A Comparison of Calculated and Experimental Parameters as Sources of Structural Information: The Case of Lipophilicity-Related Descriptors

Giulia Caron* and Giuseppe Ermondi

Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via P. Giuria 9, I-10125 Torino, Italy

Abstract: This review is organized in three parts: firstly there is a general overview of recent developments in lipophilicity written to induce medicinal chemists to question what they want to obtain from this kind of study; secondly, the state-of-the-art of experimental and computational determination of log P is briefly reviewed; finally, some applications are discussed to illustrate how much information can be extracted from lipophilicity, and to highlight the difficulty of obtaining a reliable, general method to work with.

Keywords: Log P, log D, partitioning system, CLOGP, distribution profile, folding.

1. GENERAL CONSIDERATIONS

The comparison of calculation results (how things should be) with experimental data (how things are) is a very general procedure. For instance it is well known that the goal of linear and nonlinear regression is to fit a model to data to find the best-fit values of the variables in the model (rate constants, affinities, receptor number, etc.). Some programs automatically fit the data to hundreds of equations and then present the equation(s) that fit the data best. The problem is that the program has no understanding of the scientific context of the experiment.

Analogously, a number of in silico (see article by Petrauskas et al. in this issue) and experimental tools are now available to obtain log P data, but again, sophisticated software and automated analytical instruments should be directed by a people who know the scientific background and the information that can be obtained. In this paper we hope to be able to show to medicinal chemists how much structural information is encoded in lipophilicity descriptors, and to describe some simple strategies to obtain the best results.

1.1. Reasons for Estimating Lipophilicity

The important role played by lipophilicity and the related descriptors in governing pharmacokinetic (see article by Lombardo *et al.* in this issue) and pharmacodynamic events has been extensively underlined in recent years [1-4]. The estimation of lipophilicity can be performed at different levels according to how the data will be used:

- The estimation of log P to choose or discard structures in the context of *in silico* screening, for example to test against Lipinski's rule-of-five [5]; in this case accurate log P measurements are not necessary, and calculated lipophilicity values will suffice,
- The correct measurement of lipophilicity for a few promising drug candidates; in this case it is easier to perform measurements,
- The understanding of as much as possible of what determines the final parameter values and their relationship with structural features of molecules; in this case a good strategy consists in comparing experimental with calculated descriptors and verifying eventual findings with the help of common molecular modeling tools.

*Address correspondence to this author at the Dipartimento di Scienza e Tecnologia del Farmaco, Università di Torino, Via P. Giuria 9, I-10125 Torino, Italy; Fax 0039 011 670 7687; Email: giulia.caron@unito.it

1.2. Neutral and Ionised Species: Which Contribute to Partitioning?

Many drug molecules contain one or more ionizable groups. It has been recently reported that of 51596 compounds listed in the World Drug Index, 32437 contain ionizable groups. Of these, 14.5% are acids, 67.5% are bases and 14.6% are ampholytes [6]. Most drugs are partly or largely ionized at physiological pH, and ions are much more polar than neutral compounds, due to their positive or negative charges. Correspondingly, the degree of dissociation and protonation has a significant influence on the lipophilicity of acids and bases. As a consequence, two series of lipophilicity parameters can be determined: the partition coefficients expressed as log P, which are valid for a single electrical species, to be specified (log P^N for neutral forms and log P^I for monoions), and the distribution coefficients expressed as log D^{pH}, which are pH dependent for ionizable solutes and result from the weighted contributions of all electrical forms present at this pH, as illustrated by Eq. 1

$$\log D = \log [f^N \bullet P^N + \Sigma(f^I \bullet P^I)] \qquad \text{Eq. 1}$$

where f^N and f^I are the respective molar fractions of the neutral and ionized forms. Eq. 1 includes all species present at a given pH.

In view of the difficulty of obtaining values for the lipophilic contributions of all species, it may be asked whether it is possible to neglect the contribution of any species. The relevance of partitioning of monocations and monoanions has been underlined in several papers [7,6,8-10], and in most situations has to be considered. To neglect the partitioning of the ionized species to log D, two conditions must be verified: a) the compound must be not too lipophilic, i.e. log $P^N < 1$ and b) the difference pH - pK_a must be less than 2. If a) and b) are verified then Eq. 1 can be simplified in Eq. 2

1389-5575/03 $41.00+.00

$$D = f_N \bullet P^N \quad \text{Eq. 2}$$

The lipophilicity of dications in *n*-octanol/water is more or less 6 (3 for each additional charge) log P units lower than the lipophilicity of the corresponding neutral species, as demonstrated by dibasic hydroxyzine in [11]. In other isotropic systems the difference may be even larger than 6 [7]. In the light of these observations, the contribution to molecular lipophilicity of dianions and dications and therefore of multi-ionized species (except for ampholytes, see below) can be neglected, except in anisotropic systems where the mechanisms of interaction have been not yet completely rationalized [12,13].

Lipophilicity of ampholytes and zwitterions has been reviewed [14], and some have been found to have a relatively high log P_{oct}. These findings together with other reported data [15-18] underline the necessity of taking into account the contribution of ampholytic and zwitterionic species in lipophilicity determinations.

1.3. Partitioning Systems: Which are the More Relevant?

In recent years lipophilicity has been determined both in isotropic systems (two phases – water plus organic solvent) and in anisotropic systems (two phases – water plus mono- or bilayer structure). Among these, the octanol/water system remains the reference and thus should always be determined at least as a comparison. Different isotropic solvents, such as *n*-octanol-water, alkane-water, chloroform-water and dibutylether-water express the components of lipophilicity (hydrophobic forces, dipolarity/polarizability, and hydrogen bonding) to different degrees [19,20]. To overcome experimental problems caused by the low alkane solubility of many compounds, the use of 1,2-dichloroethane/water (1,2-dce) has recently been proposed to replace the alkane–water system and to characterize H-bonding properties of solutes [21].

While the goals of the study should dictate how many systems are investigated, it should be kept in mind that the more data is available, the more time will be required to evaluate it. A good compromise is to determine lipophilicity in two very different isotropic systems. For instance, the tendency of solutes to form internal H-bonds is usually comparable in octanol and in water while non polar solvents (e.g., 1,2-dichloroethane) strongly favor internal H-bonds [22].

Anisotropic systems express the same forces as isotropic systems (hydrophobic forces, dipolarity/polarizability, and hydrogen bonding) but a supplementary, not yet completely rationalized, electrostatic contribution has also to be considered [12,23-26].

2. MEASURING LIPOPHILICITY

Methods for measuring lipophilicity descriptors have been widely reviewed [9,27]. Experimental data today are much more precise than before, mainly because of automation. Notably, the potentiometric titration method for measuring lipophilicity of ionizable molecules has become widely adopted as a reference method [6,8,28-30]. This technique gives information about both the ionization and the partition behavior of a molecule; moreover, this is accompanied by a good deal of detail in the calculation and interpretation stages. If a sample is soluble and well behaved, it is possible to determine all its pK_a values, its log Ps (for neutral and ionized species) and the whole distribution profile (log D *vs* pH plot).

