

Quantitative structure–pharmacokinetic relationship modelling: apparent volume of distribution

Taravat Ghafourian, Mohammad Barzegar-Jalali, Nasim Hakimiha and Mark T. D. Cronin

Abstract

The purpose of this study was to develop a quantitative structure–activity relationship (QSAR) for the prediction of the apparent volume of distribution (Vd) in man for a heterogeneous series of drugs. The relationship of many computed, and some experimental, structural descriptors with Vd, and the Vd corrected for protein binding (unbound Vd), was investigated. Models were constructed using stepwise regression analysis for all the 70 drugs in the dataset, as well as for acidic drugs and basic drugs separately. The predictive power of the models was assessed using half the chemicals as a test set, and revealed that the models for Vd yielded lower prediction errors than those constructed for the unbound Vd (mean fold error of 2.01 for Vd compared with 2.28 for unbound Vd). Moreover, the separation of the compounds into acids and bases did not reduce the prediction error significantly.

Introduction

Modern drug design focuses not only on the pharmacological activity of a compound but also considers a range of other properties including its pharmacokinetic behaviour. A successful drug candidate should demonstrate, among other characteristics, the ability to be absorbed and to reach its site of action and should have a suitable half-life. In recent years, there has been an enormous interest in the prediction of human pharmacokinetic properties using different methods ranging from computational approaches to using in-vitro and in-vivo data. This is due to the fact that a large proportion of drugs fail in development due to poor absorption, distribution, metabolism and elimination (ADME) properties (Kennedy 1997). The aim of these studies is to provide screening tools for drugs at a very early stage of development. Animal pharmacokinetic studies are a routine tool to predict drug behaviour in man. Furthermore, human-derived cellular or subcellular systems have been developed to measure permeation, membrane transport, absorption, distribution, metabolism and excretion; the systems as summarised by Balant & Gex-Fabry (2001) include partition coefficients, Caco-2 cell monolayers, plasma protein binding, microsomes, hepatocytes and enzyme systems. QSAR methods provide models not only for prediction of pharmacological activity but also for toxicological end-points, membrane passage and pharmacokinetics parameters. Prediction of oral absorption has been a hot research topic in recent years, with Lipinski's rule of five (Lipinski et al 1997) at an early stage leading to the application of both linear and non-linear QSAR methods to the prediction of oral absorption (for more recent works see Raevsky et al (2002), Palm et al (1997), Ghafourian & Barzegar-Jalali (2002), Van de Waterbeemd (2002), Klopman et al (2002), Zhao et al (2002)). A detailed review of the methods used to predict pharmacokinetic properties has been published by Egan & Lauri (2002). Other routes of absorption, including skin (Ghafourian & Fooladi 2000; Moss et al 2002), as well as permeability through cell lines (Tantishaiyakul 2001; Kulkarni et al 2002), artificial membranes (Agatonovic-Kustrin et al 2001; Kansy et al 2001) and the blood–brain barrier (Abraham et al 1999), have also been explored.

Volume of distribution, clearance and elimination half-life are three of the most important pharmacokinetic properties. Various quantitative structure–pharmacokinetic

Tabriz University of Medical Sciences, Tabriz 51664, Iran

Taravat Ghafourian, Mohammad Barzegar-Jalali, Nasim Hakimiha

School of Pharmacy and Chemistry, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK

Taravat Ghafourian, Mark T. D. Cronin

Correspondence: T. Ghafourian, School of Pharmacy and Chemistry, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK.
E-mail: t.ghafourian@livjm.ac.uk

Acknowledgement: This work was supported by the research council of Tabriz University of Medical Sciences.

relationships (QSPRs) have been developed for several congeneric series of molecules and different mathematical models have been proposed (Gobburu & Shelver 1995; Van der Graaf et al 1999; Turner et al 2003). Models have been proposed for non-aromatics, aromatics and hetero-aromatics to allocate the chemicals into three ratings of volume of distribution (Hirono et al 1994) and neural network (Ritschel et al 1995) and multivariate (Karalis et al 2002) methods have been employed to model volume of distribution. This study has focused on the distribution process of drugs. The volume of distribution of the central compartment (V_C) is used to correlate plasma concentration of a drug at time zero (C_0) to the amount of drug in the body (X) (Shargel & Yu 1999) by the expression:

$$X = V_C * C_0 \quad (1)$$

Two different terms have been used to describe the volume of distribution for drugs that follow multiple exponential decay. The first, designated V_{area} , is calculated as the ratio of clearance to the rate of decline of concentration during the elimination (final) phase of the logarithmic concentration-versus-time curve:

$$V_{area} = \text{Dose}/k.AUC \quad (2)$$

The second volume term is the volume of distribution at steady state (V_{ss}) which represents the volume in which a drug would appear to be distributed during steady state if the drug existed throughout that volume at the same concentration as that in the measured fluid (plasma or blood). When using pharmacokinetics to make drug dosing decisions, the difference between V_{area} and V_{ss} is not usually clinically significant (Wilkinson 2001). In this investigation, QSAR methods have been employed to predict the apparent volume of distribution (V_d) of drugs in man. The compounds used in this study were drug entities with unrelated chemical structures. Predictions have been made based on different QSAR models for acidic drugs and basic drugs separately, as well as for all the drugs together. The V_d parameter was used either untransformed or corrected for plasma protein binding. A comparison with previous QSAR, and other prediction methods, is made.

Materials and Methods

Pharmacokinetic data

Data for the volume of distribution at steady state (V_d) and the plasma protein binding (ppb) of 70 drugs, belonging to different chemical groups, were collected from the literature (Raaflaub & Speiser-Courvoisier 1974; Ritschel & Hammer 1980; Greenblatt 1981; Moffat et al 1986; Sonne et al 1988; Durnas et al 1990; Schoerlin et al 1990; Glare & Walsh 1991; Fulton & Sorkin 1995; Ritschel et al 1995; Lam et al 1997; Potter & Hollister 2001; Thummel & Shen 2001; Perry 2002). Where several V_d values were

available for a compound, the mean was used in the analyses. The volume of distribution of free drug, $V_{du} = V_d/(1 - \text{ppb})$, was also used for the development of predictive models. The compounds and the pharmacokinetic data are listed in Table 1 together with the relevant references.

Physicochemical and structural properties

A total of 75 structural descriptors for these compounds were obtained from various software packages. Table 2 lists the descriptors calculated for the drugs and the software used. The COSMIC force field in the NEMESIS software was used for energy minimisation before molecular mechanical parameter calculations. The software was distributed by Oxford Molecular Ltd (Oxford, UK), although it is no longer available. For calculation of molecular orbital parameters, the three-dimensional structures of the drugs were imported from NEMESIS and minimised using the MNDO Hamiltonian in MOPAC version 7.0 (QCPE, Department of Chemistry, Indiana University, 800 East Kirkwood Ave., Bloomington, IN 47405-7102). SMILES strings were entered into the MOLCONN-Z software. MOLCONN-Z was used to calculate topological descriptors. ACD/Log D Suite release 7.0 was used to calculate, among others, log P and log D at pH 1 and 7.4. The fraction of the un-ionised form of drugs and log D at pH 7.4 were also calculated from experimental pK_a ($pK_{a(\text{Exp})}$) taken from the references (see Table 3) and log P^* values taken from the Biobyte database by following the formulae below. Note that as the aim was calculation of the un-ionised fraction of drugs, only the first acidic pK_a and the first basic pK_a were considered.

For weak bases:

$$fiB = 100/(1 + \text{antilog}(7.4 - pK_{a(\text{Exp})})) \quad (3)$$

$$\log D_{7.4\text{calc}} = \log P^* + \log(1 + \text{antilog}(7.4 - pK_{a(\text{Exp})})) \quad (4)$$

For weak acids:

$$fiA = 100/(1 + \text{antilog}(pK_{a(\text{Exp})} - 7.4)) \quad (5)$$

$$\log D_{7.4\text{calc}} = \log P^* + \log(1 + \text{antilog}(pK_{a(\text{Exp})} - 7.4)) \quad (6)$$

$$fu_{\text{calc}} = 100 - (fiA + fiB) \quad (7)$$

In equations 3–7, fiB , fiA and fu_{calc} are, respectively, percents of cationic, anionic and un-ionised drug at pH 7.4.

