Energy. Several groups of workers have reported equations that can be successfully applied to real situations where ligand–protein or enzyme–substrate binding affinities are known [17, 36, 51, 54, 55, 59–64]. Essentially, the main differences between these various approaches center around the number of terms considered as being important to the binding energy and, in some cases, the methods employed in their estimation. The study by Bohm [36] is particularly interesting in view of the relatively large number of protein–ligand complexes considered, and the robust nature of the analyses performed. Moreover, good correlations ($r = 0.87$) were obtained between different combinations of the main terms involved and actual experimental values for ligand-complex binding energies [36], together with satisfactory comparisons between predicted and observed binding energies for a number of systems, including P450 complexes. The overall expression employed in this work contained essentially four terms, governing contributions from hydrogen bond, ionic, lipophilic, and bond rotation energies where the best correlations were obtained for the following equation:

$$\Delta G_{bind} = -4.7\Delta G_{hb} - 8.3\Delta G_{ionic} - 0.17\Delta G_{lipo} + 1.4\Delta G_{rot} + 5.4$$

where ΔG_{hb}, ΔG_{ionic}, ΔG_{lipo}, and ΔG_{rot} refer to the hydrogen bond, ionic, lipophilic, and bond rotation terms, respectively, which are expressed in kJ $\cdot$ mol^{-1} [36].

This equation exhibited a correlation of 0.873 for a data set comprising all 45 protein–ligand complexes considered and, although the constant term was ascribed to represent the average contribution from the loss in translational and rotational freedom, setting this term to zero also resulted in a good correlation ($r = 0.871$) with the experimental data [36]. Nevertheless, expressions of the type outlined above clearly provide a satisfactory explanation of binding affinity and may be employed predictively, although it may be necessary to make a number of minor modifications to the various terms when applied to particular enzyme–substrate interactions such as for the microsomal P450s.

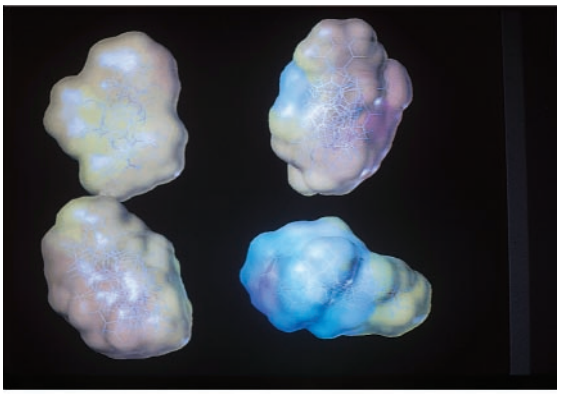
(3) Substrate Templates and Molecular Electrostatic Potential Surfaces

For the 48 chemicals listed in Table 2, the various substrates of each P450 isoform have been superimposed on the basis of their interaction within the putative active sites of the relevant enzymes concerned. Furthermore, colour-coding the solvent-accessible surface by electrostatic potential energy provides an indication of the likely interaction between substrate and enzyme, as has been demonstrated for omeprazole [15]. Performing this calculation on the entire template of six superimposed substrates for each P450 isoform tends to show a reinforcement of the hydrogen bond donor/acceptor sites and, additionally, gives an indication of the general region where oxidative metabolism may occur, as this usually coincides with a positive area of electrostatic potential energy.

