Recent Methodologies for the Estimation of N-Octanol/Water Partition Coefficients and their Use in the Prediction of Membrane Transport Properties of Drugs

Gilles Klopman* and Hao Zhu

Department of Chemistry, Case Western Reserve University, 10900 Euclid Ave, Cleveland, Ohio 44106, USA

Abstract: The lipophilicity of drug molecules (represented as the logarithm of the n-octanol/water partition coefficient) often strongly correlates with their pharmacological and toxic activities. It is therefore, not surprising that there is considerable interest in developing mathematical models capable to accurately predict their value for new drug candidates.

In this review, current major approaches for estimating partition coefficients are described and some of their advantages and disadvantages are discussed. Recent uses of these partition coefficient algorithms in the development of membrane transport models are also discussed.

Keywords: Lipophilicity, partition coefficient, membrane transport, QSAR.

INTRODUCTION

The logarithm of the partition coefficient between n-octanol and water (logP) is often used to represent the lipophilicity of a molecule. In many quantitative structure-activity relationship (QSAR) studies (e.g. biodegradation rate, soil sorption coefficient, bioconcentration factor and many phamacological endpoints), logP has been used as a key parameter to estimate the biological activities of organic compounds. Although the experimental measurement of the logP value is reliable for most compounds, this process may be costly and time-consuming. Furthermore, it is sometimes important to know the lipophilic properties of a compound before it is even synthesized. A computational model, which can give a reliable estimation of the logP value for new compounds is thus important, particularly for molecular modeling and drug design.

Methods for estimating logP can be divided into two classes. The first type of method is classified as "group contribution" approaches, which includes "atom-based" methods and "fragment-based" methods. Using these methods, molecules are cut into atoms or fragments, and the logP values are calculated from the sum of the contributions of relevant groups present in the molecule. The second kind of method is based on molecular properties such as molecular lipophilicity potentials (MLP), electrostatic potential, solvent-accessible-surface area, and so on.

The first model for calculating logP was the 'hydrogen replacement' (π) system developed by Hansch and Fujita in 1964 [1]. After that, Rekker [2] and Broto [3] proposed the first fragment-based contribution method and the first atom-based contribution method, respectively. Since that time, many new methods have been proposed, and the accuracy and coverage of prediction have been significantly improved [4]. Methods for estimations of LogP values developed before the 1980s were reviewed by Lyman in 1990 [5]. Mannhold and Waterbeemd reviewed most of the logP models published before the year 2000 [4]. In this paper, we report the newest improvement of the most popular logP models in recent years. A logP model will not be listed in this review if there is no relevant report concerning this model after 1990.

It was reported that around 40% of drug candidates fail because of pharmacokinetic problems [6]. For orally administered drugs the process of absorption can be viewed as the diffusion of the drug molecule through a number of membranes and aqueous microenvironments. If a drug has little ability to pass through membranes, then its bioavailability will be low. Such a problem represents one of the major difficulties of pharmacokinetics.

Drugs can cross a membrane either by passive diffusion or with the help of protein carriers inside the membrane. Lipophilicity has been found to be one of the most important factors in determining the interactions of drug molecules with both membranes and protein carriers. It has been used to predict passive drug transport across cell membranes as well as oral absorption. Recent progress in the study of the relationships between lipophilicity and membrane transport is reviewed in the third part of this paper.

LOGP PREDICTION METHODOLOGIES

Group Contribution Approaches

Generally in a group contribution model, the logP values are calculated from equation such as:

$$\log P = a + \sum_{i=1}^{M} b_i B_i + \sum_{j=1}^{N} c_j C_j \quad (1)$$

where logP is the partition coefficient, a, b and c are regression coefficients. B_i is the number of occurrence of the

*Address correspondence to this author at the Department of Chemistry, Case Western Reserve University, 10900 Euclid Ave, Cleveland, Ohio 44106, USA; Tel: +1-216-368-3618; Fax: +1-216-368-5921; Email: gxk6@po.cwru.edu

1389-5575/05 $50.00+.00

*i*th basic group and C_j is the number of occurrence of the *j*th correction factor.

In atom-based methods, the basic groups (or descriptors) are atoms along with the number and type of bonds attached to it. They are often referred to as 'atom cores'. The logP value of a molecule is obtained by adding the contributions of each atom core of the molecule. The advantage of the atom contribution methods is that the relationship between the logP and the 'atom cores' is straightforward and easy to explain. But a clear shortcoming of these methods is that they cannot handle the effects of long-range interactions (e.g. intramolecular H-bond) on the logP values.

In contrast, methods using larger fragments as descriptors instead of atom cores, can incorporate significant intramolecular electronic interactions. Nevertheless, the definition of the fragments used in the calculation of logP remains somewhat arbitrary and are much more difficult to identify. Current group contribution approaches often use both atom cores and fragments as their descriptor sets.

