

COMPARISON OF FOUR METHODS OF PREDICTING NEWLY MEASURED
OCTANOL/WATER COEFFICIENTS ($\log K_{ow}$) FOR HETEROCYCLIC NITROGEN
COMPOUNDS AND THE PARTITION MECHANISMLI LI,[†] LIANSHENG WANG,*[†] SHUOKUI HAN,[†] and ZHENG ZHANG[‡][†]State Key Laboratory of Pollution Control and Resource Reuse, Department of Environmental Science and Engineering[‡]Department of Chemistry, Nanjing University, Nanjing 210093, China

(Received 27 October 1998; Accepted 30 March 1999)

Abstract—Log K_{ow} values for 14 heterocyclic nitrogen compounds (by traditional shake-flask method) were reported. Reverse phase–high-performance liquid chromatography (RP-HPLC) method and three estimation models—linear solvation energy relationship, molecular connectivity indices (MCIs), and quantum chemical descriptors—were used to study the partition mechanism and to establish the correlations of the models with the measured $\log K_{ow}$ values by stepwise variable regression. Research on the partition mechanism showed that the molecular bulk properties and the hydrogen bond interaction were the most important factors influencing the partition process. Comparison of the quality of the models was made among the four methods. According to the simplicity of the method, the availability of the descriptors, and the closeness of correlation with the measured $\log K_{ow}$ values, the RP-HPLC method was not satisfactory, while the MCIs method introducing the non-disperse force factor was the most satisfactory.

Keywords—*n*-Octanol/water partition coefficient Heterocyclic nitrogen compounds Partition mechanism Quantitative structure–activity relationship

INTRODUCTION

The partition coefficient is defined as the ratio of the concentration of compounds equilibrated in two immiscible phases, normally oil and aqueous phases. *n*-Octanol/water partition coefficient (K_{ow}) has been proven to be a valuable tool in environmental sciences and related areas [1,2]. As a measure of the hydrophobic character of compounds, K_{ow} has been used for predicting their environmental fate and toxicity for a large variety of organic chemicals [3–5]. However, for heterocyclic nitrogen compounds, their $\log K_{ow}$ values (by shake-flask method) are rarely available. Heterocyclic nitrogen compounds, including pyridine, triazine, and pyridazine, are widely used in agriculture as insecticides and herbicides and are also used in industry as adhesives for textiles, as chemicals, as solvents, and as catalysts [6]. Because of their high bioactivity, they play an important role in the environment. In this paper, we report $\log K_{ow}$ values for 14 heterocyclic nitrogen compounds that were tested using the traditional shake-flask method.

Because the shake-flask method has its own disadvantages and because the measurement of every pollutant is not trivial work, methods that allow for better experimentation and prediction have been proposed. The reverse phase–high-performance liquid chromatography (RP-HPLC) method is one such experimental method that has been proposed as an alternative to the shake-flask method. The advantages that the RP-HPLC method offers (as compared to the conventional shake-flask method) include speed of determination, increased reproducibility, ease of automation, the requirement for only a very small amount of the sample to be determined, and the capability of separating the impurities in the sample simultaneously with the determination. In the RP-HPLC system, the capacity factor, k , is defined as $k = (t_r - t_0)/t_0$, where t_r is the retention

time and t_0 is the dead time. Logarithms of k' ($\log k'$) in different ratios of binary eluent were tested. The $\log k'$ value in pure water, $\log k'_w$, is obtained by extrapolating the $\log k'$ value in binary eluent to pure water eluent. Some scientists [7] think that the $\log k'_w$; for many kinds of compounds may be considered to be a good replacement for their $\log K_{ow}$. In our work, the correlation of $\log k'_w$ values with $\log K_{ow}$ values for 14 heterocyclic nitrogen compounds was observed.