Development of QSARs

Stepwise regression analysis was used to determine statistically significant relationships between structural parameters and the volume of distribution. The statistical analyses were performed using the MINITAB (release 13.1) statistical software. To avoid the risk of chance correlations, loss of interpretability and predictability, the number of parameters in the models was kept as low as possible. Accordingly, the

Table 1 Drugs considered and the apparent volume of distribution (Vd) and protein binding (ppb) collected from the literature together with the corresponding partition coefficients, distribution coefficients at pH 1 and pH 7.4, dipole moments and the number of aromatic carbon atoms

Name	Vd (L kg ⁻¹)	Ref.	ppb	Ref.	log P	log D ₁	log D _{7.4}	μ _{SD1}	C _{arom}
Acetanilide	0.161	(b)	*		1.08	1.08	1.08	3.88	6
Alprazolam	0.813	(a, b, k)	0.700	b	2.50	0.99	2.50	6.72	12
Amphetamine	3.773	(b, b, g)	0.275	b	1.81	-1.29	-0.63	1.40	6
Amitriptyline	10.643	(a, b, g)	0.940	b	4.92	1.82	3.10	0.89	12
Amobarbital	1.000	(b, g)	0.550	b	2.05	2.05	1.99	0.47	0
Bromazepam	1.183	(b, d, i)	0.550	b	2.06	1.01	2.06	4.34	12
Bupivacaine	1.000	(b)	0.900	b	3.64	0.53	2.80	4.16	6
Bupropion	13.133	(a, k)	0.850	b	3.47	0.37	3.27	1.33	6
Butorphanol	5.000	(b)	0.900	b	3.77	0.67	3.10	1.10	6
Caffeine	0.533	(b, b, g)	0.350	b	-0.13	-0.28	-0.13	4.21	0
Carbamazepine	1.133	(a, b, b)	0.750	b	2.67	2.66	2.67	3.93	12
Chlordiazepoxide	0.337	(a, b, g)	0.935	b	2.49	1.22	2.49	4.52	12
Chlorphentermine	2.500	(b)	*		2.75	-0.35	0.46	3.00	6
Clobazam	1.120	(d)	0.850	b	2.34	1.70	2.34	5.15	12
Clomipramine	14.667	(b, n)	0.925	b	5.53	2.37	3.50	2.39	12
Clonazepam	3.127	(a, b, g)	0.850	b	2.34	1.70	2.34	3.13	12
Cloral hydrate	0.600	(b)	0.350	b	1.27	1.27	1.27	1.09	0
Clorazepate	0.720	(a, b, k)	0.970	b	3.70	1.80	0.08	6.61	12
Desipramine	33.667	(a, b, n)	0.800	b	4.13	0.62	1.39	1.58	12
Diazepam	1.920	(a, c, g)	0.985	b	2.96	1.11	2.96	4.00	12
Doxepine	10.267	(a, a, b)	0.800	b	3.86	0.76	2.07	0.76	12
Ethchlorvynol	2.500	(b)	0.600	b	2.06	2.06	2.06	0.92	0
Ethosuximide	0.677	(b, g, g)	0.000	b	0.38	0.38	0.38	2.42	0
Etidocaine	2.000	(b)	0.940	b	3.77	0.66	3.30	4.53	6
Fentanyl	3.600	(d)	0.800	b	3.89	0.79	2.23	4.04	12
Flunitrazepam	4.000	(b)	0.780	b	1.25	0.51	1.25	3.00	12
Fluoxetine	41.000	(a, b, k)	0.940	b	4.09	0.99	1.56	2.75	12
Glutethimide	3.077	(b)	0.540	b	2.70	2.70	2.70	3.47	6
Haloperidol	16.007	(a, b, e)	0.900	b	3.01	-0.09	2.11	3.89	12
Ibuprofen	0.100	(b)	0.990	b	3.72	3.72	0.80	4.41	6
Imipramine	17.247	(a, b, e)	0.900	b	4.80	1.29	2.75	1.24	12
Indometacin	0.963	(b)	0.950	b	3.10	3.10	-0.01	3.56	12
Ketamine	4.000	(b)	0.350	b	2.18	-0.92	2.13	3.38	6
Lidocaine	1.443	(b, c, h, o)	0.700	b	2.36	-0.75	1.20	4.69	6
Lorazepam	1.203	(a, b, g)	0.900	b	2.47	2.42	2.47	4.58	12
Maprotiline	22.000	(a, b)	0.900	b	4.51	1.41	1.65	1.20	12
Meprobamate	0.700	(b)	0.200	b	0.70	0.70	0.70	0.11	0
Meptazinol	5.470	(d)	0.270	b	3.70	0.60	1.61	1.35	6
Methadone	3.940	(b, g, k)	0.900	b	4.20	1.10	2.56	1.82	12
Methaqualone	6.000	(b)	0.850	b	2.50	0.84	2.50	3.07	12
Metoclopramide	3.000	(b)	0.650	b	2.22	-0.95	0.05	5.30	6
Midazolam	1.913	(b, e)	0.950	b	3.93	1.13	3.92	5.46	12
Moclobemide	1.290	(e)	0.500	e	0.84	-2.26	0.72	3.88	6
Morphine	3.210	(b, g, h, j)	0.250	b	0.43	-2.67	-0.49	2.23	6
Naloxone	3.000	(b)	0.400	b	1.45	-1.65	1.36	3.37	6
Nitrazepam	2.280	(b, d, g)	0.865	b	2.18	0.39	2.18	3.67	12
Nortriptyline	21.333	(a, b, b)	0.925	b	5.65	2.55	3.11	1.29	12
Oxazepam	0.803	(a, g, l, m)	0.950	b	2.31	1.57	2.31	3.38	12
Oxyphenbutazone	0.107	(b, b, g)	0.990	b	2.72	2.72	1.98	1.50	12
Paracetamol	1.203	(d, g)	*		0.34	0.31	0.34	3.22	6
Paroxetine	25.333	(a, b, n)	0.950	b	3.63	0.53	1.42	2.77	12
Pethidine (meperidine)	4.953	(d, d, g)	0.450	b	2.35	-0.75	1.15	2.62	6
Phenacetin	1.310	(g)	0.300	b	1.63	1.61	1.63	1.91	6
Phenazone (antipyrine)	0.590	(d, g)	0.100	b	0.27	-0.14	0.27	0.00	6
Phencyclidine	6.000	(b)	0.725	b	4.88	1.78	3.18	6.08	6
Phenobarbital	0.827	(c, c, g)	0.500	b	1.67	1.67	1.56	0.40	6
Phenylbutazone	0.159	(b, g)	0.990	b	3.16	3.16	2.11	1.73	12
Phenytoin	0.613	(c, g, k)	0.900	b	2.52	2.52	2.48	2.65	6
Prazepam	1.500	(b)	0.970	b	3.86	1.99	3.86	4.26	12

(Continued)

Table 1 (Continued)

Name	Vd (L kg ⁻¹)	Ref.	ppb	Ref.	log P	log D ₁	log D _{7.4}	μ _{MIM}	C _{arom}
Primidone	0.600	(b)	0.200	b	0.40	0.40	0.40	4.47	6
Propofol	3.500	(f)	0.975	f	4.16	4.16	4.16	1.38	6
Protriptyline	14.763	(a, b, g)	0.920	b	5.06	1.96	2.21	1.15	12
Salicylamide	0.147	(b, b, g)	0.750	b	1.41	1.41	1.37	4.31	6
Sertraline	20.000	(b)	0.980	b	4.81	1.71	2.77	2.34	12
Temazepam	1.257	(b)	0.970	b	2.20	1.53	2.20	4.57	12
Theobromine	0.750	(b)	*		-0.72	-0.84	-0.72	3.06	0
Tramadol	3.000	(b)	0.050	b	2.51	-0.59	0.36	2.23	6
Triazolam	1.113	(a, e)	0.780	b	3.45	1.15	2.62	8.05	12
Valproic acid	0.173	(a, b, h)	0.900	b	2.72	2.72	0.16	1.65	0
Viloxazine	1.000	(b)	0.865	b	1.10	-2.00	0.04	0.71	6

^aPerry (2002); ^bMoffat et al (1986); ^cLam et al (1997); ^dDurnas et al.(1990); ^eSchoerlin et al (1990); ^fFulton & Sorkin (1995); ^gRitschel & Hammer (1980); ^hRitschel et al (1995); ⁱRaaflaub & Speiser-Courvoisier (1974); ^jGlare & Walsh (1991); ^kThummel & Shen (2001); ^lSonne et al (1988); ^mGreenblatt (1981); ⁿPotter & Hollister (2001); ^oNattell et al (1987).