In the past ten years, many group contribution approach models for the estimation of logP values have been developed and implemented in computer programs. Current popular logP models include ClogP [7,8], KlogP [9], AlogP [11], HlogP [12], ∑f system [13,14], LOGKOW [15], ACD/LogP [16] and XlogP [17,18] (Table **1**).

Following the pioneering work of Hansch on the relationships between n-octanol/water partition coefficients and biological activities, Leo and Hansch built the first ClogP model in 1982, and ClogP became a widely used computerized logP calculation model [7]. In ClogP 4.0 [8], the newest version of this model, a new algorithm called FRAGCALC was proposed to calculate the contribution values of fragments if they were not part of the defined descriptor set. This algorithm was based on a set of around 600 dependably measured descriptors, which only have aliphatic or aromatic bonds. It was reported that the average deviation of this method is ±0.31 logarithm unit [8].

Klopman's model for the calculation of logP (KlogP) uses atomic groups as basic parameters and fragment groups as correction factors [9]. First, a model is developed with the basic parameter set only, then an artificial intelligence system, Computer Automated Structure Evaluation (CASE)

Table 1. Group Contribution Approaches and Property-Based Approach Methods for the Estimation of logP Values of Diverse Organic Chemicals

Model	Descriptors	Size of Training Set	Newest Released Software	Internet Web Page
Group Contribution Approaches				
ClogP	Atomic and fragmental	12727*	ClogP 4.0	www.biobyte.com
KlogP	Atomic and fragmental	1663	K-PRO	www.multicase.com
AlogP	Atomic	9920	Galaxy	www.am-tech.com
HlogP	Molecular hologram	-	-	-
∑f system	Fragmental	-	PrologP 6.0_cdr	www.compudrug.hu
LOGKOW	Atomic and fragmental	10589	KowWin	esc.syrres.com
ACD/LogP	Atomic and fragmental	3601	ACD/LogP	www.acdlabs.com/
XlogP	Atomic	1853	XlogP 2.0	cheminfo.pku.edu.cn/ calculator/xlogp/manual/compile.html
Property-based Approaches				
BlogP	General molecular properties	302	-	-
QlogP	Molecular Volume	320	-	-
VlogP	Electrotopological State Values	6675	Topkat 3.0	www.accelrys.com
MlogP	Number of lipophilic groups and number of hydrophilic groups	1230	SYBYL 6.0.3	www.tripos.com
AUTOLOGP	Topological indices	7200	AUTOLOGP 4.0	-
CLIP	Molecular lipophilicity potential	-	CLIP 1.0	www-ict.unil.ch
HINT	Hydrophobic atom constant	-	HINTLOGP 2.35	www.eslc.vabiotech.com
SLIPPER	Polarizability and hydrogen bond acceptor strength	10937	SLIPPER-2001	www.timtec.net
SciLogP	Electrotopological state indices	8909	SciLogP Ultra	www.scivision.com
ALOGPS	Electrotopological state indices	12908	ALOGPS 2.1	vcclab.org

* Base on the newest information on http://clogp.pomona.edu/medchem/chem/clogp/starlist/index.html

methodology [10], is used to help identify additional correction factors to account for the calculation errors. The final model was derived by correlating 94 parameters with a database containing 1663 chemicals using regression analysis (N=1663, Rsq=0.928, S=0.38, F=218). The newest version of KlogP is available from MULTICASE Inc. , as part of the software 'K-Pro'.

AlogP uses a pure atom-based contribution method to calculate the logP values of organic chemicals. In a recent report about the AlogP model [11], a parameter set containing 68 atom type descriptors was used to create the model. Using these descriptors to correlate the logP values of 9920 organic chemicals, an Rsq value of 0.918 and a standard deviation of 0.68 were obtained.

Another model based on the same methodology as AlogP, was proposed by Viswanadhan *et al.* in 2000 under the name of HlogP [12]. In this approach, a new type of fragment known as 'molecular hologram' was used and partial least square regression was used with these hologram descriptors to create the model. When compared to the AlogP and ClogP models, HlogP method was shown to give a better prediction when dealing with drug like molecules [12].

The hydrophobic fragmental constant approach ($\sum$f system) represents the first fragmental contribution methodology calculation of logP. A total of 169 hydrophobic fragmental descriptors were defined by Mannhold [13, 14]. In addition, 13 correction factors were also used. The clear advantage of the $\sum$f system is its ease of use, even without the help of a computer, because of the relatively simple and limited descriptor set.

LOGKOW, created and updated by Meylan *et al.* [15], is a quite recent group contribution approach. In this model, a new methodology called the Experimental Value Adjusted (EVA) approach was used. In contrast to the classic group contribution models, the EVA approach evaluates the contribution of fragments by comparing closely related analogs. Challenged with a new molecular structure, the program identifies the closest analog from the molecules of the learning set and calculates the logP of the new structure by adding (or subtracting) the contribution of the groups that have to be added (or removed) in order to transform the structure of the analog to that of the unknown molecule.