Besides the experimental method, three QSAR prediction models have been presented in order to establish the correlation of the $\log K_{ow}$ values of the studied compounds with their structural descriptors or other physicochemical properties. The first is the linear solvation energy relationship (LSER) model, which is proposed by Kamlet and coworkers [8] in order to describe properties dependent on the solute–solvent interaction. In this model, chemical properties are related to molecule structure through energy required to surround the solute with solvent molecules and through energies gained or lost through the formation of electrostatic and hydrogen bond between the chemicals and the medium (to stabilize the solvent molecules and to keep the compound inside it). The chemical properties dependent on the solute–solvent interaction are the results of the contribution of component groups that form the molecule. Usually, the chemical properties can be described by the following predictive equation:

$$XYZ = XYZ_0 + mVi/100 + s\pi^* + b\beta + a\alpha \quad (1)$$

The endoergic energy term $mVi/100$ measures the cavity term, and π^* , β , and α are the solvatochromic parameters that measure solute dipolarity/polarizability, hydrogen bond acceptor basicity, and hydrogen bond donor acidity, respectively. Many chemical properties have been successfully correlated, rationalized, and predicted by the LSER model [9–11]. The advantage of this model is that it can provide information on the

* To whom correspondence may be addressed (ecls@nju.edu.cn).

nature and the solute–target system interaction affecting the property of interest.

The second model we dealt with is the molecular connectivity indices (MCIs) system, which is the most successful system of topological quantitative modeling. The information used in the MCIs calculation includes the number and the type of the atoms and bonds as well as the number of all electrons and valence electrons for nonhydrogen atoms. There are many extensive reviews on the MCIs method within the literature [12,13]. Here we only present a brief description of the valence zero-order and first-order MCIs. In the valence approximation, nonhydrogen atoms are described by their atomic valence delta values (δ^v), which are calculated from their electron configuration using the following equation: $\delta^v = (Z^v - h)/(Z^v - Z - 1)$. In this equation, Z^v is the number of valence electrons in an atom, Z is the atom's atomic number, and h is the number of hydrogen atoms bound to the same atom. The symbol ${}^0\chi^v$ represents a zero-order valence index, whereas ${}^1\chi^v$ represents a first-order valence index. ${}^0\chi^v$ and ${}^1\chi^v$ are calculated from the electronic input information (atomic δ^v values) based on the following equation: ${}^0\chi^v = \sum \delta_i^v$ for all atoms and ${}^1\chi^v = \sum (\delta_i^v \delta_j^v)^{-0.5}$ for all bonded pairs of atoms, where i and j correspond to each adjacent individual nonhydrogen atom and where summation is made over all nonhydrogen atoms. Since MCIs are nonempirical parameters and can be calculated quickly and accurately from chemical structures with no need for experimentation, they are widely used in QSAR.

The third method we studied is the quantum chemical descriptors method, which is a kind of semiempirical structural descriptor that can be calculated by the chemical structure, with no experimental needs. The availability of the molecular mechanics and the quantum chemical software makes it feasible to calculate all kinds of descriptors. Their advantages are obvious: They are not restricted to closely related compounds and they describe clearly defined molecular properties. With the development of the quantum chemical software, this method developed quickly in the QSAR studies [14,15].

In this paper, we tried the RP-HPLC method as well as the three estimating methods in order to study the partition mechanism and to evaluate the ability of these methods to predict log K_{ow} values for heterocyclic nitrogen compounds as well.

MATERIALS AND METHODS

Samples

Fourteen heterocyclic nitrogen compounds, including triazines, pyridazine, and pyridines, were synthesized in our laboratory. Their purities were monitored by HPLC. The chemical structures and names of these compounds are listed in Table 1.

Determination of partition coefficients

K_{ow} values were determined using the shake-flask method, as described by the Organization for Economic Cooperation and Development's guidelines for testing of chemicals [16]. *n*-Octanol (analytical grade) saturated by water was used as aqueous phase, and water saturated by *n*-octanol was used as aqueous phase. The volume ratio of organic phase to aqueous phase was 1:9 ml. Octanol solutions of each compound of two different known concentrations were prepared. Three tubes were prepared for each concentration. Blanks were prepared in an identical manner, with the exception that no compounds were added. Compounds in the organic phase was allowed to come into equilibrium with compounds in aqueous phase in

the thermostatic oscillator for 8 h at 12°C. After equilibrium was achieved, the mixed solution was centrifuged, and the concentration of compounds in aqueous phase was determined with UV 2201 UV-VIS spectrophotometer. K_{ow} values were calculated from the quotient of the volume-based octanol and water concentrations. The average reproducibility of each determination was better than 1% relative.

Estimation of LSER parameters

The parameters were estimated by Hickey's "Rule of Thumb" [17]. The LSER values of the triazine ring were calculated using the benzene ring substituted by three aromatic nitrogen atoms. The estimated values included $Vi/100$, π^* , β , and α .