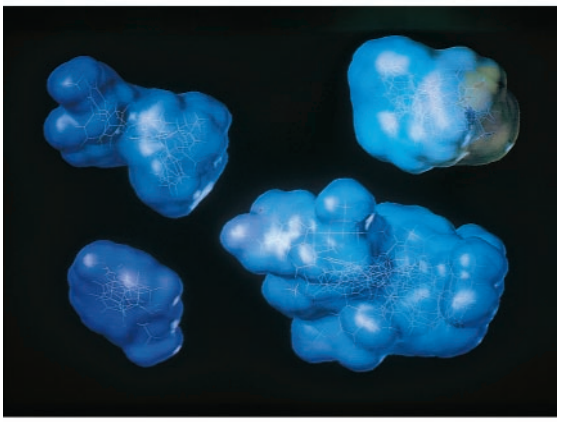
Consequently, production of electrostatic potential surfaces for all eight substrate templates enables differentiation of certain key features likely to govern enzyme specificity and location of the preferred metabolic site. Figure 2 shows molecular electrostatic potential energy surfaces for substrates of (a) CYP1A2, (b) CYP2A6, (c) CYP2B6, (d) CYP2C9, (e) CYP2C19, (f) CYP2D6, (g) CYP2E1, and (h) CYP3A4; these also provide a means of visualizing the shape characteristics of each structural template. The size and shape characteristics of each set of P450 substrates, together with the disposition of positive and negative electrostatic potential energy centers, probably reflect the essential topographical features of the particular P450 isoform concerned, and there is a general agreement between the ways in which these substrate templates match the corresponding active site region of each human P450 isoform studied to date. This situation is certainly the case in structurally characterized P450s, such as $P450_{cam}$ (CYP102) for example, where the substrate, camphor, displays a large negative potential energy associated with its hydrogen bond contact with an active site tyrosine (Tyr-96) and a high positive potential energy centered around the site of metabolism [16]. It is, therefore, reasonable to assume that similar combinations of molecular shape and electronic structural properties govern the recognition of potential P450 substrates for the enzymes that tend to metabolize them most effectively.

CONCLUSIONS

Clearly, there are structural determinants of P450 selectivity which are reflected in substrate molecular geometries. Table 6 summarizes the general characteristics of human hepatic microsomal P450 substrates based on the information presented in Table 1, and from other studies reported in the literature. Together with the scheme proposed in Fig. 1 as a decision tree for evaluating P450 specificity, it is probable that a fairly satisfactory differentiation can be achieved for most chemicals. However, modelling the active site interactions for individual compounds should



A



B

FIG. 2. (A) Molecular surfaces of substrate templates for CYP1A2 (top left), CYP2A6 (top right), CYP2B6 (bottom left), and CYP2C9 (bottom right) colour-coded according to electrostatic isopotential (EIP) energy, where positive EIP values are shown in red and negative EIP values are in blue, with intermediate colours representing EIP values between the two extremes. (B) Molecular surfaces of substrate templates for CYP2C19 (top left), CYP2D6 (top right), CYP2E1 (bottom left), and CYP3A4 (bottom right) colour-coded according to EIP energy values using the same colour scheme as for panel A.

enable more precise definition of a likely metabolic route, and consideration of a relatively small number of readily calculated structural properties (formulated in Results and Discussion) would facilitate estimations of both binding affinity and rate of metabolism, as described in part 2 of this section. Some degree of caution should be maintained in attempting to predict P450 isoform specificity directly from the structure of the chemical concerned when there is significant conformational flexibility in the molecule. Under such circumstances, the lowest energy conformer may not necessarily represent the actual molecular geometry of the enzyme-bound substrate, because the topography of the active site and disposition of key amino acid residues may constrain the molecule significantly. This means that active site modelling can be recommended as the next stage following calculation of substrate properties and, if necessary, the characteristics of the bound molecule should also be evaluated. As a general rule, it is found that the more flexible structures tend to act as substrates for more than one P450 and, in some cases, there may be several sites of metabolism even with the same isoform. Nevertheless, via a combination of techniques including spectroscopic, computational, biochemical, and crystallographic procedures, it should be feasible to predict the P450-mediated metabolic fate of xenobiotics in humans.

TABLE 6. Characteristics of human P450 substrates

CYP	General properties displayed by most substrates
1A2	Planar molecules, moderately basic, medium volume and low ΔE values.
2A6	Non-planar molecules, medium volume with two hydrogen bond acceptors at about 2.5 Å apart and 5–7 Å from the site of metabolism.
2B6	Non-planar molecules, neutral or weakly basic, fairly lipophilic with one or two hydrogen bond acceptors.
2C9	Weakly acid, fairly lipophilic with one or two hydrogen bond donor/acceptors at 5–8 Å from the site of metabolism.
2C19	Neutral or weakly basic, moderately lipophilic with two or three hydrogen bond donor/acceptors at about 4.5 Å apart and 5–8 Å from the site of metabolism.
2D6	Basic, relatively hydrophilic, usually contain an aromatic ring and a hydrogen bond donor/acceptor, basic nitrogen at 5–7 Å from the site of metabolism.
2E1	Low volume, neutral, hydrophilic, relatively planar, structurally diverse with one or two hydrogen bond donor/acceptors at 4–6 Å from the site of metabolism.
3A4	High volume, relatively lipophilic, structurally diverse with one or two hydrogen bond donor/acceptors at 5.5–7.5 Å and 8–10 Å from the site of metabolism.

