ADME Evaluation in Drug Discovery. 2. Prediction of Partition Coefficient by Atom-Additive Approach Based on Atom-Weighted Solvent Accessible Surface Areas

T. J. Hou and X. J. Xu*

College of Chemistry and Molecular Engineering, Peking University, Beijing 100871, China

Received January 17, 2003

A novel method for the calculations of 1-octanol/water partition coefficient (log *P*) of organic molecules has been presented here. The method, SLOGP v1.0, estimates the log *P* values by summing the contribution of atom-weighted solvent accessible surface areas (*SASA*) and correction factors. Altogether 100 atom/group types were used to classify atoms with different chemical environments, and two correlation factors were used to consider the intermolecular hydrophobic interactions and intramolecular hydrogen bonds. Coefficient values for 100 atom/group and two correction factors have been derived from a training set of 1850 compounds. The parametrization procedure for different kinds of atoms was performed as follows: first, the atoms in a molecule were defined to different atom/group types based on SMARTS language, and the correction factors were determined by substructure searching; then, *SASA* for each atom/group type was calculated and added; finally, multivariate linear regression analysis was applied to optimize the hydrophobic parameters for different atom/group types and correction factors in order to reproduce the experimental log *P*. The correlation based on the training set gives a model with the correlation coefficient (*r*) of 0.988, the standard deviation (*SD*) of 0.368 log units, and the absolute unsigned mean error of 0.261. Comparison of various procedures of log *P* calculations for the external test set of 138 organic compounds demonstrates that our method bears very good accuracy and is comparable or even better than the fragment-based approaches. Moreover, the atom-additive approach based on *SASA* was compared with the simple atom-additive approach based on the number of atoms. The calculated results show that the atom-additive approach based on *SASA* gives better predictions than the simple atom-additive one. Due to the connection between the molecular conformation and the molecular surface areas, the atom-additive model based on *SASA* may be a more universal model for log *P* estimation especially for large molecules.

INTRODUCTION

Log *P* has been widely used as a measure of hydrophobicity or lipopilicity, which is the ratio of a chemical's concentration in the *n*-octanol phase to its concentration in the aqueous phase of a two-phase system at equilibrium.[1,2] Pioneering work by Hansch and Leo has led to the use of log *P* in quantitative structure−activity relationships (QSARs), as a general description of cell permeability.[3] Up to now, log *P* has been widely used to access biological properties relevant to drug action, cellular uptake, metabolism, bioavailability, and toxicity.

The advent of combinatorial chemistry and the increasing applications of QSAR evaluation have been increasing requirement for fast and accurate theoretical estimation of log *P* and other relevant molecular properties. Computational methods for the assessment of log *P* date back to 1964 when Fujita et al. correlated differences between benzene and substituted benzenes to experimental data of log *P* and extrapolated these correlations for predicting log *P* in other series, leading later on to the development of the CLOGP method.[4] Since then, tremendous efforts in the past by theoretical chemists led to several useful computational methods for estimating log *P* values of organic compounds. At present, the most widely accepted method is classified as the "additive method", where a molecule is dissected into basic fragments (functional groups or atoms) and its log *P* value is obtained by summing the contributions of each fragment. According to the basic units dissected, the additive method can be divided into two categories: the fragment-based methods and the atom-based approaches. The fragment-based method originated with Rekker and co-workers[5,6] has become a standard calculation procedure and is available in many common software packages. This method involves the estimation of log *P* based on the contributions of functional groups and fragments attached to a base molecule. Current popular fragment-additive methods include CLOGP,[7,8] KLOGP,[9] KOWWIN,[10] CHEMICALC-2,[11] etc. The atom-based method was developed by Broto and later refined by Ghose and co-workers. This method assigns to the individual atoms in the molecular additive contributions to molecular log *P*. Atom-additive methods include ALOGP,[12−14] SMILOGP,[15] XLOGP,[16] and VLOGP.[17] There are also methods that try to incorporate molecular properties into the calculation, such as HINT[18] and ASCLOGP.[19] Moreover, several attempts have been made to calculate log *P* from free energies by molecular dynamics or Monte Carlo simulations,[20−22] but their extension to larger systems is limited by available computer resources and predictive precision.

In our previous work of the relationships between the brain−blood concentration ratio (*BB*) of 96 structurally diverse compounds with a great number of structurally

* Corresponding author e-mail: xiaojxu@chem.pku.edu.cn.