Calculation of MCIs

The MCIs were calculated according to the method outlined by Kier and Hall [18]. In previous studies [19], Bahnick and Doucette developed a polar fragment correction factor, $\Delta\chi$, which described nondisperse intermolecular interactions, in order to estimate soil-adsorption coefficients for organic chemicals with the use of MCIs. The authors define $\Delta\chi$ as the nondisperse force factor, which can be computed for each type of index using the following equation: $\Delta\chi = (\chi)_{np} - \chi$. In this equation, $(\chi)_{np}$ is the MCI for the nonpolar molecular structure when the polar atoms are replaced by carbon atoms and χ is the usual MCI for the molecules of interest. Detailed descriptions and calculations of $\Delta\chi$ can be found in Bahnick and Doucette [19].

Calculation of quantum chemical descriptors

The molecular modeling package ALCHEMY II [20] was used to construct and view all molecular structures. Internal coordinates were used to write all molecular structures. Molecular geometry was optimized with the Broyden–Fletcher–Goldfarb–Shanno method using the semiempirical orbital program MOPAC 6.0 [21]. The MOPAC was run with the following keywords: AM1, PRECISE, DIPOLE, POLAR, and NOINTER. The calculated quantum chemical descriptors included the molecular weight (MW), total energy (Et), energy of the highest occupied molecular orbital (E_{homo}), energy of the lowest unoccupied molecular orbital (E_{lumo}), maximum positive atomic charge (Q_{max}^+), and the maximum negative atomic charge (Q_{max}^-), etc.

Only values of π^* , α , ${}^0\chi^v$, $\Delta^1\chi^v$, total energy, and MW, which were included in the obtained regression equations, are listed in Table 2.

Regression analysis

The regression analysis was performed using the Statgraphics program (version 4.0, Statistical Graphic Corporation, Englewood Cliffs, NJ, USA). Equation 2 from Table 3 was obtained by simple regression and Equations 3 through 5 (from Table 3) were obtained by stepwise variable regression. Stepwise variable regression is a kind of multivariable regression used to select independent variables that are significant enough to correlate with the dependent variable from a series of independent variables. In our case, the dependent variable is the log K_{ow} value, and the independent variables are the different series of descriptors calculated out using the three estimation methods. By putting one series of descriptors (such as the calculated MCIs) into the independent variable database, the stepwise variable regression will select out variables of sig-

Table 1. The studied compounds and their $\log K_{ow}$ and $\log k_w$ values

No. ^a	Structure	Chemical name	$\log K_{ow}$	$\log k_w$ ^b
1		2,6-dichloro-4-morpholino-1,3,5-triazine	1.250	2.618
2		2,6-dichloro-4-diethylamino-1,3,5-triazine	2.170	3.016
3		2,6-dichloro-4-N-phenylamino-1,3,5-triazine	2.320	2.947
4		2,6-dichloro-4-N-methyl-phenylamino-1,3,5-triazine	2.430	3.247
5		3,6-dichloro-pyridazine	0.700	0.310
6		3-amino-2-chloro-pyridine	0.043	0.482
7		2-methyl-pyridine	0.327	1.133
8		2,6-dimethyl-pyridine	0.536	1.503
9		2-phenylvinyl-pyridine	1.960	3.113
10		2-amino-3-methyl-pyridine	0.0054	1.260
11		2-amino-5-methyl-pyridine	-0.059	1.298
12		2-chloro-5-chloromethyl-pyridine	0.856	1.947
13		Diethyl-2,6-pyridinedicarbonyl ester	0.150	0.934
14		2-amino-pyridine	-0.544	0.929

^a The numbers assigned in this table are used to identify the corresponding structures in subsequent tables.

^b $\log k_w$ values are from Li et al. [31].

nificance (such as ${}^0\chi^v$ and $\Delta^1\chi^v$) and will establish their correlation with the dependent variable.