Though accurate, the classic manual shake-flask method of measuring lipophilicity may be very time-consuming. Several examples are reported in literature in which authors using the shake flask method find values for the same compound that vary by as much as 1 log unit (see ref [31]). It must also be underlined that some solutes of current interest are more difficult to measure than the standard compounds, such as marketed drugs; if solutes are very hydrophilic or very lipophilic, they will require water-solvent ratios of 1,000 to 1 or greater for experimental determination. Finally, while it may be possible to measure log P values > 6 in octanol-water for extremely lipophilic solutes, they may not be included in some QSAR/QsPkR models where lipophilicity descriptors have been used successfully (e.g., bioaccumulation), as models are believed to break down above log P_{oct} of 6 (Leo, personal communication).

3. CALCULATING LIPOPHILICITY

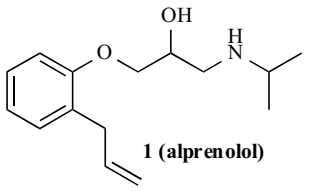
Recently there have been a number of attempts at predicting log P using different algorithms (see article by Petrauskas *et al.* in this issue). Most of them have been developed to predict the lipophilicity of neutral species in the octanol/water system (log P^N_{oct}). Methods to calculate log P can be classified according to various criteria; in particular classifications based either on the methodology applied [32-34] or the number of molecular dimensions considered [22] are the most adopted.

It could be argued that a calculated log P result is reliable if the calculated value is about 0.5 units lower or higher than the experimental value. In the absence of an experimental value, users should remember to check their calculations with similar known molecules with measured log P values and/or to compare results arising from various calculation methods (see below).

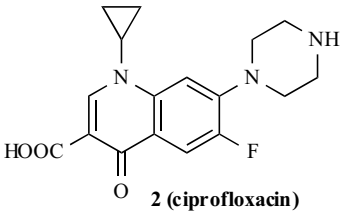
3.1. Calculating the Lipophilicity of Neutral Species

3.1.1. Calculation Referring to n-Octanol/Water System

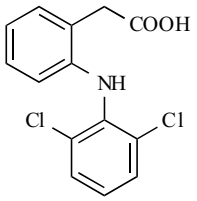
To calculate log P in the absence of experimental values, well-designed software tools based on different methodologies should be used, and results should be compared. By connecting to Tetko's web site (http://146.107.217.178/lab/alogps/index.html, it is possible from a SMILES input [35] to calculate a number of log Ps (ALOGP [36], IA logP (details in www.logp.com), CLOGP [37], LOGKOW [38], XLOGP [39]), check calculations details, look for availability of experimental data, and to include log P values calculated by other software. Further details of this approach are published [40], and a detailed explanation of the method is provided in the article by Tetko *et al.* in this issue. For a single compound calculated data dispersion can be checked by histograms. Two opposite situations can happen; they are illustrated using the chemical structures drawn in Fig. (**1**).



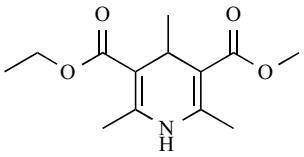
1 (alprenolol)



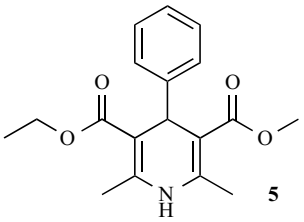
2 (ciprofloxacin)



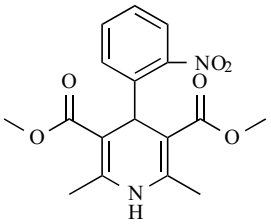
3 (diclofenac)



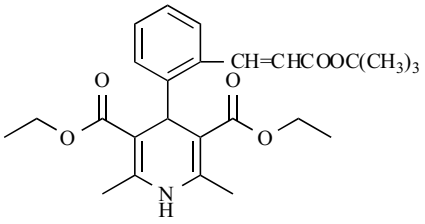
4



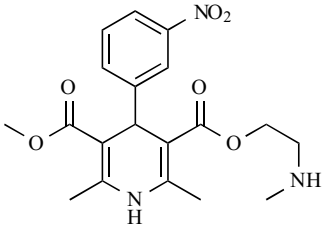
5



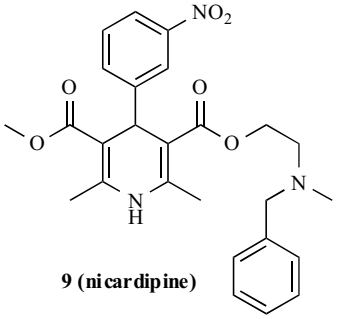
6 (nifedipine)



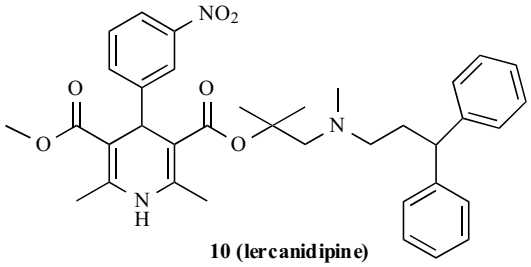
7 (lacidipine)



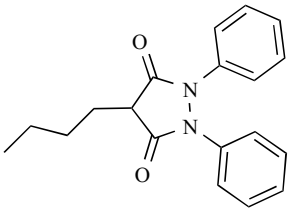
8



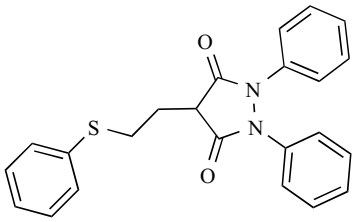
9 (nicardipine)



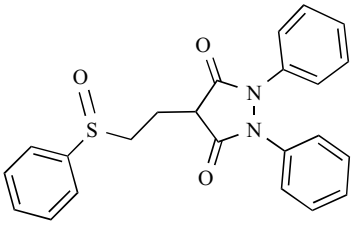
10 (lercanidipine)



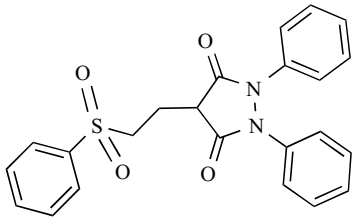
11 (phenylbutazone)



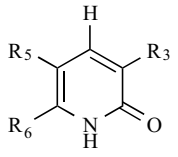
12



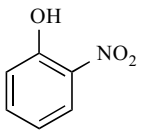
13 (sulfinpyrazone)



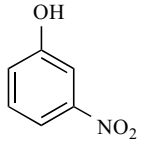
14



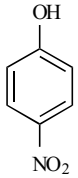
15



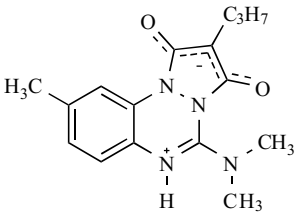
16



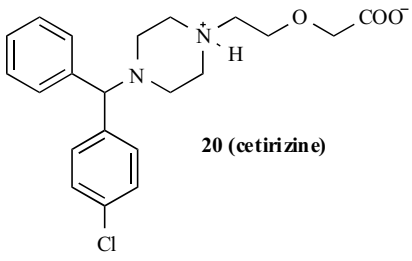
17



18



19 (azapropazone)



20 (cetirizine)

(Fig. 1). contd.....