10.1021/ci034007m CCC: $25.00
Published on Web 04/11/2003

derived descriptors,[23] we found that log *P* was very crucial to log *BB*. When we constructed the linear correlation models of log *BB*, the ALOGP approach proposed by Crippen et al.[12−14] was used to calculate log *P* using the Cerius2 molecular simulations package.[24] The reliability of the commercial software prevents us to develop an automatic program software to estimate *BB* as a high throughput fashion. So the first aim of the paper is to develop a simple procedure to estimate log *P* as an automatic fashion, thus the program can be embedded into other programs developed by us to predict log *BB* or other ADME properties concerned with log *P*. Recently, we reported an additive-constitutive approach to predict aqueous solvation.[25,26] Our method is based on atom-weighted solvent accessible surface area (*SASA*). Moreover, we also found that the solvation model based on the addition of *SASA* is much better than the solvation model based on the simple atomic addition of the number of atoms (*NA*). Especially for large molecules such as proteins, the solvation model based on simple atomic contributions using *NA* nearly does not bear any predictive ability. So the second aim of this paper tries to construct a prediction model of log *P* based on *SASA*. And we also want to know if the prediction model of log *P* based on *SASA* is better than that based on *NA* for organic molecules.

METHODS

Data Set. The reliability of the training set is crucial to the accuracy of the final empirical model for log *P* calculation. Among all the 1850 compounds in the training set, the structures of the former 1831 ones were obtained from the combined collections of Suzuki and Kudo[11] and Klopman.[9] These 1831 compounds have also been used as a training set in developing the XLOGP model.[16] Since there were only several phosphorus-containing compounds in the data set afforded by refs 9 and 11, we have added some additional phosphorus-containing compounds. Here, the experimental log *P* values were obtained from Hansch and Leo's compilation.[27] It should be noted that all compounds in the training set do not include any metal atoms or ions. The molecular geometries of all compounds were fully minimized using molecular mechanism with MMFF force field.[28] The models were then saved in a MACCS sd database for further analysis. The molecules and the experimental log *P* values are listed in Table A. Table A and the MACCS sd database file are available in the Supporting Information.

Atom Typing Rules and Solvent Accessible Surface Area (*SASA*) Calculations. The final atom classification system has 100 basic atom/group types. It is comprehensive for the elements commonly found in organic molecules (C, H, O, N, P, S, and halogens). To allow for portability and simple implementation of the classification system, all atom types are presented in SMARTS strings (in Table 1). The atom types represented by SMARTS strings were determined by using the SMARTS system included in OELib.[29] SMARTS is a language that allows you to specify substructures using rules that are straightforward extensions of SMILES. In fact, almost all SMILES specifications are valid SMARTS targets. As SMILES, in SMARTS one can use atomic and bond symbols to specify a graph. However, in SMARTS the labels for the graph's nodes and edges (its "atoms" and "bonds") are extended to include "logical operators" and special atomic and bond symbols; these allow SMARTS atoms and bonds to be more general. Using SMARTS, flexible and efficient substructure-search specifications can be made in terms that are meaningful to chemists. In the current work, a parameter file was used to store the SMARTS strings defined for all atom/group types. If we want to add some new typing rules or modify the typing rules, we only need to make minor modifications to this parameter file.

If the log *P* of a molecule is calculated from the simple atomic contributions of *NA*, it can be described by

$$\log P = \sum_i n_i a_i \tag{1}$$

where a_i is the contribution of atom type *i*, and n_i is the number of atoms with atom type *i* in a molecule. The contribution for each atom type was determined by using the multiple linear correlations. Equation 1 has been widely used in most atom-additive approaches.

In the current work, we applied another atom-additive model, and the log *P* values are not simple contributed from the number of atoms with atom type *i*, while from the total *SASA* of atom type *i*. So the log *P* of a molecule is described as

$$\log P = \sum_i b_i s_i \tag{2}$$

where b_i is the contribution of atom type *i*, and s_i is the total *SASA* for atom type *i*. Molecular solvent accessible surface areas were calculated using the MSMS program,[30] and the probe radius was set to 0.5 Å with density of 3.0 vertex/Å2. In the calculations, the surface component for each atom was outputted.

Correction Factors. For many compounds, the model described by eq 2 can give reasonably good results. But the whole is often more than the sum of its parts, and it has become apparent in log *P* calculations. Usually, for various compounds, the log *P* values obtained by summing the atom/fragment contributions alone deviate significantly from the experimental values. This is sometimes explained by the inter- or intramolecular group−group interactions. In the current work, to consider the intermolecular hydrophobic interactions and intramolecular hydrogen bonds, we introduced two correction factors.

(1) Hydrophobic Carbon. For many compounds with hydrocarbon chains, their hydrophobicities are often underestimated by only using the summation of atomic contributions alone. The large deviation between experimental and predicted hydrophobicity may be introduced by the aggregation of these compounds in aqueous phase. This correction factor has been widely introduced by most additive methods.[16] Here, we defined the sp^3- or the sp^2-hybridized carbon without any attached heteroatom (any atom other than carbon) with the 1−4 relationship as the "hydrophobic carbon" (see Figure 1). It should be noted that sp^2-hybridized aromatic carbons were not considered as hydrophobic carbons. Moreover, the sp^2-hybridized carbon in the ring was also not considered as a hydrophobic carbon, because the sp^2-hybridized carbon in the ring was relatively rigid and it was not easy to adjust conformation to form aggregation.