For example, the logP value of 1,2-dichlorobenzene can be calculated from chlorobenzene (with a known logP value of 2.84) and adding the coefficient of an aromatic chlorine (0.64 in LOGKOW model) to the experimental logP value of chlorobenzene, thus giving logP = 2.84+0.64=3.48 for 1,2-dichlorobenzene. As can easily be anticipated, the accuracy of the predicted value will greatly depend on whether there is a good starting analog of the target compound in the learning set.

In the latest report of LOGKOW model, 150 atom/fragment and 250 correction factors were reported, and an Rsq of 0.943 was obtained when the program was used to recalculate the logP values of the learning set of 10589 chemicals.

See www.multicase.com

ACD/LogP was developed using fragmentation rules that were based on a definition of the isolating carbon (IC) [16]. The IC parameters are similar to the parameters used in ClogP but differ in a number of ways. First, sp^2 carbon atoms attached to two aromatic heteroatoms (like a 2-pyrimidinyl carbon) or to a sp carbon (like an acetylenyl carbon) are not defined as IC's. Second, Hydrogen atoms are included in the ICs. For example, $-CH_2-$, >CH-, >C< and =CH- groups are 4 individual ICs in the ACD/LogP model. But in ClogP, they are represented by the same IC (>C<) with relevant structural correction factors. Over 500 basic descriptors (f) and over 2000 correction factors (F) were included in the newest release of the ACD/LogP software. When a new chemical is submitted to this model, its structure is fragmented and f and F constants of fragments are identified from descriptor set. If there are unknown fragments in the molecule, the f and F value of this unknown fragment are calculated from a fragmental increment equation similar to the Ghose-Crippen method [19] and a multi-linear equation similar to the Hammett equation. Then the logP value of the chemical is obtained by adding the f and F values of all the fragments found in the molecular structure.

XLOGP was first reported as a pure atomic contribution approach by Wang in 1999 [17]. In a recently released version of the model, XLOGP v2.0 [18], the correlation result (Rsq) for a database containing 1853 organic chemicals is listed as 0.946, with a standard deviation of 0.35. Besides 90 atom type descriptors, 10 correction factors were added into its descriptor set. The remarkable advantage of XLOGP is its simplicity of use. Furthermore, because of the simple descriptors used in this methodology, the 'missing fragment' problem does not usually occur.

Property-Based Approaches

The partition coefficient of chemical between two relatively immiscible solvents is defined as:

$$\log P = k \Delta G \quad (2)$$

In this equation, k is a constant at a fixed temperature. This equation shows that the logP is proportional to the molar Gibbs free energy of transfer between octanol and water. Therefore, it is understandable that the partition coefficients may be dependent on some molecular properties, which contribute to this free energy. It seems that the first attempt to calculate the logP values of chemicals from other molecular properties was reported by Rogers [20] in 1969. Since then, many new methods have been reported, and many new algorithms, such as Partial Least Squares (PLS), neural networks and so on, were proposed for use in these kinds of models.

Klopman *et al.* [21] first used quantum mechanical calculations based on the MINDO program and on Huckel-type calculations to estimate logP. Based on this methodology, Bodor and Huang created the BlogP model in 1992 [22]. Instead of MINDO, AM1 methodology was used for conformational analysis and the prediction of molecular properties. Eighteen descriptors were identified using a training set containing 302 chemicals, and a correlation coefficient (Rsq) of 0.95 was obtained. However, no further work on BlogP could be found after this publication.

Since the cavity term is a major contributor of the free energy of transfer, Bodor and Buchwald developed the QlogP model, which utilizes molecular volume (V) as its central descriptor to calculate logP [23, 24]. Besides molecular volume, another correction parameter N was used to account for the hydrogen bonding effect between the solvent molecules and oxygen- and nitrogen-containing functional groups of the solute molecules. The Authors found that the experimental logP values of alkanes are always higher than their predicted results, so an alkane indicator (I_{alkane}) was also included in their final equation (Eq. 3).

$$logP = aV + bN + cI_{alkane}V \quad (3)$$

Using this model, a correlation coefficient (Rsq) of 0.978 and standard error of 0.214 was obtained for a database containing 320 chemicals.

Gombar and Enslein built the VLOGP model based on a linear free energy relationship (LFER) approach in 1996 [25,26]. Electrotopological state values (E values) were used as numerical quantifiers of molecular structure in this model. The E value of an atom or a group contains information about its electron content, topology and environment. Furthermore, size-corrected E values were also used to account for molecular bulk attributes and topological shape descriptors were included to quantify the molecular shape attribute. 363 descriptors were generated from a training set containing 6675 chemicals, and a standard error of 0.20 was found when the authors retrofitted their final model.