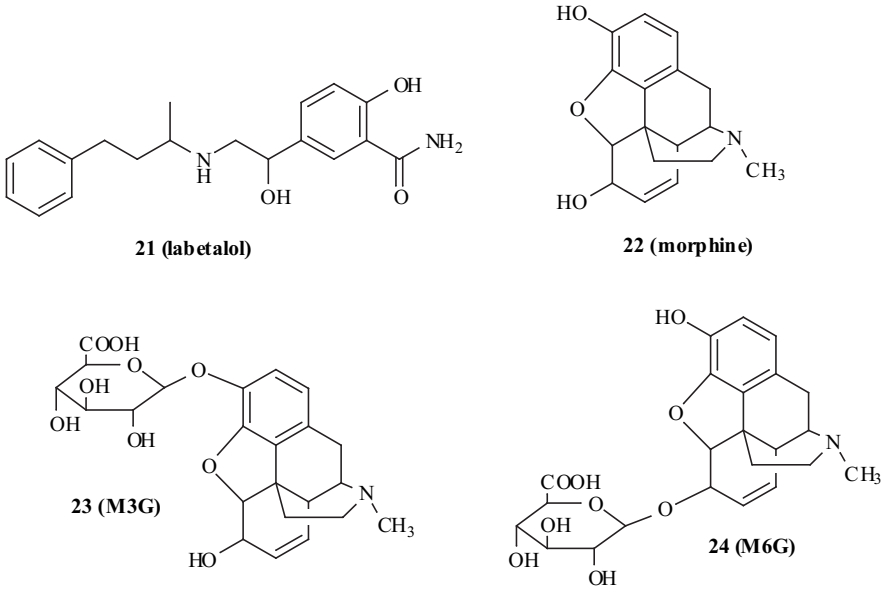


Fig. (1). Chemical structures of investigated compounds.

If there is a limited dispersion of log P data (alprenolol, histogram on the left in Fig. (**2**): all tested methods give very similar log Ps), it can be supposed that the simulation is acceptable; alternatively in the presence of a wide dispersion (ciprofloxacin, histogram on the right in Fig. (**2**): calculated log Ps vary considerably according to the selected method), a good understanding (i.e., knowledge of the theoretical background and advantages and limits of the software, together with careful output reading, especially in the case of warning messages) of at least one of the methods (e.g., CLOGP) is necessary to understand the source of errors. In situation B, measurement of log P is recommended if compound is available

A major problem in interpreting log P data [32] is that software users do not generally know which log P values serve as training sets to develop the various algorithms. It should be noted that:

- There can be inter-laboratory variations in log P measurements;
- The accuracy of prediction for very lipophilic or hydrophilic compounds strongly depend on the number, distribution and quality of data at the extreme conditions;
- Uncommon structures (i.e., platinum drugs [41], disulfides [42], see article by Maiocchi in this issue) are poorly represented, which leads to bias in statistics. In some more recent software (i.e., ALOGP, see Tetko's paper), users can improve the accuracy of log P prediction by training the system with their own good quality experimental pK_a, log P or log D data.

Another problem arising from calculation concerns interaction factors [33]. In particular, while short range interactions (inductive, resonance, alpha effects …) are more

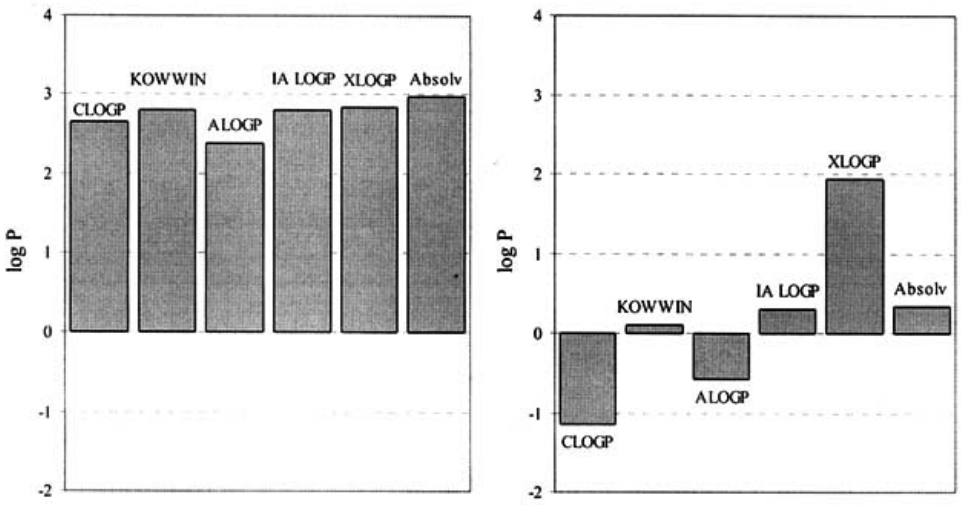


Fig. (2). Comparison between calculated log Ps_{oct}: alprenolol on the left and ciprofloxacin on the right. For further details see text.

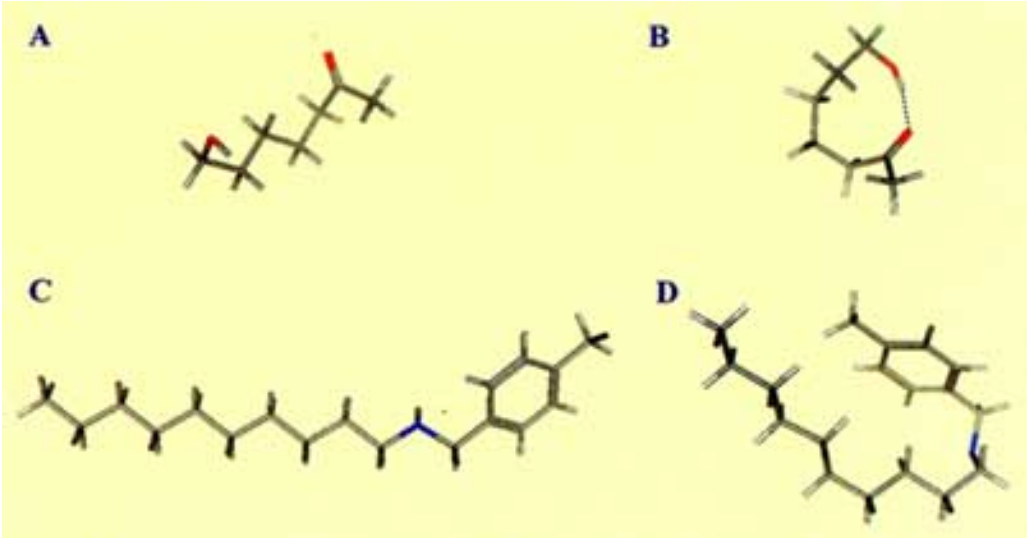


Fig. (3). Examples of hydrophilic folding (upper plate) and hydrophobic collapse (lower plate): A and B represent two different conformers for 7-hydroxy-heptan-2-one but only for B an hydrophilic folding occurs; analogously C and D represent two different conformers of decyl-(4-methyl-benzyl)-amine and only in D an hydrophilic folding is present.

or less fixed by several softwares, long-range interactions governed by molecular flexibility are far from being completely rationalized [37] and thus accounted for in a computer algorithm.