(2) Intramolecular Hydrogen Bond. As being well-known, the intramolecular hydrogen bond in a compound

Table 1. Atom Typing Rules and Their Contributions to log *P* in SLOGP

type	description[a]	no. of compds	freq of use	contribution 1[d]	contribution 2[e]	type	description[a]	no. of compds	freq of use	contribution 1[d]	contribution 2[e]
				sp³ Carbon in							
1	**C**H$_4$(π=0),[b] **C**H$_3$R(π=0), **C**H$_2$R$_2$(π=0), **C**HR$_3$ (π=0)	565	1507	0.0103	−0.1021	6	**C**A$_2$X$_2$	4	4	−0.0566	−0.7663
						7	**C**AX$_3$	80	87	−0.0404	−0.4750
						8	**C**X$_4$	8	8	−0.0018	−0.1718
2	CH$_3$R,CH$_2$R$_2$,CHR$_3$ ($\pi\neq$0) A$_3$−**C**−C=R	414	542	0.0102	−0.1297	9	**C**H$_3$X	344	486	−0.0018	−0.4927
						10	**C**H$_2$AX, **C**H$_2$X$_2$	564	804	−0.0080	−0.4586
3	A$_3$−**C**−C=[N,O,S]	284	325	−0.0093	−0.5937	11	**C**HA$_2$X	190	348	−0.0036	−0.2718
4	A$_3$−**C**−C≡[C,N]	10	11	−0.0226	−0.7072	12	**C**H$_2$AX, **C**H$_2$X$_2$	75	75	−0.0103	−0.1822
5	**C**A$_3$X	30	32	−0.3533	−0.2801						
				sp² Carbon in							
13	R=**C**H$_2$	41	52	0.0095	0.0240	22	[O,N]=**C**A$_2$	86	88	0.0566	0.2512
14	R=**C**HA	85	130	0.0074	−0.0099	23	A−**C**OO	203	211	0.0352	0.0089
15	R=**C**HX	29	32	−0.0018	−0.1420	24	O=**C**C$_{sp2}$[c]	58	68	0.0299	−0.1111
16	N=**C**HA	8	8	0.0221	0.0629	25	O=**C**H−c, O=**C**H−n	169	185	0.0628	0.2858
17	R=**C**H−c, R=**C**H−(C=C)	48	58	0.0250	0.2423	26	O=**C**(A)-N	203	242	0.0053	−0.1933
18	R=**C**A$_2$	10	10	−0.1267	−0.4647	27	O=**C**HA	22	22	0.0253	−0.0080
19	R=**C**(A)-c, R=C(A)-C=*, R=**C**(A)-C≡*	8	10	−0.0303	−0.1129	28	O=**C**HX	14	14	0.0054	−0.3979
						29	[O,N]=**C**X$_2$	174	176	0.0244	0.0095
20	R=**C**(A)-C=O	12	15	−0.0867	−0.4331	30	S=**C**H$_2$, S=**C**HA, S=**C**A$_2$	12	12	−0.0354	0.2134
21	A=**C**AX, A=**C**X$_2$	20	23	0.0300	0.0443	31	O=S−**C**A$_3$, O=P−**C**A$_3$	22	24	−0.0244	−0.7622
				sp¹ Carbon in							
32	R≡**C**H	4	4	0.0011	−0.0229	33	A≡**C**A	6	8	0.0505	0.6135
				Aromatic Carbon in							
34	c···**c**H···c	1382	6103	0.0166	0.1154	41	X$_r$···**c**(X)···X$_r$	83	87	0.0595	0.6371
35	c···**c**H···X$_r$	254	362	−0.009	−0.3929	42	c···**c**(c)···c	165	309	0.0549	0.2835
36	A$_r$···**c**H···X$_r$, A$_r$···**c**(A)···X$_r$	76	93	0.0234	0.3106	43	c···**c**(F)···c	43	53	−0.0011	−0.0355
37	c···**c**(R)···c	738	943	0.0177	0.0525	44	c···**c**(Cl)···c	176	270	0.0507	0.2518
38	c···**c**(X)···c	1017	1552	0.0034	−0.0701	45	c···**c**(Br)···c	60	77	0.0660	0.2647
39	c···**c**(R)···X$_r$	57	63	−0.0411	−0.4528	46	c···**c**(I)···c	34	35	0.0767	0.4389
40	c···**c**(X)···X$_r$	143	186	0.0191	0.0263						
				sp³ Oxygen in							
47	R−**O**H	201	287	−0.0314	−0.5419	50	R−**O**−R	308	369	−0.0015	0.0882
48	c−**O**H	200	218	0.0093	−0.0791	51	A−**O**−C=O	444	463	0.0061	0.1049
49	N−**O**H	10	10	−0.0575	−0.7673	52	R−**O**−X, X−**O**−X	28	60	0.0098	0.1144
				sp² Oxygen in							
53	**O**=C	774	923	−0.0546	−0.2391	55	**O**=[N,O,P]	106	201	−0.0574	−1.6017
54	**O**=c	74	103	−0.0938	−0.8095	56	**o**	12	12	0.0388	0.3368
				Hydrogen in							
57	H	1838	16182	0.0587	0.2239	59	**H**−N	692	1217	−0.0438	−0.4089
58	**H**−OH, **H**−SH	406	520	−0.0671	0.0000	60	**H**−OC=O	219	230	−0.0712	−0.1697
				sp³ Nitrogen in							
61	R(π=0)−**N**H$_2$	46	46	−0.0303	0.1165	65	X−**N**H−R, X−**N**H−c, X−**N**H−X	410	4	0.0974	0.8524
62	R(π=0)−**N**H−R(π=0)	35	36	−0.0607	−0.0624						
63	**N**R$_3$(π=0)	40	46	−0.1887	−0.2975	66	**N**R$_2$X, **N**RX$_2$, **N**X$_3$	24	27	−0.2057	−0.9097
64	X−**N**H$_2$	5	5	−0.0352	−0.0760						
				sp² Nitrogen in							
67	N	51	62	−0.0356	0.0583	73	O=C−**N**−R(π=0)	110	111	−0.0375	0.0364
68	c−**N**H$_2$, R(π=0)−**N**H−c	190	203	−0.0135	0.3777	74	O=C−**N**A$_2$	4	6	−0.2054	−0.2618
69	c−**N**H−c	9	9	0.0193	0.4608	75	O=C−**N**A−R(π=0)	21	21	0.0590	−0.0593
70	R(π=0)−**N**−R$_2$($\pi\neq$0), NR$_3$($\pi\neq$0), R$_2$(π=0)−**N**−R($\pi\neq$0)	14	14	0.3811	0.6875	76	O=C−**N**−R$_2$(π=0)	36	36	−0.3997	−0.3826
						77	R−**N**=R	30	30	−0.0889	−0.4186
						78	X−**N**=R	44	45	0.0572	0.7570
71	O=C−**N**H$_2$	89	95	−0.0077	0.3266	79	X−**N**=X	4	8	0.0191	0.1769
72	O=C−**N**H	133	154	0.0393	0.4108	80	A$_{2-}$**N**−S=O, A$_{2-}$**N**−P=O	75	87	−0.0412	0.0457
				Aromatic Nitrogen in							
81	n	220	243	−0.0084	0.1309	83	**n**···c···n	144	359	−0.0564	−0.2981
82	**n**···n	32	52	0.0019	0.0570	84	**n**···n···n	13	16	−0.0283	−0.3096
				Sulfur in							
85	A−**S**H	5	5	0.0214	0.6413	88	A−**S**O−A	5	5	0.0027	0.7619
86	**S**A$_2$	77	85	0.0320	0.7129	89	A−**S**O$_{2-}$A	81	88	0.1834	2.8385
87	**S**=R	23	23	0.0128	−0.1006						
				Phosphorus in							
90	P	29	29	0.1444	0.6216						
				Halogens in							
91	F	140	333	0.0419	0.5425	93	Br	87	110	0.0218	0.8958
92	Cl	238	421	0.0230	0.7049	94	I	40	41	0.0246	1.0340