RESULTS AND DISCUSSION

Structural consideration of $\log K_{ow}$ values

The chemical structures of the studied compounds and their measured $\log K_{ow}$ values are shown in Table 1. According to the structures and the $\log K_{ow}$ values listed in Table 1, these compounds can be classified into two small groups. The first group, compounds 1 through 5, 7 through 9, and 12, are among those whose structures do not contain the functional group $-\text{NH}_2$. The $\log K_{ow}$ values of these compounds increase with an increase in their molecular volume. In fact, compounds 1 through 4 also contain nitrogen atoms outside the triazine ring. However, because their nitrogen atoms are linked to functional

groups of great volume, such as phenyl, morpholino, and several ethyls, respectively, the steric effect retards the formation of hydrogen bonds, and they still act as hydrophobic compounds. The second group, compounds 6, 10, 11, and 14, are among those whose structures include the functional group $-\text{NH}_2$. For these compounds, their $\log K_{ow}$ values are small, probably because it is easy to form hydrogen bonds between the nitrogen atoms in their molecules and the water molecules in the aqueous phase, which means that these compounds have an affinity for aqueous phase. The $\log K_{ow}$ value of compound 14 is the smallest among the group, as it only contains functional group $-\text{NH}_2$. Compared to that of compound 14, the $\log K_{ow}$ values of compounds 6, 10, and 11 are larger because of the inclusion of hydrophobic functional groups such as $-\text{Cl}$ and $-\text{CH}_3$. Compound 13 does not contain $-\text{NH}_2$, but it con-

Table 2. Descriptors selected in the regression equations^a

No.	π^*	α	${}^0\chi^v$	symbol $\Delta^1\chi^v$	Et	Mol wt
1	1.750	0.000	8.790	1.063	-2,900.79	235.07
2	1.710	0.000	8.968	0.448	-2,608.14	221.09
3	2.300	0.170	8.993	0.177	-2,808.27	241.08
4	2.300	0.000	9.940	0.640	-2,963.49	255.11
5	0.430	0.000	5.314	0.597	-1,699.83	148.98
6	1.100	0.260	4.889	0.351	-1,496.43	128.56
7	0.830	0.000	4.257	0.141	-1,071.14	93.13
8	0.790	0.000	5.179	0.098	-1,226.95	107.16
9	1.560	0.050	7.798	0.140	-2,021.34	181.24
10	0.960	0.260	4.757	0.351	-1,292.12	108.14
11	0.960	0.260	4.757	0.351	-1,292.13	108.14
12	1.530	0.120	6.151	0.141	-1,791.35	162.02
13	1.210	0.240	9.226	1.785	-3,026.40	220.20
14	1.000	0.260	3.834	0.352	-1,136.25	94.12

^a ET = total energy; mol wt = molecular weight.

tains four oxygen atoms, which also readily form hydrogen bonds with water molecules.

Partition mechanism

Table 3 lists the regression equations based on the four methods (respectively) and provides a comparison of their statistical analysis results. As the regression equations were obtained using the stepwise variable regression, which selects out the variables that have the closest correlation with the $\log K_{ow}$ values into regression equation, the variables selected demonstrate the most important factor influencing the partition process.

The correlation coefficient R of Equation 2 is 0.881, which illustrates the unsatisfactory correlation of $\log k'_w$ with $\log K_{ow}$ values. From Table 1, we find that for the first group of compounds, the changes of the $\log k'_w$ agree with the changes of the $\log K_{ow}$. However, for the second group of compounds, $\log k'_w$ values do not change in accordance with the change of the $\log K_{ow}$ values, because the RP-HPLC partition system is different from the octanol/water system (because of the addition of the octadecyl-silylated column in the RP-HPLC system). As previously discovered [22–24], it is impossible to couple alkyl chains with all hydroxyl groups on the silica gel surface so that, depending on the bonding procedure, a variable number of residual silanol sites remain accessible to silanophilic interactions with solute molecules. These may involve dipole interactions with the polar or polarizable molecules, hydrogen bonding interactions, and electrostatic interactions with charged molecules that are attributable to the acidic nature of the silanol groups. For hydrophobic compounds, these interactions have little effect on the $\log k'_w$ values of the compounds. However, for compounds in which it is easy to form hydrogen bonds, the hydrogen bonding interactions in the RP-

HPLC system become more sophisticated than those in the octanol/water system, which makes the prediction of $\log K_{ow}$ values from $\log k'_w$ values unreliable for them. As it is easy to form hydrogen bonds in many of the studied compounds, it is not difficult to understand the unsatisfactory correlation of $\log k'_w$ with the $\log K_{ow}$ values.