Solving these problems requires a comparison between calculated and experimental data (see below), often combined with indications arising from molecular potentials. Interesting indications can also arise from algorithm outputs, especially CLOGP warnings. Because of its impact on drug disposition and activity [1-3], a major task of log P calculation is currently to predict when a partitioning (lipophilic and hydrophilic) environment encourages flexible compounds to adopt a preferred conformation. In general, flexible compounds with suitable moieties may exhibit hydrophobic collapse (conformational change by which a solute maximizes the superposition of hydrophobic interactions) in polar solvents, and hydrophilic folding (conformational change by which a solute maximizes both the number and the strength of internal electrostatic interactions, mainly H-bonds [22]) in non-polar environments (Fig. (**3**)).

Hydrophobic folding is partially included in the CLOGP algorithm by the FRAGBRANCH correction, which is based on the idea that when a fragment is at the centre of three alkane chains, the chains will limit the amount of surface to be solvated. Some structures that appear to encourage hydrophobic surfaces to overlap one another are not covered by the definition of FRAGBRANCH in the current version of CLOGP; CLOGP gives a warning message when it identifies the possibility of such behavior. It should be noted that in the output of CLOGP, the pictures of compounds are colored to give information about the electronically active components of the molecule. Isolating carbons are in cyan, while polar fragments are in yellow. In this way, even if the 3D structure is not displayed, the operator can at least check whether some collapse is possible, and eventually go on with a molecular dynamics simulation.

Finally log P values can be extrapolated from experimental log D data. Note, however, that the accuracy of the result depends on the accuracy of the pK_a value/s used. Moreover the extrapolation is sometimes not correctly performed; Fig. (**4**) shows the errors made in the estimation of log P^N of diclofenac (chemical structure in Fig. (**1**)) using various log D values when the partitioning of the ionized form is taken into account or not (Eq. 3 and 4, respectively).

$$P^N = D^{pH} \bullet \left(1 + \frac{K_a}{[H^+]}\right) - P^I \bullet \frac{K_a}{[H^+]} \qquad \text{Eq. 3}$$

$$P^N = D^{pH} \bullet \left(1 + \frac{K_a}{[H^+]}\right) \qquad \text{Eq. 4}$$

3.1.2. Calculations Referring to Other Isotropic and Anisotropic Systems

While many methods have been developed for the prediction of log P_{oct}, less attention has been directed towards prediction of other partitioning systems, despite their relevance to gain structural information on compounds.

To date, it is possible to calculate log P_{alk} by the Rekker approach [43] and to calculate a number of log Ps by Absolv [44,45] (Sirius Analytical Instruments Ltd, Forest Row, East Sussex, UK). Absolv is based on the factorization of log P by solvatochromic equations (whose general form is in Eq. 3) [46]

$$\log P = v \bullet V + p \bullet \pi^* + b \bullet \beta + a \bullet \alpha + c \qquad \text{Eq. 5}$$

In this equation, log P is the logarithm of a partition coefficient determined in a given system, V is the steric parameter (namely the molar volume to assess the solute's capacity to elicit non-polar interactions, such as hydrophobic bonds, and to some extent dispersive forces), π^* is a measure of the solute's dipolarity/polarizability (orientation

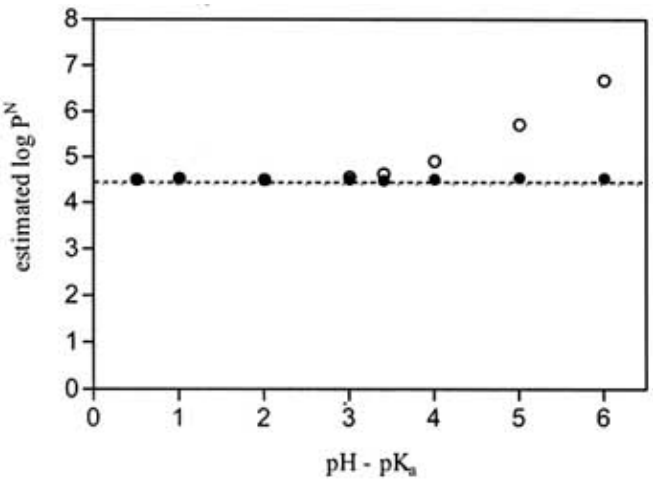


Fig. (4). Prediction of log P^N for monoacid diclofenac: full circles represent data obtained by Eq. 3, empty circles are data obtained by Eq. 4. Experimental log P^N (4.51 [51]) is represented by the dotted line. The following experimental values have been used: pK_a = 3.99 [51]; log P^I = 0.68 [51] and twelve couples of pH/log D data (2.0/4.51; 3.0/4.46; 4.5/3.87; 5.0/3.48; 6.0/2.49; 7.0/1.55; 7.4/1.22; 8.0/0.90; 9.0/0.71; 10.0/0.68; 10.9/0.68).

and induction forces); α and β are the solute's H-bond donor acidity and H-bond acceptor basicity, respectively; and v, p, b, a are the regression coefficients and c is a constant. Abraham has collected more that 20 solvatochromic equations for as many log Ps [47].

Solvatochromic descriptors (V, π*, α and β) are general, and are themselves composed of group contributions [45,46]. With one calculation of solvatochromic descriptors, Absolv [48] can predict a wide range of partition processes. The main limitation of the method is represented by the lower number of high quality experimental log P values used to train the software for some solvent partitions, which do not allow all solvatochromic equations to have the same statistical significance.

Absolv can also calculate log P in anisotropic systems, such as micelles/water and liposomes water. As pointed out above, however lipophilicity in anisotropic media has a great interest for ionized species, but is less interesting for neutral compounds.

3.2. Calculating Lipophilicity of Ionized Species

A direct calculation of the lipophilicity of ionized species (log P^I) can be obtained by Rekker approach [43] starting from the experimental value of a related structure. Alternatively log P^I can be estimated either from log P^N or from log D^{pH} provided that accurate pK_a values are known.