Table 1 (Continued)

type	description[a]	no. of compds	freq of use	contribution 1[d]	contribution 2[e]	type	description[a]	no. of compds	freq of use	contribution 1[d]	contribution 2[e]
				United Atom Types							
95	A−**NO$_2$**	28	84	−0.0029	−0.0362	98	−**NO**	37	80	0.0063	0.2041
96	c−**NO$_2$**	114	387	0.0083	0.1176	99	−**NCS**	23	72	0.0260	0.5191
97	−**CN**	79	168	0.0010	0.0141	100	−**NH$_2$**, −C**OO**H	14	42	−0.0927	0.7949

[a] Description: R represents any group linked through carbon; A represents any atom except hydrogen; X represents any heteroatom (O, N, S, P, and halogens); c represents aromatic carbon; n represents aromatic nitrogen; X$_r$ represents aromatic atom except aromatic carbon; o represents aromatic oxygen; − represents single bond; = represents double bond; ≡ represents triple bond; ··· represents aromatic bond. The atom described is shown in bold. [b] π=0 represents that the atom has π electrons; $\pi\neq0$ represents that the atom has not π electrons. [c] sp^2 represents the hybridized state. [d] The hydrophobicity using the atom-additive approach based on *SASA*. [e] The hydrophobicity using the simple atom-additive approach based on the number of atoms.

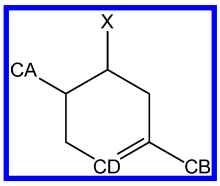

Figure 1. The definition of hydrophobic carbons. Here X represents a heteroatom. According to our definition, CB is a hydrophobic carbon, CB is not because a heteroatom is within four atoms, and CD is not because CD is sp^2-hydridization and in a six-membered ring.