In Equation 3, which is based on the LSER model, symbols π^* and α are selected in the regression equation. In the LSER model, π^* is defined as the dipole/polarizability term. As is suggested by Kamlet and coworkers [8], π^* , which indicates the exoergic effects of solute/solvent dipole/dipole, dipole/induced dipole, and dispersion interactions, measures the ability of the solvent to stabilize a charge or a dipole by virtue of its dielectric effect. The π^* comprises the overall information of induced dipole moment, polarizability, and molecular volume, whereas induced dipole moment and polarizability are related to molecular size and shape. That is to say that π^* depends to a large extent on the molecular bulk properties, which in turn indicates the importance of the molecular bulk properties in the partition process. Aromatic amino is a kind of compound that can be hydrogen bond acceptor and hydrogen bond donor as well. The fact that symbol β is not included in Equation 3 shows that these compounds act as hydrogen bond donors in the partition process. From Table 2, we find that for hydrophobic compounds, α values almost all equal zero, whereas α values equal 0.260 for the compounds containing $-NH_2$. The α value of compound 13 equals 0.240, as it contains four oxygen atoms. The greater the α value, the better the ability of the compound to form hydrogen bonds. Thus, it is reasonable that $\log K_{ow}$ values decrease with the increase in α values.

In Equation 4, which was obtained by MCIs, ${}^0\chi^v$ and $\Delta^1\chi^v$ are included in the regression equation. The ${}^0\chi^v$ is deemed to be a molecular volume descriptor [25,26]. Non-disperse force factor, $\Delta^1\chi^v$, is a parameter that describes the polarity of the compounds [19]. If $\Delta^1\chi^v$ is not included in the equation, the following equation is obtained by simple regression:

$$\log K_{ow} = -1.538 + 0.363{}^0\chi^v$$

$$n = 14, R = 0.800, F = 21.57, \text{ and SE} = 0.619 \quad (6)$$

The significance of the equation was greatly reduced without the inclusion of $\Delta^1\chi^v$, which demonstrates that the introduction of non-disperse force factors in the correlation equation allows for precise prediction of $\log K_{ow}$ values of the studied heterocyclic nitrogen compounds. ${}^0\chi^v$ and $\Delta^1\chi^v$ in Equation 4 can be deemed to be nonpolar and polar correction terms, respectively, which indicates the influence of the molecular bulk properties and the polarity in terms of the partition process.

The total energy and MW of these compounds are included in Equation 5, which was obtained using the quantum chemical descriptors. Since total energy provides information about the bulkiness of the solute and since MW is also a molecular size

Table 3. The obtained regression equations^a

Method	Regression equation	Eqn. R	SE	F
RP-HPLC	$\log K_{ow} = -0.721 + 0.908 \log k'_w$	(2) 0.881	0.493	41.63
LSER	$\log K_{ow} = -0.370 + 1.286 \pi^* - 3.932 \alpha$	(3) 0.918	0.393	36.02
MCIs	$\log K_{ow} = -1.912 + 0.518 {}^0\chi^v - 1.390 \Delta^1\chi^v$	(4) 0.961	0.276	79.35
Quantum method	$\log K_{ow} = -1.651 + 3.932 \times 10^{-3} \text{Et} + 6.194 \times 10^{-2} \text{MW}$	(5) 0.945	0.327	55.17

^a RP-HPLC = reverse phase-high-performance liquid chromatography; LSER = linear solvation energy relationship; MCIs = molecular connectivity indices; Et = total energy; MW = molecular weight.