Eq. 6 can be used to obtain log P^I from log P^N

$$\log P^I = \log P^N - K \qquad \text{Eq. 6}$$

where K is a constant depending on the partitioning system and the compound's characteristics. Suggested values for K are 3 for acids in the octanol/water system and 3.5 for bases

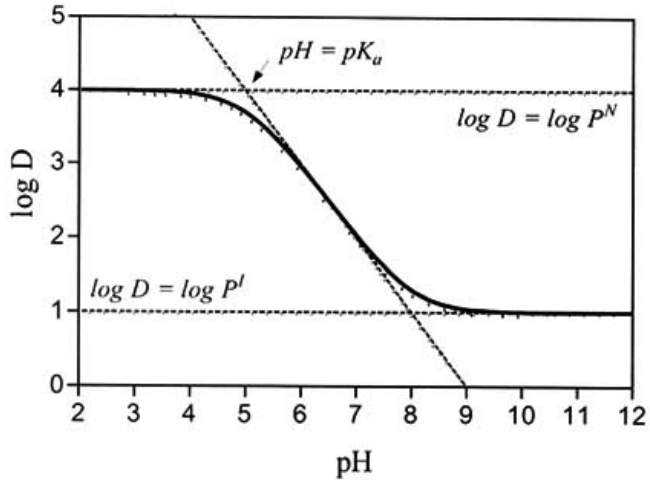


Fig. (5). Lipophilicity profile for an acid with pK_a = 5, log P^N = 4 and log P^I = 1.

[49,50]. A more detailed classification has been recently proposed for the 1,2-dichloroethane system [7].

Sometimes experimental data determined at a given pH is available (e.g., log D^{pH}, generally taken from literature). If the pK_a is known, it is possible to check whether log D = log P^I, and thus directly to obtain log P^I, or to apply Eq. 4 if log D = log P^N.

Finally there is a comment on anisotropic systems. For neutral compounds, partition coefficients in isotropic and anisotropic media are known to be comparable [26,51]. In contrast, charged species and particularly cations are known to partition better into anisotropic lipid membranes than into bulk *n*-octanol [50,52,53]. The mechanisms governing this behavior have not been unraveled, and thus to date no computational method is available to predict the anisotropic lipophilicity of ionized species.

3.3. Obtaining Lipophilicity Profiles

The lipophilicity profile is the plot of log D vs pH as obtained by applying Eq. 1 [8,29]. An example of this plot is shown in Fig. (**5**) for a generic acid HA; it represents the complete lipophilicity behavior of a substance in a given system [6].

Some software (e.g., ACD/Log*D* Suite) enables log D to be calculated at any pH if log P values are known. A similar service can be performed on-line at http://www.raell.demon.co.uk/chem/calcs/index.htm.

3.4. Computational Resources and Operator Skills Required to Perform Lipophilicity Calculations

Most softwares available today to estimate lipophilicity run under a Windows environment, and some programs are also available as on-line WWW versions (see article by Tetko in this issue). To perform calculations, either SMILES [35], traditional input formats, such as mol file, or the recently developed JAVA editor (i.e., JME editor of Peter Ertl [54]) are required. In other words everyone can calculate log P, though to make sense of it all, they need some "lipophilic" knowledge. As the knowledge level grows, a supplementary skill in informatics is necessary to combine these data with molecular modeling tools, most of which run under Linux systems.

4. HOW TO COMBINE CALCULATED AND EXPERIMENTAL PARAMETERS

4.1. The use of Calculated Data Before Experimental Determination: Setting-Up of Experiments

Independent of the method used, an important step in the experimental determination of log P and/or log D is the choice of the solvent/water ratio [55]. This ratio depends on the lipophilicity of the molecule in that solvent/water system, and thus a good predicted value allows the experimental design to be optimized.

4.2. The Use of Calculated Data After Experimental Determination: Identification of the Best Predictive Methods for a Structurally-Related Series of Compounds

If experimental values are available, it is possible to verify which computational tools work better by comparing their output with experimental values (log P* or recently published values), and thus to choose the optimal theoretical method for predicting lipophilicity of related compounds.

In Table 1, experimental and computational log P data for a number of 1,4-dihydropyridine (1,4-DHP) calcium channel antagonists [56] have been compiled.

Compounds **4** and **5** are simple models, and thus it is easier to check with them which calculation works better. It is not surprising for **4** that, of the programs tested CLOGP provides the closest fit between calculated and observed log P values, because the log P* value was determined by the same person who developed the rules for CLOGP. Interestingly for **5** there is a difference of about 0.5 between experimental log P (obtained in Pomona labs) and CLOGP;

Table 1. Lipophilicity Data in *n*-Octanol/Water for a Series of 1,4-DHPs

Compound	exp [a]	CLOGP [b]	ALOGP [b]	IAlogP [b]	XLOGP [b]	KOWWIN [b]
4	3.27	3.23	2.08	1.28	1.66	1.96
5	3.74	4.27	2.84	2.36	2.90	3.17
6 (nifedipine)	2.86	3.41	2.32	2.06	2.37	2.5
7 (lacidipine)	5.56	6.21	4.76	4.36	4.84	5.39
8	2.86 [c]	3.06	1.90	1.21	1.99	1.98
9 (nicardipine)	4.65 [d]	5.51	4.34	3.53	3.94	3.9
10 (lercanidipine)	6.0 [e]	8.28	6.42	6.58	6.82	6.88

a) Experimental value from Pomona database (log P star), otherwise indicated

b) Calculated data obtained from Tetko's website (see text)

c) Taken from [68]

d) Taken from [69]

e) Taken from [70]

other software gives less correct results. According to these observations, it is reasonable to assume that CLOGP is the better tool to use in predicting log P of 1,4-DHPs, but it should be noted in mind that an overestimation of experimental log P of about 0.5 is expected. This is confirmed by results for nifedipine, lacidipine, **8**, and nicardipine; the range of variation is 0.2-0.9, the most deviant compounds being the bases **8** and **9,** for which the influence of pK_a is more relevant. Interestingly for lercanidipine an extremely large (> 2) overprediction of experimental data is noted. A careful examination of CLOGP output for lercanidipine indicates the presence of a warning message "very high log P unrealistic in nature". According to Leo, this means that the calculation itself is correct, but for many biological models (QSAR) where log P has been used successfully, the model breaks down above log P_{oct} of 6 (personal communication, cited above). Does lercanidipine present a peculiar lipophilic behavior due to its molecular structure (hydrophobic collapse?), or is the experimental data doubtful because of its high value?. Studies are way in our laboratory to rationalize this finding.

4.3. The Use of Calculated Data After Experimental Determination: a Source of Structural Information

As previously mentioned the combined use of calculated and experimental parameters allows researchers to understand as much as possible of what determines the lipophilicity parameters. The term *diff*(log $P^{exp-calc}$) is the difference between an experimental log P and the corresponding value calculated by a favorite or best known method [9]. A standard range is $-0.5 < diff(\log P^{exp-calc}) < +0.5$

The *diff*(log $P^{exp - calc}$) parameter has been used to bring some tautomeric equilibria to light. If a structure can exist in two tautomeric forms, the equilibrium is often quite different in water than in wet octanol. Of course the partitioning equilibrium constant necessarily depends upon the tautomeric ratio in two phases and this is not easily predicted. In the case of keto/enol tautomers, it is the keto form which is preferred by water, because it has the higher H-bond acceptor strength β. Conversely, it is the enol form which is preferred by octanol [57].