MlogP, created by Moriguchi *et al.* [27], used two basic descriptors: the sum of lipophilic atoms (all carbons and halogens) and the sum of hydrophilic atoms (all nitrogen and oxygen atoms). With a database of 1230 compounds, they found that by using only these two parameters, they could account for 73% of the variance in the experimental logP values. After the addition of 11 correction factors, 91% of the variance in the experimental logP values of this database could be covered. A clear advantage of MlogP is that it is simple and can be easily programmable [28].

AUTOLOGP developed by Devillers [29, 30] uses novel topological indices to describe molecular structures and generate descriptors for calculating logP values. Three major types of vectors were used in this model: vector H (representing lipophilicity); vector MR (molar refractivity) and vectors HBA/HBD (describing hydrogen bond acceptors/donors). A three-layer back-propagation neural network was used as the statistical tool for this model. In the newest version (AUTOLOGP 4.0) [30], thirty-five autocorrelation descriptors were generated from a training set that consisted of 7200 chemicals.

Hydrogen Bond Thermodynamics (HYBOT) is a program to calculate quantitative hydrogen bond descriptors. Based on the relationships between HYBOT descriptors and logP, Raevsky *et al.* [31, 32] proposed that logP values could be estimated by using the polarizability (α) and hydrogen bond acceptor strength (ΣC_a) of the molecule. Using only these two descriptors, a correlation coefficient (Rsq) of 0.941 and standard error of 0.23 was found by fitting the database consisting of 2870 chemicals. They developed a software package called SLIPPER (Solubility, Lipophilicity and PERmeability) based on this methodology. In the newest release, SLIPPER2001 [32], a new tool which is similar to the EVA approach reported by Meylan [15] was used to calculate the logP value of a new chemical from its nearest neighbor (the chemical with the most similar molecular structure) in the training set with a known logP value. With the help of this new tool, the correlation coefficient (Rsq) for fitting an extended database of 10937 chemicals was increased to 0.945.

Electrotopological state (E-state) indices were introduced by Hall and Kier [33] to capture electronic, topological and valence state information. Using the E-state indices, two individual logP models, SciLogP and ALOGPS were recently developed by the SciVision company and Tetko *et al.* [34, 35], respectively. The former model used 2D E-State descriptors and a training set of 8909 chemicals. Neural Network (NN) is the basic statistic tool for the generation of the model. However, SciLogP has not yet been published. The latter model utilizes a new statistic methodology called Associative Neural Network (ASNN). This new method includes a memory functionality, and by simply including new data into the memory, it is possible to improve the prediction result without needing to retrain the neural network ensemble. The descriptor set of this model consists of 73 E-state indices and a number of hydrogen and non-hydrogen atoms, which were generated from a training set of 12908 chemicals.

CLIP_logP was created by Gaillard *et al.*, and is based on Molecular Lipophilicity Potential (MLP) [36]. The MLP defined in this model was based on the Broto's atomic descriptors, and a modification of the distance function used by Fauchère [37]. The summation of positive MLP values (ΣMLP^+) and the summation of negative MLP values (ΣMLP^-) were used to represent the 'lipophilic' and 'hydrophilic' parts of the molecule. The logP values were calculated by a linear correlation equation in which ΣMLP^+ and ΣMLP^- were used as the two basic parameters.

The Hydropathic INTeractions (HINT) model utilizes three-dimensional descriptors in calculating logP [38, 39]. The key parameter in the HINT model is the hydrophobic atom constant ($\mathbf{a_i}$). The logP value of a chemical is calculated as the sum of the $\mathbf{a_i}$ values of all the atoms of the molecule. The $\mathbf{a_i}$ value of each atom is presented as representing how it interacts with both the other atoms of the molecule and the solvent molecules/atoms. Because $\mathbf{a_i}$ values directly relate to the free energy of atom transfer between two solvents, the logP values can be calculated by simply summing the $\mathbf{a_i}$ values of the molecule.

QSAR STUDIES OF MEMBRANE TRANSPORT USING LOGP MODELS

Gastrointestinal Absorption

Gastrointestinal (GI) absorption is an important factor in determining the pharmacokinetic behavior of drugs. Caco-2 cell monolayers are a standard tool used in screening for GI absorption [40]. It was proposed that low permeability across cell monolayers could be used to identify drug candidates with absorption problems.

ACD/logP was used to study the permeability of 17 drugs across Caco-2 monolayers by Osterberg and Norinder

[41]. With the help of a PLS analysis, lipophilicity (shown as ACD/logP value) was shown to be the most important descriptor in calculating transport across the Caco-2 monolayers as high lipophilicity favors the permeability of drugs. Similar result was obtained for another data set consisting of 17 peptides in the same report [41].