Table 2 The physicochemical and structural descriptors used in the study

Method	Parameter
Nemesis	Solvent accessible surface area (SA), dipole moment calculated by the Charge-2 method (μ _{MIM}), the highest and the lowest electrostatic potentials on the solvent accessible surface (ESP ⁺ and ESP ⁻ , respectively)
MOPAC 7.0 (MNDO Hamiltonian)	The energies of the highest occupied and the lowest unoccupied molecular orbitals (E _{HOMO} and E _{LUMO} , respectively), dipole moment (μ), the highest and the lowest atomic charges in the molecule (Q ⁺ and Q ⁻), principle moments of inertia (IM), length of the molecule (l), molecular weight (MW)
MOLCONN-Z	Simple and valence corrected molecular connectivity indices including zero- through fourth-order path (⁰ χ _p ^{v-4} χ _p and ⁰ χ _p ^{v-4} χ _p ^v), fourth order path-cluster (⁴ χ _{pc} and ⁴ χ _{pc} ^v), third- and fourth-order cluster (³ χ _c , ⁴ χ _c , ³ χ _c ^v and ⁴ χ _c ^v) and fifth- and sixth-order chain (⁵ χ _{ch} and ⁶ χ _{ch}), the highest atomic electrotopological index (S(I)), molecular shape indexes (⁰ κ ⁻³ κ and ⁰ κ _a ⁻³ κ _a), and delta connectivity indexes (X-X3)
ACD/LOGD	Calculated partition coefficient (log P), calculated distribution coefficient at pH 7.4 and 1 (log D _{7.4} and log D ₁), calculated pK _a (pK _a), the fraction unionised at pH 7.4 (fu), polarisability (α), molecular volume (V), molar refractivity (MR) and parachor (PA)
Experimental parameters	Experimental pK _a obtained from the literature ((pK _{a(Exp)}), see Table 3 for the references); log P obtained from the Biobyte database star list (log P*); log D _{7.4calc} is log D calculated using pK _{a(Exp)} and log P values at pH 7.4; fiB, fiA and fu _{calc} are, respectively, cationic, anionic and un-ionised percents at pH 7.4 calculated using pK _{a(Exp)} values
Other parameters	The total number of oxygen and nitrogen atoms (N _{N+O}), number of hydrogen atoms connected to oxygen or nitrogen (N _H), number of double bonds (N ₌), number of rotatable bonds (N _{rotat}), total number of bonds (N _{bond}), number of aromatic carbon atoms (C _{arom}), number of aliphatic carbon atoms (C _{alip}), logarithms of molecular surface area and weight (log SA and log MW), logarithm of the unionised fraction at pH 7.4 divided by the experimental pK _a (log(fu _{calc} /pK _{a(Exp)})), molecular volume divided by the length (V/l) and molecular weight divided by the volume (MW/V)

stepwise was cut short when addition of the third or fourth parameter did not add to the interpretability and predictability of the models. QSARs were sought for the whole dataset and also for the acidic and basic drugs separately. A compound was allocated to the acidic group of drugs if the percent ionised as an acid (anionic percent, fiA) was higher than the percent ionised as a base (cationic percent, fiB) at pH 7.4 and was allocated to the basic group if fiB was higher than fiA. While deletion of outliers often improves

the statistics of a QSAR, it was decided to keep all the compounds in the study, unless they affected the coefficients of equations significantly.

To test the predictive power of the models, the data sets were divided into the two equal groups, a training set and test set. To this end, the data were ranked based on the ascending Vd values and every other compound was allocated in the test set and the remaining compounds were assigned into the training set. Stepwise regression on the

Table 3 Experimental pK_a values for acids and bases, percent ionised as an acid and as a base (fiA and fiB, respectively) and percent un-ionised ($f_{u,calc}$), together with the values of Q^- , MW/V, $^3\chi_p$, S(I), ESP^- and $^3\kappa_a$ for the drugs in this study