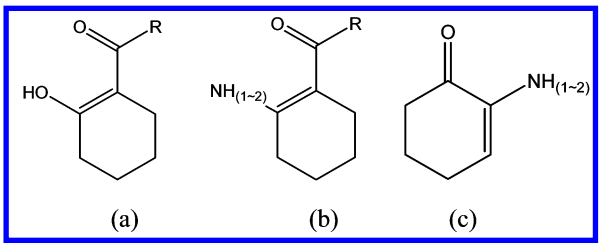

Figure 2. The three kinds of intramolecular hydrogen bonds considered in SLOGP.

may weaken the electrostatic interactions between compound and water and thus increase the hydrophobicity of a molecule. It should be noted that the hydrogen bonds between organic compounds and water are relatively significant, so only a very strong intramolecular hydrogen bond can effectively influence the H-bond interactions between organic compound and water. Here, we only defined intramolecular hydrogen bonds for three kinds of compounds shown in Figure 2, and we think that intramolecular hydrogen bonds in these compounds are strong enough to effect the interaction between compounds and water. We have tried other definitions in our analysis, but they do not work as well as this definition. The number of intramolecular hydrogen bonds in each compound is calculated by using substructure searching.

After including two correction factors, the log *P* is described as

$$\log P = \sum_i b_i s_i + \sum_j c_j B_j \qquad (3)$$

where b_i and c_i are regression coefficients, s_i is the total *SASA* of aotm type *i*, and B_j is the number of the correction factor of atom type *j*.

RESULTS AND DISCUSSION

The program, SLOGP v1.0, was developed in C++. The program can read a single molecule or multiple molecules (represented in single SYBYL/mol2 file, single MACCS/mol file, SYBYL/mol2 database file, or MACCS/sd database file), performs atom typing and surface calculation, detects correction factors, and then calculates log *P* using the parameters from multiple linear regression analysis. For each molecule, the estimation of log *P* takes about 0.5 s on an SGI O2 R10000 workstation. So our program can screen a large database and construct the subdatabase meeting the required range of log *P* values. Soon, the SLOGP program will be embedded into our ADME prediction program as a subroutine.

Prediction of log *P*. The initial model was based on the summation of contribution of *SASA* as eq 2, in which a total of 100 atom types were used. This model yielded fairly satisfactory results, $n = 1850$, $r = 0.985$, $s = 0.414$, $F = 585.989$, which was comparable to or even better than those obtained by other methods using similar strategies. From the prediction, we found that for many hydrocarbons or compounds with long hydrophobic aliphatic chains and some compounds with intramolecular hydrogen bonds, the predicted values deviate much from the experimental values. In such cases, we introduced two correction factors to account for the possible intra- or intermolecular interactions. The two correction factors are found to be statistically significant in multivariate regression analysis. The final model for log *P* calculations was obtained by correlating the total *SASA* of 100 atom/group types and the frequencies of two correction factors with the experimental log *P* values. The list of contributions for the final fit and number of molecules containing each atom type and correction factors is included in Tables 1 and 3. The final model with eq 3 produced better results than those with eq 2: $n = 1850$, $r = 0.988$, $SD = 0.368$, $F = 702.218$, which bears comparative statistical significance with the latest CLOGP model ($n = 12\,546$; $r = 0.986$; $SD = 0.30$).[31] Figure 3 shows the correlation between the experimental and calculated log *P* values. Figure 4 shows a histogram of the deviation of the calculated values from the experimental results, where a near-Gaussian error distribution curve centered as zero can be seen. To further test the robust of the model, we have performed leave-one-out cross-validation on the whole training set, which give nearly the same results with the multivariate regression analysis ($q = 0.983$). The experimental and calculated log *P* values using the final model are summarized in Table A in the Supporting Information. The final model predicts well for most of the 1850 compounds in training set, but as listed in Table 4, 29 compounds showed deviations greater than 1.0 log unit. Now

Table 2. Prediction Model Using Different Probe Radius

	1	2	3	4	5	6	7	8	9[a]
probe radius (Å)	0.5	0.6	0.7	0.8	0.9	1.0	1.2	1.4	
r	0.988	0.988	0.988	0.988	0.988	0.987	0.987	0.987	0.987
SD	0.368	0.375	0.372	0.372	0.376	0.381	0.381	0.383	0.378
F	707.218	682.813	694.670	693.924	677.485	665.420	665.404	666.408	679.418
unsigned mean error	0.261	0.264	0.264	0.263	0.267	0.269	0.269	0.270	0.273

[a] The statistical significance of the model from the simple atom-additive approach based on the number of atoms.