Another similar work was reported by Bravi and Wikel [42]. Instead of logP, MS-WHIM descriptors were used to correlate the experimental permeability values throughout Caco-2 cells of 17 drugs, again using a PLS analysis. A correlation coefficient (Rsq) of 0.98 was obtained for their best model. Moreover, using the same MS-WHIM descriptor set, the logP values of 268 organic chemicals could be predicted. It is expected that a similar relationship between the lipophilicity and permeability of these chemicals may exist as well.

Lipophilicity was used as one of the four major parameters to study drug permeability by Lipinski *et al.* [28]. They built a database, which contains around 2,500 drugs with relevant information about clinical exposure. Using the ClogP model to predict the logP values of the chemicals in this library, they found that only about 10% of the drugs have a ClogP value over 5. Therefore, they proposed that poor permeation will be observed for drugs with a logP value over 5. This rule became one of components of the 'rule of 5', which is an approximate method used to study the permeability of a new drug candidates.

Multidrug Resistance

Human P-glycoprotein is a 1280 amino acid surface ATP-binding cassette transport protein. Efflux out of cells by P-glycoprotein (PGP) decreases the active concentration and the effect of pharmaceuticals, especially for anti-cancer agents. Therefore, identification of potential PGP substrates among drug candidates is important for understanding their bioavaliability. Moreover, it is also important for designing potential PGP inhibitors (multidrug resistance modulators).

Lipophilicity has been used as the major descriptor in almost all SAR and QSAR researches about PGP substrates. An early review about physical-chemical properties shared by compounds that modulate multidrug resistance in human leukemic cells proposed a weak but noticeable relationship (Rsq=0.62) between the logP and the modulation activity [43] of 12 drugs. However, while logP was reported in some studies to be an important descriptor in QSAR studies of PGP substrates/inhibitors [44-49], others did not find such relationship [50, 51].

A recent study of propafenone type PGP inhibitors using lipophilicity was done by Ecker and Chiba [47]. Using logP values calculated by MOLGEN (Ghose and Crippen method [19]), an excellent correlation (Rsq=0.98) between inhibition of rhodamine 123 efflux and lipophilicity was reported for eleven propafenone analogs. However, several outliers were excluded from this correlation. Similar results were obtained in follow-up studies [48, 49]. Seelig proposed that partition into the membrane was the rate-limiting step for binding with PGP, and the strength and number of hydrogen bonds formed between the substrate/inhibitor and the transporter (PGP) determined the dissociation of the PGP-substrate complex [52]. It was therefore suggested that two steps (partition and interaction) were involved in the process of efflux by PGP, and lipophilicity could only describe the first step. In a recent publication of Ecker and Chiba [49], two new descriptors in addition to logP were proposed to correlate with the MDR reversal activity of 15 chemicals (Eq. 4).

$$\log(1/EC_{50})=0.82(\pm0.11)\text{logP} - 50.24(\pm24.14)Ch - 0.32(\pm0.10)L - 21.52(\pm9.66) \quad (4)$$

N=15, r=0.98, s=0.17, F=87.17

The descriptor *L* was used to account for the structural difference and *Ch* was the charge descriptor, which was used to calculate the hydrogen bond acceptor strength of the molecule.

Blood Brain Barrier

The blood brain barrier (BBB) keeps homestasis within the central nervous system (CNS) by separating the brain from the systemic circulation. It is important to evaluate the permeability of a drug, which must pass through the BBB to reach its target. For drugs target at other sites, high BBB permeability may cause unwanted side-effect. For example, the tubulin-binding agent 1069C85 was developed to overcome multidrug resistance associated with existing anti-tubulin agents, but severe central neurotoxicity was found during its phase I studies. This ultimately resulted in the failure of this drug [53].

A good correlation between lipophilicity and the degree of BBB penetration (expressed as $\log(C_{brain}/C_{blood})$ in most studies) was found by Habgood *et al.* for 18 hydroxypridinones [54]. However, even in this small homological data set, several outliers were excluded from the correlation. This indicated that a model based exclusively on logP could not be sufficient to explain BBB permeability.

Because of his success in modeling human intestinal absorption with polar surface area (PSA), Clark used lipophilicity and polar surface area (PSA) to predict $\log(C_{brain}/C_{blood})$ for a set of 55 diverse organic compounds [55]. Both ClogP and MlogP were used to build two individual models in this study for comparison. Using ClogP values and PSA as the descriptors, a correlation coefficient (R) of 0.887 and a standard error (S) of 0.354 was found. If MlogP values are substituted for the ClogP values, the correlation coefficient (R) is 0.876 and the standard error (S) is 0.369. Using the same chemicals and the ACD/logP and ACD/ChemSketch descriptors, Österberg and Norinder obtained a good PLS model (Rsq=0.956) for their $\log(C_{brain}/C_{blood})$ [41]. In this model, the dominant PLS coefficients were associated with the ACD/logP and hydrogen bonding descriptors.