Name	$pK_{a(Exp)}$ (acid)	$pK_{a(Exp)}$ (base)	fiA	fiB	$f_{u,calc}$	Q^-	MW/V	$^3\chi_p$	S(I)	ESP^-	$^3\kappa_a$
Acetanilide	*	0.50 ^b	0E+00	1E-05	100.0	-0.39	1.34	2.53	10.50	-36.2	2.33
Alprazolam	*	2.40 ^a	0.000	0.001	100.0	-0.25	1.25	8.58	6.26	-58.0	2.29
Amfetamine	*	9.94 ^b	0.000	99.712	0.3	-0.27	0.97	2.53	5.62	-31.1	2.55
Amitriptyline	*	9.40 ^b	0.000	99.010	1.0	-0.41	1.17	7.26	2.42	-26.1	3.09
Amobarbital	7.94 ^b	*	22.385	0.000	77.6	-0.41	1.26	5.32	11.79	-29.5	2.56
Bromazepam	11.00 ^a	2.90 ^a	0.025	0.003	100.0	-0.39	2.13	6.81	11.66	-44.1	2.47
Bupivacaine	*	8.00 ^a	0.000	79.924	20.1	-0.47	1.18	6.80	12.54	-44.0	3.69
Bupropion	*	7.00 ^b	0.000	28.475	71.5	-0.36	1.07	4.31	11.92	-34.0	3.73
Butorphanol	*	8.60 ^b	0.000	94.065	5.9	-0.46	1.18	10.44	11.95	-40.2	2.05
Caffeine	14 ^b	0.60 ^b	3E-05	2E-05	100.0	-0.43	2.24	5.88	11.49	-38.1	1.09
Carbamazepine	14.07 ^b	-0.46 ^b	2E-05	0E+00	100.0	-0.41	1.37	6.35	11.68	-49.2	1.60
Chlordiazepoxide	*	4.80 ^a	0.000	0.251	99.7	-0.37	1.55	7.57	12.57	-39.6	2.74
Chlorphentermine	*	9.60 ^b	0.000	99.373	0.6	-0.26	1.46	3.04	5.75	-33.6	3.63
Clobazam	8.59	*	6.060	0.000	93.9	-0.40	1.18	7.88	12.48	-39.9	2.49
Clomipramine	*	9.46 ^a	0.000	99.137	0.9	-0.45	1.15	7.61	6.27	-24.9	3.54
Clonazepam	*	1.57 ^a	0.079	0.000	99.9	-0.38	1.48	7.70	11.67	-32.8	2.52
Cloral hydrate	10.00 ^a	*	0.251	0.000	99.7	-0.32	2.30	1.73	8.05	-22.3	3.38
Clorazepate	3.50 ^a	*	99.987	0.000	0.0	-0.37	1.76	7.92	11.31	-44.9	2.86
Desipramine	*	10.2 ^b	0.000	99.842	0.2	-0.37	1.18	7.09	3.24	-28.5	2.79
Diazepam	*	3.30 ^a	0.000	0.008	100.0	-0.42	1.57	7.98	12.10	-41.6	2.62
Doxepine	*	8.00 ^b	0.000	79.924	20.1	-0.41	1.19	7.26	5.98	-29.2	3.07
Ethchlorvynol	12.06	*	0.002	0.000	100.0	-0.32	3.43	2.56	9.22	-23.6	2.55
Ethosuximide	9.30 ^a	*	1.243	0.000	98.8	-0.42	1.16	3.57	10.70	-34.3	1.12
Etidocaine	*	7.70 ^a	0.000	66.614	33.4	-0.41	1.02	6.42	12.57	-47.6	4.42
Fentanyl	*	8.40 ^b	0.000	90.909	9.1	-0.47	0.89	8.64	12.52	-42.5	4.77
Flunitrazepam	*	1.80 ^a	0E+00	3E-04	100.0	-0.42	1.48	8.74	14.14	-38.1	2.67
Fluoxetine	*	10.06	0.000	99.782	0.2	-0.35	1.22	6.65	12.60	-26.9	4.37
Glutethimide	4.52 ^b	*	99.868	0.000	0.1	-0.42	1.23	5.92	12.03	-34.9	1.82
Haloperidol	*	8.30 ^b	0.000	88.818	11.2	-0.41	1.45	9.30	13.18	-41.8	4.92
Ibuprofen	5.20 ^b	*	99.373	0.000	0.6	-0.31	1.06	4.49	10.76	-40.5	3.48
Imipramine	*	9.50 ^b	0.000	99.212	0.8	-0.45	1.12	7.26	2.52	-26.8	3.12
Indometacin	4.50 ^a	*	99.874	0.000	0.1	-0.38	2.10	9.12	12.99	-36.8	3.27
Ketamine	*	7.50 ^a	0.000	55.731	44.3	-0.37	1.28	6.02	12.18	-43.6	1.91
Lidocaine	*	7.86 ^b	0.000	74.254	25.7	-0.44	1.06	5.35	11.89	-46.2	3.95
Lorazepam	*	1.3 ^a	0.008	0.000	100.0	-0.35	1.90	7.71	11.79	-49.5	2.86
Maprotiline	*	10.02 ^b	0.000	99.761	0.2	-0.35	1.19	8.41	3.31	-28.9	1.80
Meprobamate	13.39	*	1E-04	0E+00	100.0	-0.41	1.32	3.03	10.37	-38.4	5.49
Meptazinol	*	8.70 ^a	0.000	95.227	4.8	-0.43	2.31	5.43	9.58	-25.1	2.65
Methadone	*	8.25 ^b	0.000	87.623	12.4	-0.44	2.16	8.07	13.17	-38.3	3.53
Methaqualone	*	2.54 ^a	0.000	0.001	100.0	-0.36	1.43	7.25	12.67	-43.7	1.97
Metoclopramide	*	9.00 ^a	0.000	97.550	2.5	-0.42	1.16	6.58	12.11	-50.5	4.82
Midazolam	*	6.20 ^a	0.000	5.935	94.1	-0.35	1.53	8.71	14.23	-40.3	2.10
Moclobemide	*	6.89	0.000	23.608	76.4	-0.43	1.35	5.78	11.83	-40.0	4.36
Morphine	9.26 ^a	8.18 ^a	1.362	85.766	12.9	-0.31	1.38	9.70	10.44	-27.9	1.26
Naloxone	*	7.94 ^b	0.000	77.615	22.4	-0.35	0.80	10.61	12.60	-32.8	1.54
Nitrazepam	10.80 ^a	3.2 ^a	0.040	0.006	100.0	-0.39	1.43	7.54	11.69	-34.9	2.62
Nortriptyline	*	9.73 ^b	0.000	99.534	0.5	-0.35	1.21	7.09	3.23	-28.8	2.71
Oxazepam	11.10 ^a	1.80 ^a	0.020	0.000	100.0	-0.37	1.64	7.83	11.81	-44.1	2.74
Oxyphenbutazone	4.70 ^b	*	99.801	0.000	0.2	-0.28	1.62	9.43	13.04	-33.9	2.41
Paracetamol	9.50 ^b	*	0.788	0.000	99.2	-0.38	2.16	2.94	10.52	-35.1	2.49
Paroxetine	*	9.72	0.000	99.524	0.5	-0.03	1.42	8.89	13.56	-22.2	3.52
Pethidine	*	8.70 ^b	0.000	95.227	4.8	-0.44	1.23	6.33	12.39	-40.0	2.65
Phenacetin	*	2.20 ^b	0.000	0.001	100.0	-0.39	1.78	3.47	10.68	-34.1	3.36
Phenazone	*	1.40 ^b	0E+00	1E-04	100.0	-0.31	1.25	5.10	11.60	-21.3	1.51
Phencyclidine	*	8.50 ^b	0.000	92.641	7.4	-0.32	1.04	6.39	2.79	-21.9	2.80
Phenobarbital	7.40 ^b	*	50.000	0.000	50.0	-0.41	1.33	6.33	11.98	-30.4	1.69
Phenylbutazone	4.50 ^b	*	99.874	0.000	0.1	-0.28	1.34	8.96	13.00	-33.8	2.31
Phenytoin	8.33 ^a	*	10.514	0.000	89.5	-0.44	1.41	7.11	12.31	-37.5	1.78
Prazepam	*	2.70 ^a	0.000	0.002	100.0	-0.42	1.19	8.27	12.59	-43.3	3.07
Primidone	12.26	*	0.001	0.000	100.0	-0.45	2.09	6.09	11.99	-39.8	1.60

(Continued)

Table 3 (Continued)

Name	pK _{a(Exp)} (acid)	pK _{a(Exp)} (base)	fiA	fiB	fu _{calc}	Q ⁻	MW/V	³ χ _p	S(I)	ESP ⁻	³ κ _a
Propofol	11.00	*	0.025	0.000	100.0	-0.26	1.02	4.14	9.93	-25.2	2.10
Protriptyline	*	10.61	0.000	99.938	0.1	-0.35	1.13	6.84	3.28	-29.8	2.34
Salicylamide	8.20 ^b	*	13.681	0.000	86.3	-0.36	1.63	3.03	10.48	-24.4	1.41
Serteraline	*	9.47	0.000	99.156	0.8	-0.36	1.41	7.67	6.02	-23.5	2.61
Temazepam	*	1.60 ^a	0E+00	2E-04	100.0	-0.43	1.93	8.49	12.18	-43.6	2.72
Theobromine	10.05 ^b	*	0.223	0.000	99.8	-0.40	2.69	4.24	11.12	-39.2	0.76
Tramadol	*	8.30 ^a	0.000	88.818	11.2	-0.44	1.18	6.56	11.21	-27.4	2.93
Triazolam	*	8.19 ^b	0.000	86.045	14.0	-0.24	1.70	9.05	6.40	-58.4	2.48
Valproic acid	5.00 ^b	*	99.603	0.000	0.4	-0.37	0.91	2.26	10.48	-32.1	4.16
Viloxazine	*	8.10 ^a	0.000	83.366	16.6	-0.34	1.20	5.40	5.74	-26.6	3.78

^aMoffat et al (1986); ^bFoye et al (1995).

training set often led to the models containing parameters other than those involved in the models for the whole dataset. Therefore, for simplicity, multiple regression analysis was performed on the training set using the parameters obtained from the previous stepwise regression analyses and the model obtained was used to calculate the Vd values of the compounds in the test set. Fold error of prediction for the test set was calculated according to equation 8. Because of the skewed distribution of fold error, the geometric mean was calculated and reported as the mean fold error.

$$\text{Fold error} = \text{antilog}(|\log V_{d, \text{obs.}} - \log V_{d, \text{pred.}}|) \quad (8)$$

The following statistical details of the models were noted: n, the number of observations; r, the correlation coefficient; s, the standard deviation; F, the Fisher statistic; and the P value. The figures in parentheses with the regression coefficients were standard errors of coefficients.

Results

The apparent volume of distribution (Vd) and the extent of protein binding for the compounds used in this study are listed in Table 1, together with the relevant references. Data were obtained for a wide range of drug substances. These include central nervous system agents and other drugs, such as benzodiazepines, barbiturates, hydantoin, succinimides, morphine and related analgesics, tricyclic antidepressants, phenothiazine derivatives, butyrophenones, anaesthetics, non-steroidal anti-inflammatory drugs and so on. The Vd values cover a range of 0.1–41 L kg⁻¹. Transformation of the Vd data to a logarithmic scale leads to a normal distribution with skewness of 0.005.

QSAR model for Vd

The stepwise regression analyses for Vd resulted in the following equation:

$$\log V_d = -0.151 (\pm 0.13) + 0.364 (\pm 0.038) \log P - 0.260 (\pm 0.039) \log D_1 - 0.086 (\pm 0.027) \mu_{\text{MM}}(9)$$

$$n = 70 \quad s = 0.390 \quad r = 0.787 \quad F = 35.7 \quad P = 0.000$$

Equation 9 indicates that there is a correlation between partition coefficient and Vd, but the overall relationship is improved by the introduction of other computed parameters. The distribution coefficient at pH 1 (log D₁) has a negative effect on Vd. The distribution coefficient at this extreme pH value represents a different feature of the molecules than hydrophobicity alone. At pH 1, acids are completely in their un-ionised form and bases are fully protonated, thus acids will have a log D₁ value close to their log P, whereas for bases, log D₁ will be much lower than the true log P. In other words, log D₁ is higher (and therefore, according to equation 9, Vd is lower) for compounds which are less ionised at this acidic pH; these are either acidic drugs or drugs with lower basicities. It should be noted that the variables in the equation are not strongly correlated with each other, the highest correlation being between log D₁ and log P with r = 0.488. Log P, log D₁ and μ_{MM} used in equation 9, as well as log D_{7.4}, are listed in Table 1.