Table 3. Correction Factor Used in SLOGP

description	contribution	no. of affected compds	*r*	*F*	*SD*
Not Any Correction Factor					
			0.985	582.298	0.409
Hydrophobic Carbon					
1[a]	0.162	71	0.986	609.074	0.398
2[b]	0.192	369	0.987	658.891	0.383
Intramolecular H-Bond					
1[c]	0.792	36	0.986	612.896	0.397
2[d]	0.756	44	0.986	615.449	0.396
3[e]	0.688	60	0.986	617.534	0.395
4[f]	0.635	63	0.986	605.365	0.399
5[g]	0.592	64	0.986	609.687	0.398
6[h]	0.509	71	0.986	606.348	0.399

[a] The correction factor of "hydrophobic carbon" was applied to all hydrocarbons. [b] The correction factor of "hydrophobic carbon" was applied to all organic molecules. [c] The correction factor of "intramolecular hydrogen bond" was applied to a in Figure 2. [d] The correction factor of "intramolecular hydrogen bond" was applied to a and b in Figure 2. [e] The correction factor of "intramolecular hydrogen bond" was applied to a, b, and c in Figure 2. [f] The correction factor of "intramolecular hydrogen bond" was applied to a, b, and c in Figure 2 and d and e in Figure 5. [g] The correction factor of "intramolecular hydrogen bond" was applied to a, b, and c in Figure 2 and d, e, f, and g in Figure 5. [h] The correction factor of "intramolecular hydrogen bond" was applied to all kinds of possible hydrogen bonds defined in Figures 2 and 5.

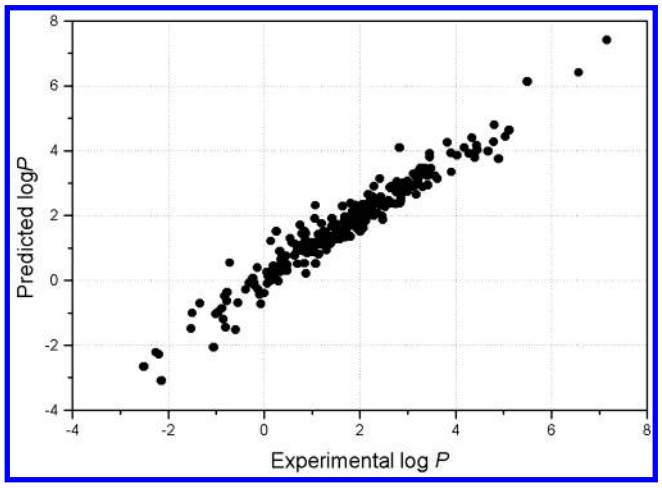

Figure 3. Correlation between the experimental and calculated log *P* values of 1850 compounds in the training set.

we cannot give an exact explanation of these deviations. They may be brought by experimental errors, inadequate atom-typing rules, or insufficient correction factors. Wang et al. have used the former 1831 compounds in this training set to develop the XLOGP model.[16] From the calculations of XLOGP, 42 compounds in the training set show deviation of 1.0 log unit. From the number of compounds with large

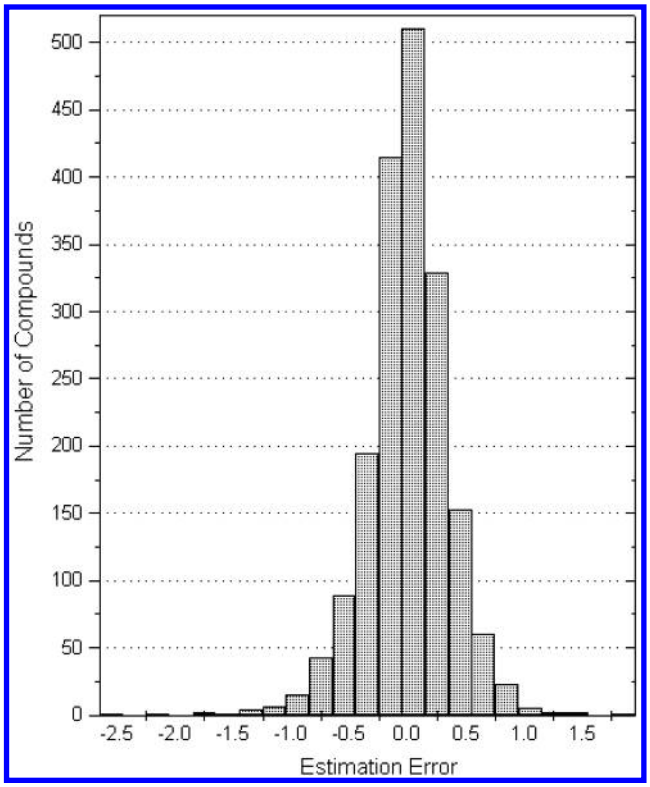

Figure 4. Distribution histogram of the estimation errors.

predicted errors, it seems that our SLOGP model works better than the XLOGP model.