Skin Permeation

Skin permeability is a crucial parameter in evaluating the absorption of non-invasive parenteral administrated drugs. Moreover, the toxicity of all industrial and environmental chemicals that may enter in contact with the skin should also be evaluated.

In 1992, Potts and Guy used logP and molecular weight to correlate the skin permeability of more than 90 chemicals

and obtained a noteworthy correlation (Rsq=0.67) [56]. Based on this work, Barratt using molecular volume (MV), meting point (mpt) and logP, was able to find a significant correlation with the skin permeability (logPC) of an equivalently sized but different data set (Eq. 5) [57]. The correlation coefficient was 0.765 for 91 diverse chemicals. The logP was shown to be the most important descriptor in this equation.

$$\log PC = -0.00484MV + 0.719\log P - 0.00515mpt - 2.665 \quad (5)$$

N=91, Rsq=0.765, S=0.601, F=94.25

A similar QSAR study was reported by Cronin *et al*. for the prediction of the skin permeability coefficients ($\log K_p$) of 114 compounds. Here, the Molecular weight (M_r) was used to represent the molecular volume and ClogP was the other descriptor (Eq. 6).

$$\log K_p = -0.00698\ M_r + 0.612\log P - 2.47 \quad (6)$$

N=114, Rsq=0.667, S=0.708, F=110

Patel *et al*. proposed a recent QSAR study of skin permeation for 158 chemicals using ClogP and other molecular property descriptors [59]. The best statistical correlation was found by using ClogP values, molecular weight (MW), the sum of absolute charges on oxygen/nitrogen atoms (**ABSQon**) and the sum of E-state indices for all methyl groups (**SsssCH**) (Eq. 7).

$$\log K_p = 0.681\mathbf{logP} - 0.00653\mathbf{MW} - 0.284\mathbf{ABSQon} - 0.268\mathbf{SsssCH} - 2.47 \quad (7)$$

N=158, Rsq=0.76, S=0.55, F=123

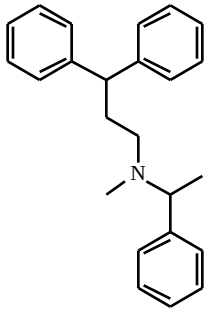
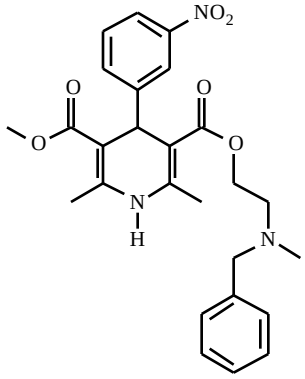
EXPECTATIONS FOR FUTURE PROGRESS

Until now, there has not been a single program capable to calculate logP values for structurally diverse compounds with consistent accuracy. When structurally diverse drugs are evaluated, the calculated logP values vary dramatically. The most common problems of the current logP models were reviewed by Leo [60]. Mannhold et al listed several comparative studies of the available logP models [61, 62]. For all models, the predicted logP values of complex drug molecules were much worse than those of simple organic compounds. The predictive power of a logP model can be improved significantly by enlarging the training set. However, in the membrane transport area, the following two additional problems cannot be ignored in future QSAR work using logP.

First, the logP value is not constant for ionizable compounds in measurements at different pH. This condition may result in pH dependent distribution and activity. In these cases the distribution coefficient (logD) should be utilized instead of logP. For example, the measured logP values of fendiline and nicardipine are nearly the same [63]. But at physiological pH (pH=7.4), the logD of fendiline is more than 1 logarithm lower than that of nicardipine. Having a high ionization constant (pKa value), Fendiline is much more hydrophilic in living systems than what would be expected from its logP value (Table **2**). This clearly indicates that the choice of partition or distribution coefficient can have a significant influence on the results of QSAR studies.

Second, the lipophilicity is only one of the major molecular properties needed to determine drug transport.

Table 2. Comparison of logP, logD and pK_a Values of Fendiline and Nicardipine (Values from ref. 63)

Drug	Structure	logP (exp)	logD (pH=7.4)	pK_a
Fendiline		4.89	3.19	9.04
Nicardipine		4.65	4.33	7.28

Based on the result of previous works, Patel *et al.* [59] concluded that the three key descriptors for predicting skin permeability were hydrophobicity, molecular size and hydrogen bonding ability. Since only hydrophobicity was well defined (logP), improved predictions of drug transport is expected to be realized only after the optimum descriptors are identified and codified for molecular size and hydrogen bond capability.