QSAR for unbound Vd

A common approach to model the volume of distribution is to correct for plasma protein binding and derive QSARs for the unbound (intrinsic) volumes (Ritschel & Hammer 1980; Blakey et al 1997). The fraction of protein-bound drug in plasma, ppb, was collected from the literature (Table 1). The unbound volume of distribution was calculated by dividing Vd by the fraction of non-protein-bound drug in plasma. Stepwise regression analysis on the unbound volume of distribution (volume of distribution of free drug, Vdu) resulted in equation 10. The ppb values of four drugs in Table 1 were not found in the literature and they were omitted from the regression analysis:

$$\log V_{du} = -0.424 (\pm 0.139) + 0.396 (\pm 0.047) \log P + 0.056 (\pm 0.016) C_{\text{arom}} \quad (10)$$

$$n = 66 \quad r = 0.847 \quad s = 0.437 \quad F = 80.0$$

where C_{arom} is the number of aromatic atoms in the molecule.

The equation shows that the volume of distribution of the free (unbound) drug is related to the partition coefficient and the number of aromatic atoms (listed in Table 1) in the molecule.

QSARs for acidic drugs

Based on the fiA and fiB values (Table 3), acidic drugs were identified as described in the Methods. There were 27 such drugs in this group. The following equations resulted from the stepwise regression analysis on log Vd and log Vdu:

$$\log Vd = -2.56 (\pm 0.63) + 0.254 (\pm 0.064) \log D_{7.4} - 4.08 (\pm 1.39) Q^- + 0.315 (\pm 0.121) MW/V \quad (11)$$

$$n = 27 \quad s = 0.333 \quad r = 0.686 \quad F = 7.0$$

$$\log Vdu = 1.92 (\pm 0.88) + 0.393 (\pm 0.054) \log P + 0.218 (\pm 0.050) {}^3\chi_p - 0.297 (\pm 0.099) S(I) \quad (12)$$

$$n = 25 \quad s = 0.259 \quad r = 0.920 \quad F = 38.6$$

Note: there were 2 missing ppb values for acidic drugs.

In equations 11 and 12, Q^- is the lowest atomic charge in the molecule, MW/V is molecular weight divided by the volume of molecule (density), ${}^3\chi_p$ is the third-order path molecular connectivity index and $S(I)$ is the highest electrotopological state index in the molecule.

The importance of lipophilicity in the distribution process is expressed by the presence of the distribution coefficient at pH 7.4 in equation 11 and the partition coefficient in equation 12. In equation 11, the lowest atomic charge in the molecule has a negative sign, which shows that the presence of a heteroatom with higher electronegativity is favoured. Bearing in mind the significance of the lipophilicity parameter, this suggests that an electronegative heteroatom without a free hydrogen atom (non-hydrogen-bond donor) will increase Vd. The third parameter in equation 11 is the density of molecules, meaning that the presence of heavy atoms (bromine, chlorine, oxygen and nitrogen) increases log Vd.

The second parameter in equation 12 is ${}^3\chi_p$, which, in addition to the size of the molecule, represents the number of three-bond fragments and adjacency of branch points of a ring in the molecule (Hall & Kier 2001). When omitted from the stepwise regression analysis, alternatives to this parameter were other connectivity parameters (fourth, second, first and zero order) and polarisability, which suggests, in this analysis, that ${}^3\chi_p$ represents mainly the size of the molecule. The negative coefficient of $S(I)$ shows the negative effect of increasing the availability of atomic electron densities. Table 3 shows the values of Q^- , MW/V , ${}^3\chi_p$ and $S(I)$ for all the drugs.

QSARs for basic drugs

The following equations were obtained for the basic drugs in the dataset:

$$\log Vd = 0.586 (\pm 0.116) - 0.325 (\pm 0.042) \log (fu_{\text{calc}}/pK_a) + 0.149 (\pm 0.044) \log D_{7.4} - 0.078 (\pm 0.031) \mu_{\text{MM}} \quad (13)$$

$$n = 43 \quad s = 0.319 \quad r = 0.834 \quad F = 29.6$$

where $\log (fu_{\text{calc}}/pK_a)$ is the logarithm of the percent of drug unionised at pH 7.4 divided by the pK_a of the bases. The inclusion of the distribution coefficient at pH 7.4 and μ_{MM} in equation 13 is analogous to equations 9 and 11.

Two new parameters are included in the QSAR for the apparent volume of distribution of free drug (note that there are two missing ppb data for basic drugs):

$$\log Vdu = -0.004 (\pm 0.38) + 0.270 (\pm 0.061) \log P + 0.105 (\pm 0.027) C_{\text{arom}} + 0.022 (\pm 0.007) ESP^- + 0.121 (\pm 0.076) {}^3\kappa_a \quad (14)$$

$$n = 41 \quad s = 0.433 \quad r = 0.854 \quad F = 24.3$$

where ESP^- is the lowest (most negative) electrostatic potential on the solvent-accessible surface of the molecule and ${}^3\kappa_a$ is the third-order kappa alpha shape index.

ESP^- may be regarded as a measure of the ability to engage in electrostatic interaction with positively charged particles. The positive coefficient shows the detrimental effect of such an ability (i.e. the higher negative potential leads to lower log Vdu). The molecular shape parameter, ${}^3\kappa_a$, describes the cyclicity of molecular graphs (Kier 1987), implying a higher volume of distribution for molecules with fewer rings. Values of ESP^- and ${}^3\kappa_a$ for all the drugs are listed in Table 3.

Validation of the QSAR models

Table 4 shows the equations obtained for the training set based on equations 9–14. The Vd values predicted for the test sets using the equations for the training sets (equations 15–20) are tabulated in Table 5 for all drugs, acidic drugs and basic drugs. The goodness of prediction has been presented by mean fold error and prediction accuracy (i.e. fraction of compounds predicted to have a Vd value within a 2-fold error from the experimental value).

Discussion

The distribution of a compound in the human body is a function of its affinity to various tissues. It is related to the extent of binding in tissues vs the extent of binding in plasma (Rowland & Tozer 1995). Plasma protein binding limits the concentration of drug available for metabolism and distribution in-vivo. A common approach to model the volume of distribution is to correct for this factor by dividing the Vd by the fraction unbound to plasma protein, and derive QSARs for the unbound (intrinsic) volumes (Ritschel & Hammer 1980; Blakey et al 1997). This approach has been criticised by Davis et al (2000) on the basis that protein binding itself is related to structural parameters, such as partition coefficient, which are commonly used in QSAR models. Lombardo et al (2002)

Table 4 QSARs developed for the validation of the models

Equation no.	Equation for training set	n	r	s	F
15 (model 9)	$\log Vd = -0.157 + 0.367 \log P - 0.240 \log D_1 - 0.101 \mu_{MM}$	35	0.812	0.377	20.0
16 (model 10)	$\log Vdu = -0.542 + 0.412 \log P + 0.063 C_{arom}$	32	0.881	0.423	50.5
17 (model 11)	$\log Vd = -2.33 + 0.290 \log D_{7.4} - 3.48 Q^- + 0.252 MW/V$	14	0.663	0.396	3.0
18 (model 12)	$\log Vdu = 1.63 + 0.565 \log P + 0.219^3 \chi_p - 0.307 S(I)$	13	0.935	0.268	20.1
19 (model 13)	$\log Vd = 0.522 - 0.373 \log(fu_{calc}/pKa) + 0.118 \log D_{7.4} - 0.033 \mu_{MM}$	21	0.849	0.306	14.6
20 (model 14)	$\log Vdu = 0.698 + 0.347 \log P + 0.047 C_{arom} + 0.023 ESP^- + 0.022^3 K_a$	21	0.856	0.430	11.0

Table 5 The observed Vd values and the Vd values predicted using the equations for the training set together with the corresponding mean fold error and prediction accuracy