In the case of sulfinpyrazone **13** [58] and its metabolites (**12** and **14** in Fig. (**1**)), for which diketo/keto-enol tautomerism exists in the neutral form of compounds, it has been proposed after analyzing the *diff*(log $P^{exp-calc}$) that the lipophilicity is affected by tautomeric and conformational equilibria. In fact, the log P values calculated by Rekker's method are based on the experimental log P of phenylbutazone **11**, where the percentage of enolic and ketonic forms in water is respectively 1.8% and 98.2% [59], whereas the compound is mostly diketonic in DMSO [60]. In other words, the calculated values mostly neglect a contribution of the more lipophilic enolic tautomer. To explore the diketo/keto-enol equilibrium ^{13}C-NMR spectroscopy was used, and it has been verified that the diketo/keto-enol ratio of the sulfide is about 3/1 and larger than that of the sulfoxide (about 1.5/1) and the sulfone (about 1/1). This difference could explain the smaller influence of the tautomeric equilibrium on the lipophilicity of sulfide and the more accurate calculation of its log P compared to oxygenated compounds.

Another application of *diff*(log $P^{exp-calc}$) to tautomerism concerns lipophilicity of pyridine-2(1H)-one cardiotonic agents (**15** in Fig. (**1**)) investigated by Altomare *et al.* in [61]. As expected, CLOGP calculations assigned an higher lipophilicity to the 'hydroxy' compared to the 'oxo' form. Experimental values were not significantly different from those calculated for 'oxo' forms ($-0.5 < diff(\log P^{exp-calc}) < 0.5$) except for derivatives bearing electron withdrawing substituents in position 6 (R_6) ($diff(\log P^{exp-calc}) > 0.5$), for which there is therefore a different keto/enol ratio compared to other derivatives.

Insights into the ionization (and thus partitioning) behavior of ampholytes can be obtained by using *diff*(log $P^{exp-calc}$). Ampholytes can be classified into two types [14]: ordinary ampholytes and zwitterionic ampholytes. Ordinary ampholytes exist in neutral or singly charged form while zwitterionic ampholytes can equilibrate between zwitterionic and neutral forms. The ratio of these species (and thus the ionization processes in terms of microconstants) is governed by K_Z; this term is a constant for a given solvent, but in the two phases will probably differ. Therefore, a question that arises when studying zwitterions is: what is the significance of the calculated value? In our opinion the answer is clear only for CLOGP, which adds (ZW-) or (ZW+) after the name of each fragment, which can participate in zwitterion formation, even when only the predominance of zwitterionic species is evident, as denoted by a large K_Z value. Some examples (azapropazone, cetirizine and labetalol, see Fig. (**1**)) are shown in Table 2.

Table 2. Lipophilicity Data in *n*-Octanol/Water and in 1,2-Dicloroethane/Water Systems for Neutral Species of Nitrophenols

Compound	log P^N_{oct}		log P^N_{dce}	
	exp [a]	Absolv [b]	exp [a]	Absolv [b]
16 (*o*-nitrophenol)	1.77	1.51	2.81	1.91
17 (*m*-nitrophenol)	2.00	1.46	0.92	0.35
18 (*p*-nitrophenol)	1.96	1.80	0.72	0.82

a) Experimental values taken from [7]

b) Values calculated by the software Absolv (see text for details)

CLOGP is the only software able to calculate reasonable log P values for all zwitterions and thus to correctly distinguish between zwitterions with large K_Z (azapropazone and cetirizine), and zwitterions with low K_Z. The zwitterionic correction has been applied cetirizine and azapropazone, but not for labetalol. This prediction could fail for unknown compounds [14], and thus it is advised to compare experimental log D at isoelectric pH with CLOGP with and without zwitterionic correction. In the absence of any fragmental errors the *diff*(log $P^{exp-calc}$) values will give some insights into K_Z values.

Intramolecular effects can be studied by the analysis of *diff*(log $P^{exp-calc}$) obtained by Absolv. As previously mentioned, with one calculation of solvatochromic descriptors Absolv can predict a number of log P values. As

Table 3. Lipophilicity Data in *n*-Octanol/Water for Selected Zwitterionic Compounds

Compound	exp [a]	CLOGP [b]	ALOGP [b]	IAlogP [b]	XLOGP [b]	KOWWIN [b]
19 (cetirizine)	1.50 [e]	2.08	2.80	3.37	2.80	-0.61
20 (azapropazone)	1.78 [d]	1.78	1.47	1.65	1.98	-0.03
21 (labetalol)	2.60 [c]	2.50	1.73	1.12	2.52	2.41

a) Experimental value taken from indicated references

b) Calculated data obtained from Tetko's website (see text)

c) Taken from [14] and referring to neutral species (zwitterion with small K_Z)

d) Taken from [60] and referring to zwitterionic species (zwitterion with large K_Z)

e) Taken from [11] and referring to zwitterionic species (zwitterion with large K_Z)

equations in the same form are used for all predictions, a comparison of data can give interesting information. An excellent example is represented by some nitrophenols (Fig. (**1**)). Having a reasonable log P_{oct} prediction (Table 3), failure in predicting log P_{dce} can be only explained by the presence of intramolecular effects predominating in 1,2-dce and not in *n*-octanol. This is in keeping with results reported in [62].

Sometimes comparison between experimental and calculated data should be made indirectly. The lipophilicity behavior of morphine and its glucuronides has been investigated by calculation and by potentiometry [63-65] to shed light on their peculiar pharmacokinetic behavior. In fact morphine-6-glucuronide (but not morphine 3-glucuronide) is a highly potent opiate receptor agonist, even though glucuronides are polar metabolites that are generally considered to be unable to cross the blood-brain barrier. The *in silico* prediction of lipophilicity has been performed for the three compounds using the molecular lipophilicity potential (MLP) by Gaillard *et al.* [63,66,67]. Findings indicate that morphine 6-glucuronide, and to a lesser extent morphine 3-glucuronide, are far more lipophilic than predicted, and in fact not much less lipophilic than morphine itself. This lipophilicity order (Table 4) has been confirmed by potentiometric determination even if there is some discrepancy between absolute values, probably due to MLP parametrization.

Table 4. Lipophilicity of Morphine and its Glucuronides

Compound	exp [a]	log P^{min} [b]	log P^{MAX} [c]
22 (morphine)	0.89	1.0	1.4
23 (M3G)	-1.10	-2.2	-1.4
24 (M6G)	-0.76	-1.6	-1.0

a) Experimental value for the neutral/zwitterionic species taken from [64]

b) The lowest virtual log P (the MLP allows the calculation of a log P_{oct} value for each distinct 3D-structure, giving access to the theoretical log P (virtual log P) of all conformers).

c) The highest virtual log P

CONCLUSION AND PERSPECTIVES

Among physicochemical properties used today in the very early stages of drug-discovery, lipophilicity has certainly assumed a very important role because of its relevance in governing most stages of drug disposition.

Because of the increasing potency of informatics, a large number of different calculation methods have been derived for estimating octanol/water partition coefficient (log P) of chemical structures, being most of them also available on the Internet (see article by Tetko in this issue).

From an experimental point of view, automation has enabled laboratories to obtain hundreds of data in one week. In addition experimental lipophilicity values are compiled in commercial databases available on the market.