CONCLUSION

Lipophilicity is a dominant physicochemical property of drugs in relation to their transport and absorption characteristics. Therefore, the ability to predict logP is definitely one of the crucial components of drug research, especially in the membrane transport area. With the continuing improvement of algorithmic methodologies for predicting logP values, these methods will continue to play an important role in helping predict the pharmacodynamic behavior of new drug candidates, even before the chemical is synthesized.

REFERENCES

[1] Hansch, C.; Fujita, T. *J. Am. Chem. Soc*., **1964**, *86*, 1616.
[2] Rekker, R. *The hydrophobic Fragmental Constant*. Elsevier: New York, **1977**.
[3] Broto, P.; Moreau, G.; Vandycke, C. *Eur. J. Med. Chem.,* **1984**, *19*, 71.
[4] Mannhold, R.; Waterbeemd, H. *J. Comp.-Aid. Mol. Design*., **2001**, *15*, 337.
[5] Lyman, W. J.; Potts, R. G. In *Handbook of Chemical Property Estimation Methods. Enviromental Behavior of Organic Compounds.* Lyman, W. J.; Reehl, W. F.; Rosenblatt, D. H., Ed. American Chemical Society, Washionton D. C., **1990**; Vol. *14*.
[6] Kennedy, T. *Discovery Today*, **1997**, *2*, 436.
[7] Leo, A. J. *Chem. Rev*., **1993**, *93*, 1281.
[8] Leo, A. J.; Hoekman, D. *Persp. Drug Discov*., **2000**, *18*, 19.
[9] Klopman, G.; Li, J.; Wang, S.; Dimayuga, M. *J. Chem. Inf. Comput. Sci.,* **1994**, *34*, 752.
[10] Klopman, G. *J. Am. Chem. Soc.,* **1984**, *106*, 7315.
[11] Wildman, S. A.; Crippen, G. M. *J. Chem. Inf. Comput. Sci.,* **1999**, *39*, 868.
[12] Viswanadhan, V. N.; Ghose A. K.; Wendoloski, J. J. *Perspect. Drug Discov*. *Design*, **2000**, *19*, 85.
[13] Mannhold, R.; Rekker, R. F.; Dross, K.; Bijloo, G.; Vries, G. *Quant. Struct-Act. Relat.,* **1998**, *17*, 517.
[14] Mannhold, R.; Rekker, R. F. *Perspect. Drug Discov*. *Design*, **2000**, *18*, 1.
[15] Meylan, W.; Howard, P. H. *Perspect. Drug Discov*. *Design*, **2000**, *19*, 67.
[16] Petrauskas, A. A.; Kolovanov, E. A. *Perspect. Drug Discov*. *Design*, **2000**, *19*, 99.
[17] Wang, R.; Fu, Y.; Lai L. *J. Chem. Inf. Comput. Sci.,* **1997**, *37*, 615.
[18] Wang, R.; Gao, Y.; Lai, L. *Perspect. Drug Discov*. *Design*, **2000**, *19*, 47.
[19] Ghose, A. K.; Crippen, G. M. *J. Chem. Inf. Comput. Sci.,* **1987**, *27*, 21.
[20] Rogers, K. S.; Cammarata, A. *Biochim. Biophy. Acta*, **1969**, *193*, 22.
[21] Klopman G.; Iroff, L. D. *J. Chem. Inf. Comput. Sci.,* **1981**, *2*, 157.
[22] Bodor, N.; Huang, M. *J. Pharm. Sci*., **1992**, *81*, 272.
[23] Bodor, N.; Buchwald, P. *J. Phys. Chem.*, **1997**, *B 101*, 3404.
[24] Buchwald, P. *Perspect. Drug Discov*. *Design*, **2000**, *19*, 19.
[25] Gombar, V.; Enslein, K. *J. Chem. Inf. Comput. Sci.,* **1996**, *36*, 1127.
[26] Combar, V.; Enslein, K. *SAR QSAR Environ. Res*., **1999**, *10*, 371.
[27] Moriguchi, I.; Hirono, S.; Liu, Q.; Nakagome, Y.; Matsushita, Y. *Chem. Pharm. Bull*., **1992**, *40*, 127.
[28] Lipinski, C. A.; Lombardo F.; Dominy, B. W.; Feeney P. J. *Adv. Drug Deliv. Rev*., **2001**, *46*, 3.
[29] Devillers, J.; Domine, D.; Karcher, W. *Polycyclic Arom. Comp*., **1996**, *11*, 211.
[30] Devillers, J.; Domine, D.; Guillon, C. *Eur. J. Med. Chem*., **1998**, *33*, 659.