Name	Vd _{obs.} (L kg ⁻¹)	Vd _{pred} for drugs in the test sets					
		Eq. 9	Eq. 10	Eq. 11	Eq. 12	Eq. 13	Eq. 14
Acetanilide	0.16					0.46	*
Alprazolam	0.81	0.70	5.28				
Amfetamine	3.77	4.74	2.77			9.47	6.50
Amitriptyline	10.64						
Amobarbital	1.00						
Bromazepam	1.18			1.38	2.25		
Bupivacaine	1.00	4.29	2.17			3.67	2.08
Bupropion	13.13	7.83	2.77			3.06	4.61
Butorphanol	5.00						
Caffeine	0.53	0.27	0.16	0.49	0.13		
Carbamazepine	1.13	0.62	5.17				
Chlordiazepoxide	0.34						
Chlorphentermine	2.50					8.31	*
Clobazam	1.12			1.11	1.03		
Clomipramine	14.67						
Clonazepam	3.13	0.95	2.27	1.05	1.68		
Cloral hydrate	0.60	0.79	0.62				
Clorazepate	0.72			0.26	2.83		
Desipramine	33.67						
Diazepam	1.92					1.53	0.37
Doxepine	10.27	10.02	12.78			3.90	19.91
Ethchlorvynol	2.50	1.03	0.81	1.77	1.32		
Ethosuximide	0.68			0.35	0.22		
Etidocaine	2.00	4.09	1.47				
Fentanyl	3.60						
Flunitrazepam	4.00	0.75	1.18			0.83	1.67
Fluoxetine	41.00	6.75	4.77			17.22	8.63
Glutethimide	3.08						
Haloperidol	16.01						
Ibuprofen	0.10						
Imipramine	17.25	14.81	15.59			16.16	23.92
Indometacin	0.96	0.76	1.55	0.32	1.21		
Ketamine	4.00						
Lidocaine	1.44	2.61	1.93				
Lorazepam	1.20	0.51	1.71				
Maprotiline	22.00						
Meprobamate	0.70	0.83	0.45				
Meptazinol	5.47	8.34	16.76			5.81	40.83
Methadone	3.94						
Methaqualone	6.00						
Metoclopramide	3.00	2.24	1.97				
Midazolam	1.91	2.91	3.41				
Moclobemide	1.29					1.23	1.41
Morphine	3.21					2.23	2.47

(Continued)

Table 5 (Continued)

Naloxone	3.00	2.70	1.63				
Nitrazepam	2.28						
Nortriptyline	21.33	14.98	26.19			21.79	31.27
Oxazepam	0.80			1.11	0.52		
Oxyphenbutazone	0.11	1.09	0.22	0.42	0.17		
Paracetamol	1.20			0.42	*		
Paroxetine	25.33	5.87	2.57			12.21	6.12
Pethidine	4.95	4.18	3.51			4.66	4.78
Phenacetin	1.31						
Phenazone	0.59					0.73	3.75
Phencyclidine	6.00	3.93	19.34			5.24	47.02
Phenobarbital	0.83						
Phenylbutazone	0.16	1.18	0.33	0.40	0.24		
Phenytoin	0.61	0.79	0.75				
Prazepam	1.50					1.78	1.42
Primidone	0.60			0.75	0.25		
Propofol	3.50	1.71	0.89				
Protriptyline	14.76	13.02	15.96			37.93	19.37
Salicylamide	0.15						
Serteraline	20.00						
Temazepam	1.26	0.67	0.40				
Theobromine	0.75	0.30	*				
Tramadol	3.00					2.77	18.41
Triazolam	1.11	1.05	9.53			3.01	3.30
Valproic acid	0.17	1.05	0.38				
Viloxazine	1.00						
Prediction accuracy		22 of 35	17 of 34	8 of 13	8 of 12	11 of 22	9 of 20
Mean fold error		2.03	2.29	1.89	1.99	1.84	2.55

*ppb value was not available.

suggested that the fraction unbound in tissues (f_{ut}) is a better end-point for QSAR analysis than Vd, and subsequently back-calculated the Vd. They employed the Oie-Tozer equation (Oie & Tozer 1979) to calculate f_{ut} based on a knowledge of protein binding and volume of distribution of drugs, as well as the constant values for volumes of plasma, extracellular fluid and remainder fluid, and the ratio of extravascular to intravascular proteins.

To investigate the effect of protein binding correction in this study, the unbound volume of distribution, Vdu, has been compared with Vd for the development of QSARs. Comparison of equations 10, 12 and 14 with equations 9, 11 and 13 shows that QSARs for unbound Vd provide statistically better models with higher r and F values. However, predictive ability, as tested by the mean fold errors for the test sets, does not reflect this statistical superiority (Figure 1). This could be due to the fact that the parameters in equations 10, 12 and 14 may partly describe the protein binding extent, leading to higher error of prediction for Vd. To explore this further, unbound Vd may be rewritten as a subtraction of the two terms, a Vd term and a protein binding term as follows:

$$\log[Vd/(1 - ppb)] = \log Vd - \log[1/(1 - ppb)] \quad (15)$$

The contribution of each term in equation 15 to the relationship with the unbound Vd (equation 10) was

examined by regressing $\log Vd$ and $\log [1/(1 - ppb)]$ against the structural descriptors in equation 10. The results (equations 16 and 17) show that while $\log P$ is a significant descriptor of both Vd and protein binding, C_{arom} is describing protein binding only. Note that in equation 16, C_{arom} is not statistically significant ($P = 0.587$).

$$\log Vd = -0.402 (\pm 0.17) + 0.234 (\pm 0.058) \log P + 0.010 (\pm 0.019) C_{arom} \quad (16)$$

$$n = 66 \quad s = 0.524 \quad r = 0.552 \quad F = 13.8$$

$$\log [1/(1 - ppb)] = 0.022 (\pm 0.131) + 0.162 (\pm 0.044) \log P + 0.045 (\pm 0.015) C_{arom} \quad (17)$$

$$n = 66 \quad s = 0.411 \quad r = 0.659 \quad F = 24.2$$

Similar conclusions can be drawn from the corresponding equations obtained for acids and bases, where, using the example of the acids, none of the parameters of equation 12 were statistically correlated to $\log Vd$ (results not presented).

All the QSARs obtained in this study include either a partition, or a distribution, coefficient parameter. This is consistent with many previously published reports that high lipophilicity is associated with large Vd (Ritschel & Hammer 1980; Van der Graaf et al 1999; Van de Waterbeemd et al

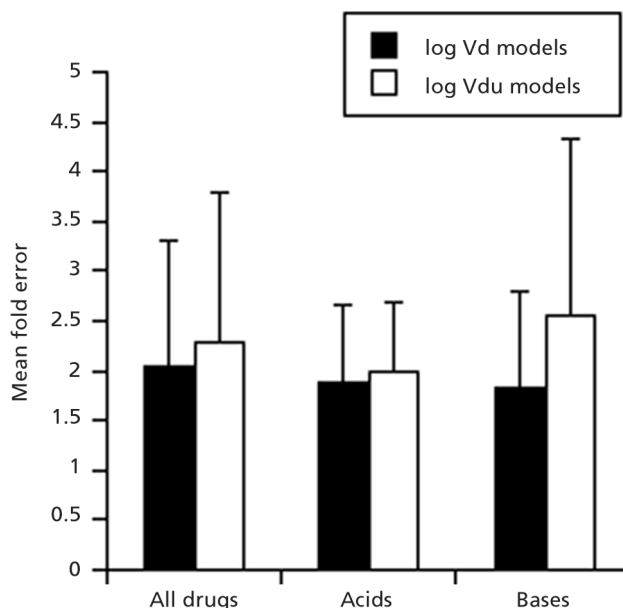


Figure 1 Mean fold error of prediction for the test set associated with different models; error bars show the average deviation.