To optimize the process of data evaluation and information extraction, the comparison of experimental results with *in silico* simulations is a very powerful procedure, which can be easily applied in many fields of medicinal chemistry and pharmaceutical sciences.

In particular in the case of lipophilicity descriptors, an almost perfect superimposition of data (the difference between calculated and measured log D is less than ±0.5) indicates that mechanism governing partitioning has been completely rationalised and correctly included in software packages. Conversely, large differences (the difference between calculated and measured log D is more than ±0.5) suggest that the particular partitioning phenomenon has not been completely understood and a supplementary number of measurement often combined with molecular modeling studies are required.

ACKNOWLEDGEMENTS

GC and GE are indebted with a number of people who contributed actively to the preparation of this review: John Comer (Sirius) for his smart comments and also for checking our Italianated English; Silvia Arpicco (University of Turin) for her careful and fruitful reading of the manuscript; and Silvio Aime (University of Turin) for encouragement.

Particular thanks are sent to the QSAR mailing list http://www.accelrys.com/user/pub-mail-lists/qsar/ for the excellent discussion.

Cosmital (Marly, CH) is acknowledged for financial support.

REFERENCES

[1] Testa, B.; Caron, G.; Crivori, P.; Rey, S.; Reist, M.; Carrupt, P. A. *Chimia,* **2000,** *54*, 672-677.

[2] Testa, B.; Crivori, P.; Reist, M.; Carrupt, P. A. *Perspect. Drug Discovery Design,* **2000,** *17,* 179-211.

[3] van de Waterbeemd, H.; Smith, D. A.; Beaumont, K.; Walker, D. K. *J. Med. Chem.,* **2001,** *44*(9), 1313-1333.

[4] Folkers, G.; Merz, A. Hydrophobic fields in quantitative structure-activity relationships, in *Lipophilicity in Drug Action and Toxicology*, Pliska, V.; Testa, B.; van de Waterbeemd, H., editors; VCH Publishers: Weinheim, **1996**, pp. 219-232.

[5] Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Deliv. Rev.,* **1997,** *23*, 3-25.

[6] Comer, J. E.; Tam, K. Y. In *Pharmacokinetic Optimization in Drug Research*, Testa, B.; van de Waterbeemd, H.; Folkers, G.; Guy, R. H., editors; Wiley-VCH: Zürich, **2001**, pp. 275-304.

[7] Bouchard, G.; Carrupt, P. A.; Testa, B.; Gobry, V.; Girault, H. H. *Chem. Eur. J.,* **2002,** *8*(15), 3478-3484.

[8] Avdeef, A. In *Lipophilicity in Drug Action and Toxicology*, Pliska, V.; Testa, B.; van de Waterbeemd, H., editors; VCH Publishers: Weinheim, **1996**, pp. 109-139.

[9] Caron, G.; Reymond, F.; Carrupt, P. A.; Girault, H. H.; Testa, B. *PSST,* **1999,** *2*(8), 327-335.

[10] Reymond, F.; Gobry, V.; Bouchard, G.; Girault, H. H. In *Pharmacokinetic optimization in drug Research*, Testa, B.; van de Waterbeemd, H.; Folkers, G.; Guy, R. H., editors; Wiley-VCH: Zürich, **2001**, pp. 327-349.

[11] Pagliara, A.; Testa, B.; Carrupt, P. A.; Jolliet, P.; Morin, C.; Morin, D.; Urien, S.; Tillement, J. P.; Rihoux, J. P. *J. Med. Chem.,* **1998,** *41*, 853-863.

[12] Plemper van Balen, G.; a Marca Martinet, C.; Caron, G.; Bouchard, G.; Reist, M.; Carrupt, P. A.; Fruttero, R.; Gasco, A.; Testa, B. *Med. Res. Rev.,* **2003**.

[13] Avdeef, A. *Current Topics in Medicinal Chemistry*, **2001,** *1*, 277-351.

[14] Pagliara, A.; Carrupt, P. A.; Caron, G.; Gaillard, P.; Testa, B. *Chem. Rev.,* **1997,** *97*(8), 3385-3400.

[15] Takacs-Novak, K.; Jozan, M.; Szasz, G. *Int. J. Pharm.,* **1995,** *113*, 47-55.

[16] Takacs-Novak, K.; Jozan, M.; Hermecz, I.; Szasz, G. *Int. J. Pharm.,* **1992,** *79*, 89-96.

[17] Takacs-Novak, K.; Avdeef, A.; Box, K. J.; Podanyi, B.; Szasz, G. *J. Pharm. Biomed. Anal.,* **1994,** *12*, 1369-1377.

[18] Scherrer, R. A. In *Pharmacokinetic Optimization in Drug Research*, Testa, B.; van de Waterbeemd, H.; Folkers, G.; Guy, R. H. editors; Wiley-VCH: Zürich, **2001**; pp. 351-381.

[19] El Tayar, N.; Tsai, R. S.; Testa, B.; Carrupt, P. A.; Leo, A. *J. Pharm. Sci.,* **1991,** *80*, 590-598.

[20] Pagliara, A.; Caron, G.; Lisa, G.; Fan, W.; Gaillard, P.; Carrupt, P. A.; Testa, B.; Abraham, M. H. *J. Chem. Soc. Perkin Trans. 2,* **1997,** 2639-2643.

[21] Steyaert, G.; Lisa, G.; Gaillard, P.; Boss, G.; Reymond, F.; Girault, H. H.; Carrupt, P. A.; Testa, B. *J. Chem. .Soc. Faraday Trans.,* **1997,** *93*, 401-406.

[22] Carrupt, P. A.; Testa, B.; Gaillard, P. In *Reviews in Computational Chemistry*, Boyd, D. B.; Lipkowitz, K. B., editors; VCH Publishers: Weinheim, **1997**; Vol. *11*, pp. 241-315.

[23] Krämer, S. D.; Jakits-Deiser, C.; Wunderli-Allenspach, H. *Pharm. Res.* **1997,** *14*, 827-832.

[24] Betageri, G. V.; Rogers, J. A. *Int. J. Pharm.,* **1988,** *46*, 95-102.

[25] Pauletti, G. M.; Wunderli-Allenspach, H. *Eur. J. Pharm. Sci.,* **1994,** *1*, 273-282.

[26] Taillardat-Bertschinger, A.; a Marca Martinet, C.; Carrupt, P. A.; Reist, M.; Caron, G.; Fruttero, R.; Testa, B. *Pharm. Res.,* **2001,** *19*(6), 729-737.

[27] Avdeef, A.; Testa, B. *CMLS,* **2002,** *59*, 1681-1689.

[28] Avdeef, A. *Quant. Struct.-Act. Relat.,* **1992,** *11*, 510-517.

[29] Avdeef, A.; Kearney, D. L.; Brown, J. A.; Chemotti, A. R. Jr. *Anal. Chem.,* **1982,** *54*, 2322-2326.

[30] Takacs-Novak, K.; Avdeef, A. *J. Pharm. Biomed. Anal.,* **1996,** *14*, 1405-1413.

[31] Medchem95 database. Daylight Chemical Information System, Inc., Irvine, California: **1995**.