Table 4. The predicted and residual values of the regression equations^a

No.	$\log K_{ow}$	RP-HPLC		LSER		MCI		Quantum method	
		Pre.	Res.	Pre.	Res.	Pre.	Res.	Pre.	Res.
1	1.250	1.594	-0.346	1.880	-0.630	1.148	0.102	1.501	-0.251
2	2.170	1.936	0.234	1.828	0.342	2.080	0.091	1.785	0.385
3	2.320	1.877	0.443	1.919	0.401	2.463	-0.143	2.235	0.086
4	2.430	2.134	0.296	2.587	-0.157	2.306	0.124	2.492	-0.062
5	0.700	-0.379	1.079	0.183	0.517	0.035	0.665	0.896	-0.196
6	0.043	-0.232	0.275	0.002	0.021	0.158	-0.115	0.432	-0.390
7	0.327	0.325	0.002	0.697	-0.370	0.127	0.200	-0.087	0.414
8	0.536	0.642	-0.106	0.646	-0.110	0.650	-0.114	0.168	0.368
9	1.960	2.020	-0.059	1.439	0.521	1.912	0.048	1.626	0.334
10	0.00535	0.434	-0.428	-0.158	0.164	0.091	-0.086	-0.027	-0.032
11	-0.059	0.466	-0.525	-0.158	0.099	0.091	-0.150	-0.027	-0.032
12	0.856	1.022	-0.166	1.125	-0.269	1.081	-0.225	1.342	-0.486
13	0.150	0.155	-0.048	0.242	-0.092	0.380	-0.230	0.090	0.060
14	-0.544	0.150	-0.694	-0.107	-0.434	-0.375	-0.169	-0.281	-0.263
Ave.	—	—	0.336	—	0.295	—	0.176	—	0.240

^a RP-HPLC = reverse phase-high-performance liquid chromatography; LSER = linear solvation energy relationship; MCI = molecular connectivity index; Pre. = predicted values calculated by the regression equation of each method, respectively. Res. = residual values, calculated by: measured values minus predicted values by each method, respectively. Ave. = average residual of all the absolute residuals for the compounds by each method, respectively.

descriptor, the inclusion of these variables illustrates the fact that molecular bulk properties are very important to the partition process. The results agree with the suggestion that has been made by some chemists [27–29] that the size and shape of the solute molecule are considered to be major factors in the determination of the solubility and partitioning behavior of the molecule.

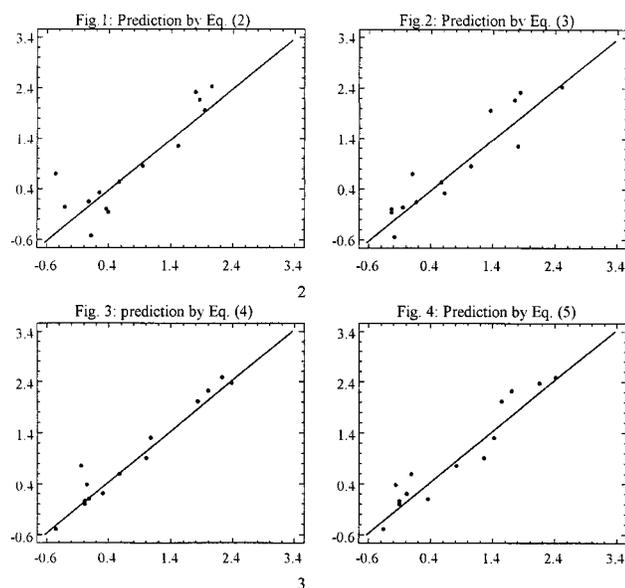
Evaluation of the models

Among the four methods used, RP-HPLC is not a prediction model. However, it can still be evaluated for its prediction quality, along with the other three estimation methods. Successful models will have to fulfill several fit criteria: the simplicity of the method, the availability of the descriptors, and the close correlation with the measured values must all be

considered. The RP-HPLC method is obviously not a simple method, because a serious consideration will have to be made in order to determine whether RP-HPLC is suitable for predicting $\log K_{ow}$ values for compounds of interest. Use of the LSER model is not as convenient as that of others, since some solvatochromic parameters for some kinds of chemicals are not available. Quantum chemical descriptors are calculated for optimized geometries obtained using a sophisticated semi-empirical quantum mechanical procedure, AM1, and the availability of the quantum chemical descriptors depends on the feasibility of the quantum chemical software, whereas MCIs can be calculated by hand with ease. Thus, the MCIs method is the simplest one of the four, and its descriptors can be easily obtained.

With regard to the closeness of correlation, three parameters can be used to test the quality of the regression equation: the correlation coefficient (R), a test of null hypothesis (F -test), and the standard error of estimate SE. For a regression equation, the higher the R and F values and the smaller the SE value, the better the correlation will be. In addition, the quality of a regression equation can also be compared based on the residual, which is the measured value minus the corresponding predicted value as calculated by the regression equations, and the average residual, which is the average of all the absolute residual values. The Organization for Economic Cooperation and Development guidelines have required that the average residual of each model suitable for calculating $\log K_{ow}$ data for testing chemicals should be less than 0.30 log units [30]. Table 4 provides the predicted values, the residuals of each compound, and the average residuals obtained using each method. All of the regression equations obtained are at the 95% confidence level. The relationships of the $\log K_{ow}$ values, determined experimentally using the shake-flask method, and the $\log K_{ow}$ values, calculated using Equations 2 through 5, respectively, are shown in Figures 1 through 4.