[31] Raevsky, O. A. *SAR QSAR Environ. Res.*, **2001**, *12*, 367.
[32] Raevsky, O. A.; Trepalin S. V.; Trepalina H. P.; Cerasimenko, V. A.; Raevskaja, O. E. *J. Chem. Inf. Comput. Sci.*, **2002**, *42*, 540.
[33] Hall, L. H.; Kier, L. B. *J. Chem. Inf. Comput. Sci*., **1995**, *35*, 1039.
[34] Tetko, I. V.; Tanchuk, V. Y.; Vila, A. E. P. *J. Chem. Inf. Comput. Sci*., **2001**, *41*, 1407.
[35] Tetko, I. V.; Tanchuk, V. Y. *J. Chem. Inf. Comput. Sci*., **2002**, *42*, 1136.
[36] Gaillard, P.; Carrupt, P. A.; Testa, B.; Boudon, A. J. *J. Comp.-Aid. Mol. Design*., **1994**, *8*, 83.
[37] Fauchère, J. L.; Quarendon, P.; Kaetterer, L. *J. Mol. Graphics*, **1988**, *6*, 203.
[38] Kellogg, G. E.; Abraham, D. J. *Analusis*, **1999**, *27*, 19.
[39] Kellogg, G. E.; Abraham, D. J. *Eur. J. Med. Chem*., **2000**, *35*, 651.
[40] Artursson, P.; Palm, K.; Luthman, K. *Adv. Drug. Deliv. Rev*., **1996**, *22*, 67.
[41] österberg, T.; Norinder, U. *Eur. J. Pharm. Sci*., **2001**, *12*, 327.
[42] Bravi, G.; Wikel, J. H. *Quant. Struct-Act. Relat*., **2000**, *19*, 39.
[43] Zamora, J. M.; Pearce, H. L.; Beck, W. T. *Mol. Pharm*., **1988**, *33*, 454.
[44] Inaba, M, Maruyama, E. *Cancer. Res*., **1988**, *48*, 2064.
[45] Emmer, G.; Crassberger, M. A.; Schulz, G.; Boesch, D.; Gaveriaux, C.; Loor, F. *J. Med. Chem*., **1994**, *37*, 1918.
[46] Kawase, M.; Saito, S.; Motohashi, N. *Int. J. Antimicrob. Agents*, **2000**, *14*, 193.
[47] Chiba. P.; Ecker, G.; Schmid, D.; Drach, J.; Tell, B.; Goldenberg, S.; Gekeler, V. *Mol. Pharm*., **1996**, *49*, 1122.
[48] Schmid, D.; Ecker, G.; Kopp, S.; Hitzler, M.; Chiba, P. *Biochem. Pharm*., **1999**, *58*, 1447.
[49] Chiba P.; Holzer, W.; Landau, M.; Bechmann, G.; Lorenz, K.; Plagens, B.; Hitzler, M.; Richter, E.; Ecker, G. *J. Med. Chem*., **1998**, *41*, 4001.
[50] Litman, T.; Zeuthen, T.; Skovsgaard, T.; Stein, W. D. *Biochim. Biophys. Acta*, **1997**, *1361*, 159.
[51] Thimmaiah, K. N.; Jayashree, B. S.; Germain, G. S.; Houghton, P. J.; Horton, J. K. *Oncology Res*., **1998**, *10*, 29.
[52] Seelig, A., Landwojtowicz, E. *Eur. J. Pharm. Sci*., **2000**, *12*, 31.
[53] Judson, I.; Briasoulis, E.; Raynaud, F.; Hanwell J.; Berry, C.; Lacey, H. *B. J. Cancer*, **1997**, *75*, 608,
[54] Habgood, M. D.; Liu, Z. D.; Dehkordi, L. S.; Khodr, H. H.; Abbott, J.; Hider, R. C. *Biochem. Pharm.*, **1999**, *57*, 1305.
[55] Clark, D. E. *J. Pharm. Sci.* , **1999**, *88*, 815.
[56] Potts, R. O.; Guy, R. H., *Pharm. Res*., **1992**, *9*, 663.
[57] Barratt, M. D. *Toxic. In Vitro*, **1995**, *9*, 27.
[58] Cronin, M. T. D.; Dearden, J. C.; Moss, G. P.; Murray-Dickson, G. *Eur. J. Pharm. Sci*., **1999**, *7*, 325.
[59] Patel H.; Berge, W.; Cronin, M. T. D. *Chemosphere*, **2002**, *48*, 603.
[60] Leo, A. J. Lipophilicity in Drug Action and Toxicology. Pliska, V., Testa, B., Waterbeemd, H. Ed. VCH Publishers, Inc. New York, **1996**, Vol. *9*, 157-172.
[61] Mannhold, R.; Dross, K., *Quant. Struct-Act. Relat*., **1996**, *15*, 403.
[62] Mannhold, R.; Cruciani, G.; Dross, K.; Rekker, R. *J. Comp.-Aid. Mol. Design*., **1998**, *12*, 573.
[63] Franke, U.; Munk, A.; Wiese, M. *J. Pharm. Sci*., **1999**, *88*, 89.