[32] Duban, M. E.; Bures, M. G.; DeLazzer, J.; Martin, Y. C. In *Pharmacokinetic Optimization in Drug Research*, Testa, B.; van de Waterbeemd, H.; Folkers, G.; Guy, R. H. editors; Wiley-VCH: Zürich, **2001**, pp. 485-497.

[33] Petrauskas, A.; Kolovanov, E. A. *Perspect. Drug Discovery Design,* **2000,** *19*, 99-116.

[34] Mannhold, R.; Rekker, R. F.; Sonntag, C.; ter Laak, A. M.; Dross, K.; Polymeropoulos, E. E. *J. Pharm. Sci.,* **1995,** *84*, 1410-1418.

[35] Weininger, D. *J. Chem. Inf. Comput. Sci.,* **1988,** *28*, 31-36.

[36] Tetko, I. V.; Tanchuk, V. Y.; Villa, A. E. P. *J. Chem. Inf. Comput. Sci.,* **2001,** *41*, 1407-1421.

[37] Leo, A.; Hoekman, D. *Perspect. Drug Discovery Design,* **2000,** *18*, 19-38.

[38] Meylan, W. M.; Howard, P. H. *J. Pharm. Sci.,* **1995,** *84*, 83-92.

[39] Wang, R.; Fu, Y.; Lai, L. *J. Chem. Inf. Comput. Sci.,* **1997,** *37*, 615-621.

[40] Tetko, I. V.; Tanchuk, V. Y.; Kasheva, T. N.; Villa, A. E. P. *J. Chem. Inf. Comput. Sci.,* **2001,** *41*, 246-252.

[41] Platts, J. A.; Hibbs, D. E.; Hambley, T. W.; Hall, M. D. *J. Med. Chem.,* **2001,** *44*, 472-474.

[42] Hashash, A.; Kirkpatrick, D. L.; Lazo, J. S.; Block, L. H. *J. Pharm. Sci.,* **2002,** *91*(7), 1686-1692.

[43] Rekker, R. F.; Mannhold, R. *Calculation of Drug Lipophilicity;* VCH Publishers: Weinheim, **1992**.

[44] Platts, J. A.; Abraham, M. H.; Butina, D.; Hersey, A. *J. Chem. Inf. Comput. Sci.,* **2000,** *40*, 71-80.

[45] Platts, J. A.; Butina, D.; Abraham, M. H.; Hersey, A. *J. Chem. Inf. Comput. Sci.,* **1999,** *39*(5), 835-845.

[46] Abraham, M. H. *Chem. Soc. Rev.,* **1993,** 73-83.

[47] Abraham, M. H.; Chadha, H. S. In *Lipophilicity in Drug Action and Toxicology*, Pliska, V.; Testa, B.; van de Waterbeemd, H. editors; VCH Publishers: Weinheim, **1996**, pp. 311-337.

[48] Abraham, M. H.; Ibrahim, A.; Zissimos, A. M.; Zhao, Y. H.; Comer, J. E.; Reynolds, D. *DDT,* **2002,** *7*(20), 1056-1063.

[49] Kubinyi, H. *QSAR: Hansch Analysis and Related Approaches;* VCH Publishers: Weinheim, **1993**.

[50] Fruttero, R.; Caron, G.; Fornatto, E.; Boschi, D.; Ermondi, G.; Gasco, A.; Carrupt, P. A.; Testa, B. *Pharm. Res.,* **1998,** *15*(9), 1407-1413.

[51] Avdeef, A.; Box, K. J.; Comer, J. E. A.; Hibbert, C.; Tam, K. Y. *Pharm. Res.,* **1998,** *15*(2), 209-215.

[52] Austin, R. P.; Davis, A. M.; Manners, C. N. *J. Pharm. Sci.,* **1995,** *84*, 1180-1183.

[53] Plemper van Balen, G.; Carrupt, P. A.; Morin, D.; Tillement, J. P.; Le Ridant, A.; Testa, B. *Biochem. Pharmacol.,* **2002,** *63*, 1691-1697.

[54] Ertl, P. *J. Mol. Graphics,* **1998,** *16,* 11-13.

[55] Dearden, J. C.; Bresnen, G. M. *Quant. Struct.-Act. Relat.,* **1988,** *7*, 133-144.

[56] Reappraisal, A. *Drugs,* **1998,** *55*(4), 509-517.

[57] Leo, A. J. *Chem. Rev.,* **1993,** *93*, 1281-1306.

[58] Caron, G.; Gaillard, P.; Carrupt, P. A.; Testa, B. *Helv.Chim.Acta* **1997,** *80*, 449-462.

[59] Stella, V. J.; Pipkin, J. D. *J. Pharm. Sci.,* **1976,** *65*, 1161-1165.

[60] Caron, G.; Pagliara, A.; Carrupt, P. A.; Gaillard, P.; Testa, B. *Helv. Chim. Acta,* **1996,** *79*, 1683-1685.

[61] Altomare, C.; Cellamare, S.; Summo, L.; Fossa, P.; Mosti, L.; Carotti, A. *Bioorg. Med. Chem.,* **2000,** *8*, 909-916.

[62] Chopineaux-Courtois, V.; Reymond, F.; Bouchard, G.; Carrupt, P. A.; Testa, B.; Girault, H. H. *JACS,* **1999,** *121*, 1743-1747.

[63] Gaillard, P.; Carrupt, P. A.; Testa, B. *Bioorg. Med. Chem. Lett.,* **1994,** *4*, 737-742.

[64] Avdeef, A.; Barrett, A.; Shaw, P. N.; Knaggs, R. D.; Davis, S. S. *J. Med. Chem.,* **1996,** *39*, 4377-4381.

[65] Carrupt, P. A.; Testa, B.; Bechalany, A.; El Tayar, N.; Descas, P.; Perrissoud, D. *J. Med. Chem.,* **1991,** *34*, 1272-1275.

[66] Gaillard, P.; Carrupt, P. A.; Testa, B.; Boudon, A. *J. Comput.-Aided Mol. Design,* **1994,** *8*, 83-96.

[67] Carrupt, P. A.; Gaillard, P.; Billois, F.; Weber, P.; Testa, B.; Meyer, C.; Pérez, S. In *Lipophilicity in Drug Action and Toxicology*, Pliska, V.; Testa, B.; van de Waterbeemd, H. editors; VCH Publishers: Weinheim, **1996**, pp. 195-217.

[68] Boschi, D.; Caron, G.; Visentin, S.; Di Stilo, A.; Rolando, B.; Fruttero, R.; Gasco, A. *Pharm. Res.,* **2001,** *18*(7), 987-991.

[69] Franke, U.; Munk, A.; Wiese, M. *J. Pharm. Sci.,* **1999,** *88*(1), 89-95.

[70] Leonardi, A.; Motta, G.; Pennini, R.; Testa, R.; Sironi, G.; Catto, A.; Cerri, A.; Zappa, M.; Bianchi, G.; Nardi, D. *Eur. J. Med. Chem.,* **1998,** *33,* 399-420.