The information represents a summarization of several studies reported by Smith [32]; Smith and Jones [33]; Smith *et al.* [6, 7]; Lewis *et al.* [11–14, 18, 19, 76]; Lewis and Lake [9, 10]. It is important to recognize that the above characteristics relate to those of most substrates of the P450 isoforms concerned, and some exceptions are known. Also, some compounds can possess properties that enable them to fit more than one P450 and, consequently, these may be substrates of several isoforms.

The financial support of GlaxoWellcome Research & Development Limited, Merck, Sharp & Dohme Limited, the European Union Biomed 2 programme, and the University of Surrey Foundation Fund is gratefully acknowledged. I would like to thank Maurice Dickins and Anne Hersey (GlaxoWellcome) for supplying some of the physicochemical data used in this study.

References

1. Nelson DR, Koymans L, Kamataki T, Stegeman JJ, Feyereisen R, Waxman DJ, Waterman MR, Gotoh O, Coon MJ, Estabrook RW, Gunsalus IC and Nebert DW, P450 superfamily: Update on new sequences, gene mapping, accession numbers and nomenclature. *Pharmacogenetics* **6:** 1–42, 1996.
2. Porter TD and Coon MJ, Cytochrome P450: Multiplicity of isoforms, substrates and catalytic and regulatory mechanisms. *J Biol Chem* **266:** 13469–13472, 1991.
3. Estabrook RW, Impact of the human genome project on drug metabolism and chemical toxicity. *British Toxicology Society Meeting*, Guildford, April 19–22, 1998.