Probe Radius. When we use a different radius of probe to calculate the molecular surface, the total *SASA* of the atom type *i* in a molecule and the obtained coefficient should be quite different. In our previous work, we have used eq 2 to calculate the aqueous solvation free energy, and we found that smaller probe radius could produce better correlation.[25] The reason is that if we use a large probe radius, the contributions of some interior atoms without accessible surface areas are neglected. Here, to reduce the shield of the hydrogen atoms to the interior atoms, in the calculations of the solvent accessible surface area, the van der Waals radius of hydrogen was manually adjusted from 1.2 to 0.9 Å.

Here, the influence of the probe radius to the calculated results was investigated, and a different probe radius from 0.5 to 1.4 Å was used (Table 2). From the calculated results, we found that a smaller probe radius could give a better model, but the effect is not very significant. Only from the mean unsigned errors and the standard deviations, a probe radius of 0.5 Å is the best, so a probe radius of 0.5 Å was applied for *SASA* calculations. The results are generally in good agreement with those in our previous work.[24] But we also found that the effect of the change of probe radius to

Table 4. Compounds with Large Calculation Errors

no.	name	log P_{actual}[a]	log P_{calcd}[b]	residue
442	Ado	1.05	2.05	1.00
451	DDAPR	−0.52	−1.61	1.09
499	5-ethyl-6-azauracil	0.22	−0.82	1.04
524	2,4,5-tribromoimidazole	1.96	2.97	−1.01
554	Pyridazine	−0.72	0.55	−1.27
557	2-pyrimidone	−1.62	−0.59	−1.03
576	2-methyl-2-imidazoline	−0.52	−0.63	1.15
688	picolinic acid	−1.50	0.59	−2.09
737	6-aminonicotinamide	−0.70	−0.68	1.38
757	pentylenetetrazole	0.14	1.22	−1.08
876	benzohydroximic acid	0.26	1.51	−1.25
906	2,4-diaminobenzoicacid	−0.31	0.87	−1.18
966	2-trifluoromethyl-5,6-dinitro-benzimidazole	3.89	2.07	1.82
967	4,5,6,7-tetrachloro-2-methyl-benzimidazole	2.83	4.10	−1.27
983	m-trifluoromethyltrifluoromethane-sulfonanilide	4.50	3.08	1.42
1392	benzoylacetone	2.52	1.40	1.12
1435	fusaric acid	0.68	2.34	−1.66
1594	hydrazobenzene	2.94	4.33	−1.39
1601	sulfsomidine	−0.33	0.68	−1.01
1647	niflumicacid	1.59	4.16	−2.57
1676	1-(2-SO2Et-5-CF3-phenylhydrazono)-1-cyanoaceticacidmethylester	4.22	2.91	1.31
1722	chlorambucil	1.70	3.44	−1.74
1740	1-methyl-4-phenyl-7-chloro-quinazolin-2-one	2.36	3.63	−1.27
1759	ambrosin	1.03	2.07	−1.04
1793	desipramine	4.90	3.75	1.15
1813	buquinolate	2.18	3.65	−1.47
1820	progesterone	3.26	4.46	−1.20
1821	pipamperone	1.07	2.31	−1.24
1824	etorphine	1.86	3.06	−1.20

[a] log P_{actual} is the experimental value. [b] log P_{calcd} is the predicted value.

log *P* is less obvious than that to aqueous solvation free energy. The reason is that we adopted a smaller van der Waals radius of hydrogen, and the interior atoms linked with hydrogen element bear wider exposure and are not very sensitive to the probe radius of solvent probe.

Atom Typing Rules. The solvation free energy of a molecule transferring from vacuum to water or *n*-octanol includes two parts: the electrostatic contribution and the nonpolar contribution. The latter contribution is usually modeled as proportional to the solvent accessible surface area. In a simple atom-additive approach based on *NA*, the electrostatic and nonpolar parts are actually taken into account implicitly using different atom types. So we should guarantee that the atoms belonging to the same atom type have similar *SASA* and charge densities. In principle, different atoms bear different partial charges. But if two atoms are located in similar chemical environments, the partial charges and *SASA* should be similar. According to the above assumption, the definition of atom types may be the most important thing in prediction of log *P*. Based on the above discussions, we defined the atom types listed in Table 1. The classification scheme differentiates atoms according to (i) element, (ii) hybridization state, and (iii) nature of the neighboring atoms. This establishes the rough theoretical support for the assumption that a certain type of atom has a specific contribution to the partition coefficient.