Even a brief examination of the average residuals shows that the RP-HPLC method, the average residual of which is 0.336, does not meet the requirement of the Organization for Economic Cooperation and Development guidelines. The fundamental cause of the unsuitability of the RP-HPLC method



Figs. 1–4. Relationships of the measured $\log K_{ow}$ values with the predicted $\log K_{ow}$ values calculated from Equations 2 through 5, respectively. The Y axis represents the observed $\log K_{ow}$ values, and the X axis represents the predicted $\log K_{ow}$ values.

for prediction is the difference in partition mechanism between the octanol/water system and the RP-HPLC system. With its average residual of 0.295, the LSER model barely meets the requirement of the guidelines, whereas the MCIs and the quantum chemical descriptors can be used to reliably predict log K_{ow} values, with their average residuals reaching 0.176 and 0.240, respectively. The accuracy of the MCIs is better than that associated with the quantum chemical descriptors method. The accuracy in prediction along with the ease of calculation makes the MCIs method the best prediction model (among the four methods) for these heterocyclic nitrogen compounds.

CONCLUSIONS

In this study, log K_{ow} values of 14 heterocyclic nitrogen compounds (using the shake-flask method) were reported. An investigation of the partition mechanism of these compounds through the RP-HPLC method and the three QSAR prediction models (LSER, MCIs, and quantum chemical descriptors) demonstrated that molecular bulk properties and hydrogen bonding interaction are the most important influencing factors in the partition process.

In order to establish the prediction equation for log K_{ow} values using the four methods, we obtained four regression equations using stepwise variable regression. According to the requirement of the 0.30 log units average residual proposed by the Organization for Economic Cooperation and Development guidelines for log K_{ow} prediction model of testing chemicals, the RP-HPLC method is not acceptable, with its average residual reaching 0.336. The fundamental cause of this unsatisfactory result is the difference between the RP-HPLC system and the octanol/water system. The LSER model is hardly acceptable, with its average residual reaching 0.295. The quantum chemical descriptors method is acceptable, with its average residual 0.240, whereas the MCIs method is the most accurate, with the average residual of 0.176. The accuracy and the ease of obtaining the descriptors make the MCIs method the best predictive model for the studied compounds.

Acknowledgement—This project was funded by the National Natural Scientific Foundation of China.