4. Kawajiri K and Hayashi S-I, The CYP1 family. In: *Cytochromes P450: Metabolic and Toxicological Aspects* (Ed. Ioannides C), Chap. 4, pp. 77–97. CRC Press, Boca Raton, 1996.
5. Ronis MJJ, Lindros KO and Ingelman-Sundberg M, The CYP2E subfamily. In: *Cytochromes P450: Metabolic and Toxicological Aspects* (Ed. Ioannides C), Chap. 9, pp. 211–239. CRC Press, Boca Raton, 1996.
6. Smith DA, Ackland MJ and Jones BC, Properties of cytochrome P450 isoenzymes and their substrates. Part 1: Active site characteristics. *Drug Disc Today* **2:** 406–414, 1997.
7. Smith DA, Ackland MJ and Jones BC, Properties of cytochrome P450 isoenzymes and their substrates. Part 2: Properties of cytochrome P450 substrates. *Drug Disc Today* **2:** 479–486, 1997.
8. Lewis DFV and Lake BG, Molecular modelling of members of the P4502A subfamily: Application to studies of enzyme specificity. *Xenobiotica* **25:** 585–598, 1995.
9. Lewis DFV and Lake BG, Molecular modelling of CYP1A subfamily members based on an alignment with CYP102: Rationalization of CYP1A substrate specificity in terms of active site amino acid residues. *Xenobiotica* **26:** 723–753, 1996.
10. Lewis DFV and Lake BG, Molecular modelling of mammalian CYP2B isoforms and their interaction with substrates, inhibitors and redox components. *Xenobiotica* **27:** 443–487, 1997.
11. Lewis DFV, Eddershaw PJ, Goldfarb PS and Tarbit MH, Molecular modelling of CYP3A4 from an alignment with CYP102: Identification of key interactions between putative active site residues and CYP3A-specific chemicals. *Xenobiotica* **26:** 1067–1086, 1996.
12. Lewis DFV, Eddershaw PJ, Goldfarb PS and Tarbit MH, Molecular modelling of cytochrome P450 2D6 (CYP2D6) based on an alignment with CYP102: Structural studies on specific CYP2D6 substrate metabolism. *Xenobiotica* **27:** 319–340, 1997.
13. Lewis DFV, Bird MG and Parke DV, Molecular modelling of CYP2E1 enzymes from rat, mouse and man: An explanation for species differences in butadiene metabolism and potential carcinogenicity and rationalization of CYP2E substrate specificity. *Toxicology* **118:** 93–113, 1997.
14. Lewis DFV, Dickins M, Eddershaw PJ, Tarbit MH and Goldfarb PS, Molecular modelling of human CYP2C subfamily enzymes CYP2C9 and CYP2C19: Rationalization of substrate specificity and site-directed mutagenesis experiments in the CYP2C subfamily. *Xenobiotica*: **28:** 235–268, 1998.
15. Lewis DFV and Lake BG, Molecular modelling of omeprazole interactions with cytochrome P450 isozymes is consistent with metabolism in human liver microsomes. *Toxicology* **125:** 31–44, 1998.
16. Lewis DFV, Models to predict drug metabolism. *Manuf Chem* **69:** 15–19, 1998.
17. Williams DH, Cox JPL, Doig AJ, Gardner M, Gerhard U, Kaye PT, Lal AR, Nicholls IA, Salter CJ and Mitchell RC, Toward the semiquantitative estimation of binding constants. Guides for peptide-peptide binding in aqueous solution. *J Am Chem Soc* **113:** 7020–7030, 1991.
18. Lewis DFV, Eddershaw PJ, Goldfarb PS and Tarbit MH, Structural determinants of P450 substrate specificity, binding affinity and catalytic rate. *Chem Biol Interact* **115:** 175–199, 1998.
19. Lewis DFV, Dickins M, Eddershaw PJ, Tarbit MH and Goldfarb PS, Cytochrome P450 substrate specificities, substrate structural templates and enzyme active site geometries. *Drug Metabol Drug Interact* **15:** 1–49, 1999.
20. Lewis DFV and Pratt JM, The P450 catalytic cycle and oxygenation mechanism. *Drug Metab Rev* **30:** 739–786, 1998.