2001). Although the experimental $\log P$ and $\log D_{7.4}$ parameters had also been used in the stepwise regression analysis, only the ACD calculated parameters appeared in the QSARs. There was generally good agreement between $\log P^*$ and the ACD calculated $\log P$ ($r = 0.881$). The correlation between ACD calculated $\log D_{7.4}$ and $\log D_{7.4\text{calc}}$ (calculated from experimental pK_a and $\log P^*$) was not as good ($r = 0.709$). This is probably due to the deviations of calculated pK_a , as well as $\log P$, from the experimental values. For example, one significant outlier from the correlation, glutethimide, is an acid with a $pK_{a(\text{Exp})}$ value of 4.52 and an ACD calculated pK_a of 11.36. The most important feature of equation 9 is that it shows a higher volume of distribution for basic drugs compared with the acids (note the negative coefficient of the $\log D_1$ parameter). This could be due partly to the high protein binding of most acidic drugs. Karalis et al (2002) observed that in a class of compounds with higher volume of distribution, the acid/base ratio was lower, whereas protein-binding extent was highest in the class with the lowest volume of distribution. In equation 11 for basic drugs, the inclusion of $\log(fu_{\text{calc}}/pK_a)$ shows that drugs with a higher cationic fraction have a higher volume of distribution. In addition, $\log fu_{\text{calc}}$ increases with increase in $\log(fu_{\text{calc}}/pK_a)$ values, and reaches a maximum where it remains almost constant with increasing $\log(fu_{\text{calc}}/pK_a)$ values (Figure 2). The maximum corresponds to compounds with low pK_a values, where fu_{calc} does not reflect changes in pK_a . For example, fu_{calc} values remain constant at 100.00 for pK_a ranges of 0.5–2.4 (see Table 3). According to equation 13, the stronger the basicity of a drug (i.e. a lower percentage in the unionised form and high pK_a values), the higher the apparent volume of distribution it will maintain; this is not due merely to the higher

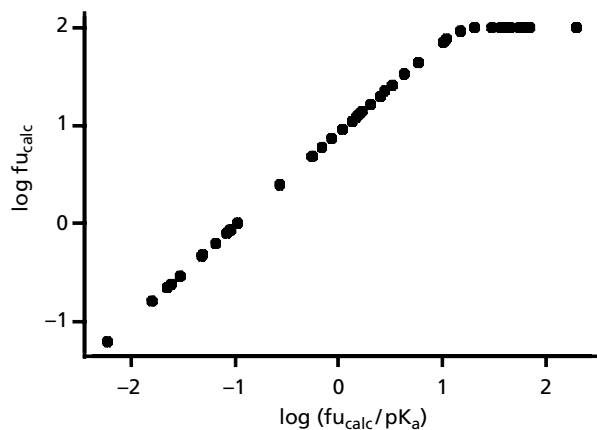


Figure 2 The plot of $\log fu_{\text{calc}}$ against $\log(fu_{\text{calc}}/pK_a)$ for all drugs considered in the analysis.

ionisation but also due to the electron directing properties of the rest of the molecule (substituents) on the basic nitrogen atom of the molecule.

According to Figure 1, the mean fold errors of prediction from the QSAR for $\log Vd$ show some decrease when the drugs are separated into acidic and basic groups. The similar mean fold errors of prediction could be due to the fact that the number of chemicals is lower in the acidic and basic drug sets. This was explored further by splitting the datasets in a 4:1 ratio of training-to-test sets. Mean fold errors of prediction using equations 11–14 (separate equations for acids and bases) were decreased to 1.44, 1.55, 1.73 and 2.06, respectively. On the other hand, the new splitting process did not change the prediction errors of equations 9 and 10 significantly. Therefore, the results suggest the use of models 11 and 13 for prediction purposes. This will require a knowledge of experimental pK_a values as well as the calculated parameters of these equations. On the other hand, the apparent volume of distribution of all drugs could be predicted using equation 9, with lower accuracy, without the need of any experimental measurement, including pK_a and the extent of protein binding. Prediction via equation 9 has a mean fold error of 2.03. A similar mean fold error has been reported in a previous study, where only basic and neutral drugs were studied and the ratio of the number of compounds in the test series to those in the training set was 14:50 (Lombardo et al 2002). In comparison with the error normally associated with the prediction using the interspecies scaling, which is reported to be in the range 1.56–2.78 (Obach et al 1997), our mean fold error is encouraging. To confirm this point further, the mean fold error for the Vd predictions of a study based on extrapolation from animals to man (Mahmood 1998) was calculated to be 1.82.

In Table 5 nortriptyline, paroxetine and fluoxetine are drugs with a Vd value higher than 20 L kg^{-1} and are included in the test sets of equations 9, 10, 13 and 14. It is only equation 13 that provides a reasonably accurate predicted Vd value for these drugs. The extreme Vd values of paroxetine and fluoxetine have been highly underestimated

by equations 9, 10 and 14; that of nortriptyline has been overestimated by equation 14. Oxyphenbutazone, phenybutazone, valproic acid and acetanilide are the drugs with extremely low Vd values (lower than 0.2) that are included in the test sets. Equations 9 and 11 have highly overestimated all the extremely low Vd values, but this is not the case for the corresponding Vdu equations 10 and 12. It could be speculated that the low Vd values of these drugs are a result of the high ppb values.

Conclusions

Statistically significant QSARs were constructed for the volume of distribution of a group of drugs as a whole and when separated into acids and bases. Unbound volume of distribution was also investigated as a possible QSAR target. Comparing the different QSARs, it was concluded that although correction of Vd for protein binding improved the statistical fit, it lessened the predictive power of the QSAR. This investigation presented a predictive model for the prediction of volume of distribution without a need for experimental measurements (equation 9). Also, a more accurate prediction is possible using equations 11 and 13 for acids and bases, but this will require the experimental pK_a as well as calculated parameters. The mean fold error associated with the prediction using equation 9 is approximately 2, which is within the range of mean fold errors of prediction using the extrapolation methods from animal species. The advantage of the model is that all the parameters are computed and there is no need for experimental measurements. The model could find use in novel drug design and high throughput screening laboratories.

References

- Abraham, M. H., Chadha, H. S., Martins, F., Mitchell, R. C., Bradbury, M. W., Gratton, J. A. (1999) Hydrogen bonding-Part 46: A review of the correlation and prediction of transport properties by an LFER method: physicochemical properties, brain penetration and skin permeability. *Pestic. Sci.* **55**: 78–88
- Agatonovic-Kustrin, S., Beresford, R., Pauzi, A., Yusof, M. (2001) ANN modeling of the penetration across a polydimethylsiloxane membrane from theoretically derived molecular descriptors. *J. Pharmaceut. Biomed. Anal.* **26**: 241–254
- Balant, L. P., Gex-Fabry, M. (2001) Modelling in preclinical and clinical drug development. In: Testa, B., van de Waterbeemd, H., Folkers, G., Guy, R. (eds) *Pharmacokinetic optimization in drug research*. Wiley-VCH, Weinheim, Germany, pp 15–32
- Blakey, G. E., Nestorov, I. A., Arundel, P. A., Aarons, L. J., Rowland, M. (1997) Quantitative structure-pharmacokinetics relationships: I. Development of a whole body physiologically based model to characterize changes in pharmacokinetics across a homologous series of barbiturates in the rat. *J. Pharmacokinetic. Biopharm.* **25**: 277–312
- Davis, A. M., Webborn, P. J. H., Salt, D. W. (2000) Robust assessment of statistical significance in the use of unbound/intrinsic pharmacokinetic parameters in quantitative structure-pharmacokinetic relationships with lipophilicity. *Drug Metab. Dispos.* **28**: 103–106
- Durnas, C., Mingloi, C., Cusack, B. (1990) Hepatic drug metabolism and aging. *Clin. Pharmacokinetic.* **19**: 359–389
- Egan, W. J., Lauri, G. (2002) Prediction of intestinal permeability. *Adv. Drug Deliv. Rev.* **54**: 273–289
- Foye, W. O., Lemke, T. L., Williams, D. A. (1995) *Principles of medicinal chemistry*. 4th edn, Lippincott Williams & Wilkins, pp 948–960
- Fulton, B., Sorkin, E. M. (1995) Propofol an overview of its pharmacology and a review of its clinical efficacy in intensive care sedation. *Drugs* **50**: 587–767
- Ghafourian, T., Barzegar-Jalali, M. (2002) Theoretical modeling of oral absorption of barbiturates. *Il Farmaco* **57**: 565–567
- Ghafourian, T., Fooladi, S. (2001) The effect of structural QSAR parameters on skin penetration. *Int. J. Pharm.* **217**: 1–11
- Glare, P. A., Walsh, T. D. (1991) Clinical pharmacokinetics of morphine. *Ther. Drug Monit.* **13**: 1–23
- Gobburu, J. V. S., Shelver, W. H. (1995) Quantitative structure-pharmacokinetic relationships (QSPR) of beta blockers derived using neural networks. *J. Pharm. Sci.* **84**: 862–865
- Greenblatt, D. J. (1981) Clinical pharmacokinetics of oxazepam and lorazepam. *Clin. Pharmacokinetic.* **6**: 89–105
- Hall, L. H., Kier, L. B. (2001) Issues in representation of molecular structure, the development of molecular connectivity. *J. Mol. Graph. Model.* **20**: 76–83
- Hirono, S., Nakagome, I., Hirano, H., Yoshii, F., Moriguchi, I. (1994) Non-congeneric structure-pharmacokinetic property correlation studies using fuzzy adaptive least squares: Volume of distribution. *Biol. Pharm. Bull.* **17**: 686–690
- Kansy, M., Fischer, H., Kratzat, K., Senner, F., Wagner, B., Parrilla, I. (2001) High-throughput artificial membrane permeability studies in early lead discovery and development. In: Testa, B., van de Waterbeemd, H., Folkers, G., Guy, R. (eds) *Pharmacokinetic optimization in drug research*. Wiley-VCH, Weinheim, Germany, pp 447–464
- Karalis, V., Tsantili-Kakoulidou, A., Macheras, P. (2002) Multivariate statistics of disposition pharmacokinetic parameters for structurally unrelated drugs used in therapeutics. *Pharm. Res.* **19**: 1827–1834
- Kennedy, T. (1997) Managing the drug discovery/development interface. *Drug Discovery Today* **2**: 436–444
- Kier, L. B. (1987) Indexes of molecular shape from chemical graphs. *Med. Res. Rev.* **7**: 417–440
- Klopman, G., Stefan, L. R., Saiakhov, R. D. (2002) ADME evaluation 2. A computer model for the prediction of intestinal absorption in humans. *Eur. J. Pharm. Sci.* **17**: 253–263
- Kulkarni, A., Han, Y., Hopfinger, A. J. (2002) Predicting Caco-2 cell permeation coefficients of organic molecules using membrane-interaction QSAR analysis. *J. Chem. Inf. Comput. Sci.* **42**: 331–342
- Lam, Y. W., Banerji S., Hatfield, C., Talbert, L. (1997) Principle of drug administration in renal insufficiency. *Clin. Pharmacokinetic.* **32**: 30–57
- Lipinski, C. A., Lombardo, F., Dominy, B. W., Feeney, P. J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **23**: 3–25
- Lombardo, F., Obach, R. S., Shalaeva, M. Y., Gao, F. (2002) Prediction of volume of distribution values in humans for neural and basic drugs using physicochemical measurements and plasma protein binding data. *J. Med. Chem.* **45**: 2867–2876
- Mahmood, I. (1998) Interspecies scaling: predicting volumes, mean residence time and elimination half-life. Some suggestions. *J. Pharm. Pharmacol.* **50**: 493–499