REFERENCES

- Sabljić A, Gusten H, Verhaar H, Hermens J. 1995. QSAR modelling of soil sorption: Improvements and systematics of log K_{oc} vs log K_{ow} correlations. *Chemosphere* 31:4489–4514.
- Bodor N, Ganyi Z, Wong CK. 1989. A new method for estimation of partition mechanism. *J Am Chem Soc* 111:3783–3786.
- Karickhoff SW. 1981. Semi-empirical estimation of sorption of hydrophobic pollutants on natural sediments. *Chemosphere* 10: 833–846.
- Kenaga EE. 1980. Correlation of bioconcentration factors of chemicals in aquatic and terrestrial organisms with their physical and chemical properties. *Environ Sci Technol* 14:553–556.
- Konemann H. 1980. Structure–activity relationships and additivity in fish toxicities of environmental pollutants. *Ecotoxicol Environ Saf* 4:415–421.
- Shinkichi S, Nanao W, Toshiaki K, Takayuki S. 1993. Pyridine and pyridine derivatives. In Barbara E, Stephen H, William R, Gail S, eds, *Ullmann's Encyclopedia of Industrial Chemistry*, 5th ed, Vol A22. Poly (Vinyl Esters) to Reduction. VCH, New York, NY, USA, pp 399–429.
- Brauman T. 1986. Determination of hydrophobic parameters by reversed-phase liquid chromatography: Theory, experimental techniques, and application in studies on quantitative structure–activity relationships. *J Chromatogr* 373:191–225.
- Kamlet M, Abraham M, Taft R. 1983. Linear solvation energy relationships. 23: A comprehensive collection of the solvatochromic parameters, π^* , α , β , and some methods of simplifying the generalized solvatochromic equation. *J Org Chem* 48:2877–2887.
- Kamlet MJ, Doherty RM, Abboud JL, Abraham MH, Taft RW. 1986. Solubility—A new look. *Chemtech* 16:566–576.
- Leahy DE. 1986. Intrinsic molecular volume as a measure of the cavity term in linear solvation energy relationships: Octanol-water coefficients and aqueous solubilities. *J Pharm Sci* 75:629–636.
- Kaliszan R. 1992. Quantitative structure–retention relationships. *Anal Chem* 64:619–631.
- Kier LB, Hall LH. 1986. *Molecular Connectivity in Structure Activity Analysis*. Research Studies Press, Chichester, UK.
- Seybold PG, May M, Bagal UA. 1987. Molecular structure property relationships. *J Chem Educ* 64:575–581.
- Lewis DFV. 1989. The calculation of molecular polarizabilities by the CNDO/2 method: Correlation with the hydrophobic parameters, Log P. *J Comput Chem* 10:145–151.
- Nevalainen T, Kolehaminen E. 1994. New QSAR models for polyhalogenated aromatics. *Environ Toxicol Chem* 13:1699–1706.
- Organization for Economic Cooperation and Development. 1987. *OECD Guideline for Testing of Chemicals*. Organization for Economic Cooperation and Development, Paris, France.
- Hickey JP, Passino-Reader DR. 1991. Linear solvation energy relationships: “Rules of Thumb” for estimation of variable values. *Environ Sci Technol* 25:1753–1760.
- Kier LB, Hall LH. 1976. *Molecular Connectivity in Chemistry and Drug Research*. Academic, New York, NY, USA.
- Bahnick DA, Doucette WJ. 1988. Use of molecular connectivity indices to estimate soil sorption coefficients for organic chemicals. *Chemosphere* 17:1703–1715.
- Tripos Associates. 1985. *A Molecular Modeling System for the IBM PC* (computer program). St. Louis, MO, USA.
- Stewart JJP. 1990. *MOPAC 6.0*. Frank J Seiler Research Laboratory, U.S. Air Force Academy, Colorado Springs, CO.
- Karch K, Sebastian I, Halasz I. 1976. Preparation and properties of reversed phases. *J Chromatogr* 122:3–16.
- Nahum A, Horvath C. 1981. Surface silanols in silica-bonded hydrocarbonaceous stationary phases I. Dual retention mechanism in reversed phase chromatography. *J Chromatogr* 203:53–63.
- Weber SG, Trampusch WG. 1983. Coation exchange characteristics of silica-based reversed-phase liquid chromatographic stationary phases. *Anal Chem* 55:1771–1775.
- Gusten H, Horvatic D, Sabljic A. 1991. Modelling N-octanol/water coefficients by molecular topology: Polycyclic aromatic hydrocarbons and their alkyl derivatives. *Chemosphere* 23:199–213.
- Sabljić A, Gusten H, Hermens J, Opperhulzen A. 1993. Modelling octanol/water partition coefficients by molecular topology: Chlorinated benzenes and biphenyls. *Environ Sci Technol* 27:1394–1402.
- Herman RB. 1971. Theory of hydrophobic bonding. I. The solubility of hydrocarbons in water, within the context of the significant structure theory of liquids. *J Phys Chem* 75:363–368.
- Herman RB. 1971. Theory of hydrophobic bonding. II. The correlation of hydrocarbon solubility in water with solvent cavity surface area. *J Phys Chem* 76:2754–2759.
- Amidon GL, Yalkowsky SH, Valvani SC. 1975. Solubility of nonelectrolytes in polar solvents. V. Estimation of the solubility of aliphatic monofunctional compounds using a molecular surface area approach. *J Phys Chem* 79:2239–2246.
- Organization for Economic Cooperation and Development. 1981. Partition coefficients (N-octanol/water, shake-flasking method). In Organization for Economic Cooperation and Development, *OECD Guidelines for Testing of Chemicals*, Vol 107. Paris, France.
- Li L, Yang H, Ding Y, Wang L, Zhang Z. 1999. Prediction of log k_w' using MCIs and LSER methods. *J Liq Chromatogr Relat Technol* 22:897–907.