21. Rendic S and Di Carlo FJ, Human cytochrome P450 enzymes: A status report summarizing their reactions, substrates, inducers, and inhibitors. *Drug Metab Rev* **29:** 413–580, 1997.
22. Parkinson A, An overview of current cytochrome P450 technology for assessing the safety and efficacy of new materials. *Toxicol Pathol* **24:** 45–57, 1996.
23. Gonzalez FJ and Gelboin HV, Role of human cytochromes P450 in the metabolic activation of chemical carcinogens and toxins. *Drug Metab Rev* **26:** 165–183, 1994.
24. Sangster J, Octanol-water partition coefficients of simple organic compounds. *J Phys Chem Ref Data* **18:** 1111–1229, 1989.
25. Hansch C and Leo AJ, *Substituent Constants for Correlation Analysis in Chemistry and Biology*. Wiley, New York, 1979.
26. Suzuki T and Kudo Y, Automatic log P estimation based on combined additive modelling methods. *J Comput Aided Mol Des* **4:** 155–198, 1990.
27. Rekker RF, ter Laak AM and Mannhold R, On the reliability of calculated log P values: Rekker, Hansch/Leo and Suzuki approach. *Quant Struct Act Relat* **12:** 152–157, 1993.
28. Hansch C, Bjorkroth JP and Leo A, Hydrophobicity and central nervous system agents: On the principle of minimal hydrophobicity in drug design. *J Pharm Sci* **76:** 663–687, 1987.
29. 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.
30. Li H and Poulos TL, The structure of the cytochrome $P450_{bm\text{-}3}$ haem domain complexed with the fatty acid substrate, palmitoleic acid. *Nat Struct Biol* **4:** 140–146, 1997.
31. Smith DA, Physicochemical properties in drug metabolism and pharmacokinetics. In: *Computer-Assisted Lead Finding and Optimization* (Eds. van de Waterbeemd H, Testa B and Folkers G), Chap. 17, pp. 265–276. Wiley, New York, 1997.
32. Smith DA, Species differences in metabolism and pharmacokinetics: Are we close to an understanding? *Drug Metab Rev* **23:** 355–373, 1991.
33. Smith DA and Jones BC, Speculations on the substrate structure-activity relationship (SSAR) of cytochrome P450 enzymes. *Biochem Pharmacol* **44:** 2089–2098, 1992.
34. Lewis DFV, The CYP2 family: Models, mutants and interactions. *Xenobiotica* **28:** 617–661, 1998.
35. Simon Z, Motoc I and Chiriac A, Molecular basis of receptor-ligand interactions. In: *Specific Interactions and Biological Recognition Processes* (Eds. Voiculetz N, Motoc I and Simon Z), Chap. 2, pp. 17–79. CRC Press, Boca Raton, 1993.
36. Bohm H-J, The development of a simple empirical scoring function to estimate the binding constant for a protein-ligand complex of known three-dimensional structure. *J Comput Aided Mol Des* **8:** 243–256, 1994.
37. Sharp KA, Nicholls A, Fine RF and Honig B, Reconciling the magnitude of the microscopic and macroscopic hydrophobic effects. *Science* **252:** 106–109, 1991.
38. Beveridge DL, Kelly MM and Radna RJ, A theoretical study of solvent effects on the conformational stability of acetylcholine. *J Am Chem Soc* **96:** 3769–3778, 1974.
39. Weast RC (Ed.), *CRC Handbook of Chemistry and Physics*. CRC Press, Boca Raton, 1991.
40. James AM and Lord MP (Eds.), *MacMillan's Chemical and Physical Data*. MacMillan, London, 1992.
41. Andrews PR and Tintelnot M, Intermolecular forces and molecular binding. In: *Comprehensive Medicinal Chemistry* (Eds. Hansch C, Sammes PG and Taylor JB), Vol. 4: Quantitative Drug Design, pp. 321–347. Pergamon, Oxford, 1990.
42. Holtje H-D and Kemmritz K, Molecular modelling of the heme environment of human prostaglandin endoperoxide synthase 1: Comparison with the crystal structure. In: *QSAR*