- Moffat, A. C., Jackson, J. V., Moss, M. S., Winddop, B. (1986) *Clarke's isolation of drug in pharmaceuticals, body fluids, and post-mortem material*. 2nd edn, The Pharmaceutical Press, London, pp 309–1060
- Moss, G. P., Dearden, J. C., Patel, H., Cronin, M. T. D. (2002) Quantitative structure-permeability relationships (QSPRs) for percutaneous absorption. *Toxicol. In Vitro* **16**: 299–317
- Nattell, S., Gagne, G., Pineau, M. (1987) The pharmacokinetics of lidocaine and β -adrenoreceptor antagonists in patients with acute myocardial infarction. *Clin. Pharmacokinet.* **13**: 293–316
- Obach, R. S., Baxter, J. G., Liston, T. E., Silber, B. M., Jones, B. C., MacIntyre, F., Rance, D. J., Wastall, P. (1997) The prediction of human pharmacokinetic parameters from preclinical and in vitro metabolism data. *J. Pharmacol. Exp. Ther.* **283**: 46–58
- Oie, S., Tozer, T. N. (1979) Effect of altered plasma protein binding on apparent volume of distribution. *J. Pharm. Sci.* **68**: 1203–1205
- Palm, K., Stenberg, P., Luthman, K., Arturson, P. (1997) Polar molecular surface properties predict the intestinal absorption of drugs in humans. *Pharm. Res.* **14**: 568–571
- Perry, P. J. (2002) Psychotropic agents. In: Schoenwald, R. D. (ed.) *Pharmacokinetics in drug discovery and development*. CRC Press, Boca Raton, FL, pp 175–197
- Potter, W. Z., Hollister, L. E. (2001) Antidepressant agents. In: Katzung, B. G. (ed.) *Basic & clinical pharmacology*. 8th edn, Lang Medical Books/McGraw-Hill, New York, p. 502
- Raaflaub, V. J., Speiser-Courvoisier, J. (1974) Zur pharmakokinetik von bromazepam beim menschen. *Arzneimittelforschung* **24**: 1841–1844
- Raevsky, O. A., Schaper, K. J., Artursson, P., McFarland, J. W. (2002) A novel approach for prediction of intestinal absorption of drugs in humans based on hydrogen bond descriptors and structural similarity. *Quant. Struct.-Act. Rel.* **20**: 402–413
- Ritschel, W. A., Hammer, G. V. (1980) Prediction of the volume of distribution from in vitro data and use for estimating the absolute extent of absorption. *Int. J. Clin. Pharmacol. Ther. Toxicol.* **18**: 298–316
- Ritschel, W. A., Akileswaran, R., Hussain, A. S. (1995) Application of neural networks for the prediction of human pharmacokinetic parameters. *Methods Find. Exp. Clin. Pharmacol.* **17**: 629–643
- Rowland, M., Tozer, T. N. (1995) *Clinical pharmacokinetics. Concepts and applications*. 3rd edn, Lippincott, Williams and Wilkins, Philadelphia, pp 143–155
- Schoerlin, M., Horber, F. F., Ferry, F. J., Mayersohn, M. (1990) Disposition kinetics of moclobemide, a new MAO-A inhibitor in subjects with impaired renal function. *J. Clin. Pharmacol.* **30**: 272–284
- Shargel, L., Yu, A. B. C. (1999) *Applied biopharmaceutics and pharmacokinetics*. Appleton & Lange, Stamford, pp 267–269
- Sonne, J., Loft, S., Dosing, M., Vollmerlarsen, A., Olesen, K. L., Victor, M., Andreassen, F., Andreassen, P. B. (1988) Bioavailability and pharmacokinetics of oxazepam. *Eur. J. Clin. Pharmacol.* **35**: 385–389
- Tantishaiyakul, V. (2001) Prediction of Caco-2 cell permeability using partial least squares multivariate analysis. *Pharmazie* **56**: 407–411
- Thummel, K. E., Shen, D. D. (2001) Design and optimization of dosage regimens: pharmacokinetic data. In: Hardman, J. G., Limbird, L. E., Gillman, A. G. (eds) *Goodman & Gillman's the pharmacological basis of therapeutics*. 10th edn, McGraw-Hill, New York, pp 1917–2024
- Turner, J. V., Maddalena, D. J., Cutler, D. J., Agatonovic-Kustrin, S. (2003) Multiple pharmacokinetic parameter prediction for a series of cephalosporins. *J. Pharm. Sci.* **92**: 552–559
- Van de Waterbeemd, H. (2002) High-throughput and in silico techniques in drug metabolism and pharmacokinetics. *Curr. Opin. Drug Dis. Dev.* **5**: 33–43
- Van de Waterbeemd, H., Smith, D. A., Jones, B. C. (2001) Lipophilicity in PK design: methyl, ethyl, futile. *J. Comput. Aided Mol. Des.* **15**: 273–286
- Van der Graaf, P. H., Nilsson, J., Van Schaick, E. A., Danhof, M. (1999) Multivariate quantitative structure-pharmacokinetic relationships (QSPKR) analysis of adenosine A1 receptor agonists in rat. *J. Pharm. Sci.* **88**: 306–312
- Wilkinson, G. R. (2001) Pharmacokinetics: the dynamics of drug absorption, distribution, and elimination. In: Hardman, J. G., Limbird, L. E., Gillman, A. G. (eds) *Goodman & Gillman's the pharmacological basis of therapeutics*. 10th edn, McGraw-Hill, New York, pp 20–22
- Zhao, Y. H., Abraham, M. H., Le, J., Hersey, A., Luscombe, C. N., Beck, G., Sherborne, B., Cooper, I. (2002) Rate-limited steps of human oral absorption and QSAR studies. *Pharm. Res.* **19**: 1446–1457