and Molecular Modelling (Eds. Sanz F, Giraldo J and Manaut F), pp. 549–555. Prous, Barcelona, 1995.
43. Christensen IT and Jorgensen FS, Molecular mechanics calculations of proteins. Comparison of different energy minimization strategies. *J Biomol Struct Dyn* **15:** 473–488, 1997.
44. Delaage M, Physico-chemical aspects of molecular recognition. In: *Molecular Recognition Mechanisms* (Ed. Delaage M), Chap. 1, pp. 1–13. VCH, New York, 1991.
45. Ellis SW, Hayhurst GP, Smith G, Lightfoot T, Wang MMS, Simula AP, Ackland MJ, Sternberg MJE, Lennard MS, Tucker GT and Wolf CR, Evidence that aspartic acid 301 is a critical substrate-contact residue in the active site of cytochrome P450 2D6. *J Biol Chem* **270:** 29055–29058, 1995.
46. Seibel GL and Kollman PA, Molecular mechanics and the modelling of drug structures. In: *Comprehensive Medicinal Chemistry* (Eds. Hansch C, Sammes PG and Taylor JB), Vol. 4: Quantitative Drug Design, pp. 125–138. Pergamon, Oxford, 1990.
47. Rigby M, Smith EB, Wakeham WA and Maitland GC, *The Forces between Molecules*. Clarendon Press, Oxford, 1986.
48. Poulos TL, Finzel BC and Howard AJ, High-resolution crystal structure of cytochrome $P450_{cam}$. *J Mol Biol* **195:** 687–700, 1987.
49. Simon Z, *Quantum Biochemistry and Specific Interactions*. Abacus, Tunbridge Wells, Kent, 1976.
50. Griffin BW and Peterson JA, Camphor binding by *Pseudomonas putida* cytochrome P-450. Kinetics and thermodynamics of the reaction. *Biochemistry* **11:** 4740–4746, 1972.
51. Andrews PR, Craig DJ and Martin JL, Functional group contributions to drug-receptor interactions. *J Med Chem* **27:** 1648–1657, 1984.
52. Atkins PW, *Physical Chemistry*. Oxford University Press, Oxford, 1990.
53. Finch SAE and Stier A, Rotational diffusion of homo- and hetero-oligomers of cytochrome P450: The functional significance of cooperativity and the membrane structure. *Front Biotransform* **5:** 34–71, 1991.
54. Williams DH, Searle MS, MacKay JP, Gerhard U and Maplestone RA, Toward an estimation of vancomycin group antibiotics. *Proc Natl Acad Sci USA* **90:** 1172–1178, 1993.
55. Williams DH, The molecular basis of biological order. *Aldrichim Acta* **24:** 71–80, 1991.
56. Searle MS and Williams DH, The cost of conformational order: Entropy changes in molecular associations. *J Am Chem Soc* **114:** 10690–10697, 1992.
57. Searle MS, Williams DH and Gerhard U, Partitioning of free energy contributions in the estimation of binding constants: Residual motions and consequences for amide–amide hydrogen bond strengths. *J Am Chem Soc* **114:** 10697–10704, 1992.
58. Juvonen RO, Iwasaki M and Negishi M, Structural function of residue-209 in coumarin 7-hydroxylase ($P450_{coh}$). *J Biol Chem* **266:** 16431–16435, 1991.
59. Kollman PA and Merz KM, Computer modelling of the interactions of complex molecules. *Acc Chem Res* **23:** 246–252, 1990.
60. Grootenhuis PDJ and Kollman PA, Crown ether-neutral molecule interactions studied by molecular mechanics, normal mode analysis, and free energy perturbation calculations. Near quantitative agreement between theory and experimental binding energies. *J Am Chem Soc* **111:** 4046–4051, 1989.
61. Bash PA, Singh UC, Langridge R and Kollman PA, Free energy calculations by computer simulation. *Science* **236:** 564–568, 1987.
62. Helms V, Deprez E, Gill E, Barret C, Hui Bon Hoa G and Wade RC, Improved binding of cytochrome $P450_{cam}$ substrate analogues designed to fill extra space in the substrate binding pocket. *Biochemistry* **35:** 1485–1499, 1996.
63. Perakyla M and Pakkanen TA, Quantum mechanical model assembly calculations of energetics of binding of ligands to protein receptors. *Biochem Soc Trans* **24:** 143S, 1996.
64. Viswanadhan VN, Reddy MR, Wlodawer A, Varney MD and Weinstein JN, An approach to rapid estimation of relative binding affinities of enzyme inhibitors: Application to peptidomimetic inhibitors of the human immunodeficiency virus type 1 protease. *J Med Chem* **39:** 705–712, 1996.
65. Gibson GG and Skett P, *Introduction to Drug Metabolism*. Chapman & Hall, London, 1994.
66. Lewis DFV, *Cytochromes P450: Structure, Function and Mechanism*. Taylor & Francis, London, 1996.
67. Whitlock JP and Denison MS, Induction of cytochrome P450 enzymes that metabolize xenobiotics. In: *Cytochrome P450* (Ed. Ortiz de Montellano PR), Chap. 10, pp. 367–390. Plenum Press, New York, 1995.
68. Kliewer SA, Moore JT, Wade L, Staudinger JL, Watson MA, Jones SA, McKee DD, Oliver BB, Willson TM, Zetterstrom RH, Perlmann T and Lehmann JM, An orphan nuclear receptor activated by pregnanes defines a novel steroid signaling pathway. *Cell* **92:** 73–82, 1998.
69. Csizmadia F, Tsantili-Kakoulidou A, Panderi I and Darvas F, Prediction of distribution coefficient from structure. 1. Estimation method. *J Pharm Sci* **86:** 865–871, 1997.
70. Desai MC, Thadeio PF, Lipinski CA, Liston DR, Spencer RW and Williams IH, Physical parameters for brain uptake: Optimizing log P, log D and p*Ka* of THA. *Bioorg Med Chem Lett* **1:** 411–414, 1991.
71. Elguero J, Gonzalez E and Jacquier R, Recherches dans la serie des azoles. XXXV. Relations basicite-structure dans la serie du pyrazole. *Bull Soc Chim Fr* 5009–5017, 1968.
72. Li N-Y, Li T and Gorrod JW, Determination of partition coefficients and ionization constants of (S)(-)-nicotine and certain metabolites. *Med Sci Res* **20:** 901–902, 1992.
73. Newton DW and Kluza RB, pK_a values of medicinal compounds in pharmacy practice. *Drug Intell Clin Pharm* **12:** 546–554, 1978.
74. Schoenwald RD and Huang HS, Corneal penetration behaviour of β-blocking agents. 1. Physico-chemical factors. *J Pharm Sci* **72:** 1266–1272, 1983.
75. Tsantili-Kakoulidou A, Panderi I, Csizmadia F and Darvas F, Prediction of distribution coefficient from structure. 2. Validation of Prolog D, an expert system. *J Pharm Sci* **86:** 1173–1179, 1997.
76. Lewis DFV, Ioannides C and Parke DV, An improved and updated version of the COMPACT procedure for the evaluation of P450-mediated chemical activation. *Drug Metab Rev* **30:** 709–737, 1998.