For the definition of the atom typing rules, we think two aspects should be considered. First, the atom types for the elements N and O should be carefully defined, because these two kinds of elements bear strong polarity and relatively complicated chemical environments. Besides the elements C and H, the elements N and O may be the most important constituent composition in organic molecules. Second, we should carefully define the atom types in the conjugate systems. The atom types in the conjugate systems show obvious irregularity due to the charge flow along the conjugate systems. For example, many compounds in the training set possess a small conjugated ring with nitrogen atoms. We know that the charge distribution of the aromatic ring with different number of nitrogen atoms should be quite different. For example, in a pyridine ring, there is only one nitrogen atom; while in a pyrazine ring, there are two conjugate nitrogen atoms. If we only define one atom type for the nitrogen atoms in a pyridine ring and a pyrazine ring, the nitrogen atoms in these two rings are actually forced to be equivalent, and the contribution of the nitrogen atoms in pyrazine to the hydrophobicity is two times that of one nitrogen in pyridine. But in fact, due to the conjugate effect, the partial charges on the nitrogen atoms in pyrazine are quite different from those on the nitrogen atom in pyridine. So the nitrogen atoms in pyridine and pyrazine should be defined to different atom types. In the old atom typing rules, we only defined one atom type for nitrogen in an aromatic ring neighboring with two connected atoms. We found that the calculated results for some compounds with heterorings were not very good. Thus we defined several new types to represent the nitrogen atom in a conjugate six-ring with different neighboring atoms (see type 81, 82, 83, and 84 in Table 1). Using the new atom typing rules, the mean unsigned error was decreased from 0.267 to 0.261. The above analyses show that the atom types in the conjugate system should be carefully defined. We think that the insufficient consideration of the conjugate systems may be one of the most important factors that influence the quantity of the prediction model of log *P*.

Any additive method, either by fragment or atom, needs a relevant scheme for fragment/atom classification. The quality of such a classification scheme can be evaluated by how well the calculated log *P* values agree with their experimental counterparts. To some extent, an additive method is the art of fragment/atom classification. The definition of atom types should be suitable, two few atom types may not effectively represent the different chemical environments. In recent work of Wildman and Crippen, the authors present a new atom type classification system with 68 atom types for use in atom-based calculation of log *P* and molar refractivity (MR).[32] The 68 atomic contributions to log *P* have been determined by fitting an extensive training set of molecules, with $r^2 = 0.918$ and $SD = 0.677$. The model proposed by Wildman et al. is obviously significantly worse than that proposed by us. The reason is that the number of basic types used by Wildman et al. is so few. Certainly, we do not mean that more atom types can produce better results. Sometimes, the two or several atom types are not fully independent, and addition of redundant atom types cannot effectively enhance the prediction of the model. The number of atom types used here is smaller than Ghose's set of 110[14] and much smaller than Broto's set of 222.[33] However, using less atom types does not weaken the power of our model which yields satisfactory results even when we use the

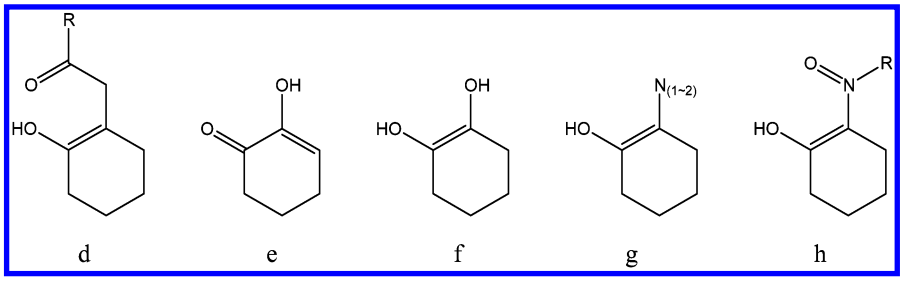

Figure 5. The other five kinds of intramolecular hydrogen bonds in our fitting process.

addition of basic atomic contribution alone. We think that after our iterative adjustment of atom types we have developed each of them into a more elaborate class for log *P* calculation.

It should be noted that the coefficients obtained by the SLOGP model are quite different from those obtained by other methods of log *P* based on *NA*, because they depend not only on the number of atoms but also on the total *SASA*. For example, the presence of a halogen atom generally increases the hydrophobicity by I > Br > Cl > F. But the coefficient of F (0.0419)) is quite smaller than those of Cl (0.0230), Br (0.0218), I (0.0246), because the *SASA* of F is quite larger than those of the other three kinds of halogen atoms. If we use the additive model of eq 1, the coefficient of halogens should be F (0.5425) < Cl (0.7049) < Br (0.8958) < I (1.0340) (see Table 1), which is in good agreement with the general concept.

Correction Factors. In the current work, we only used two simple correction factors. The consideration of these two correction factors can significantly improve the linear regression of the models. Overall, the standard deviation of the entire training set is reduced to 0.368, and the unsigned mean error is decreased from 0.294 to 0.261.

The correction factor of "hydrophobic carbon" is very important in our model. First, we only introduced this correction factor to aliphatic and aromatic hydrocarbons. After considering the hydrophobic carbon, the correlation of the model was improved obviously (see Table 3). But from the prediction of model, we also found that some compounds of heteroatom-containing series, especially those bearing long aliphatic chains, were greatly underestimated. We think these long aliphatic chains can also introduce intermolecular aggregation. So we extended the hydrophobic carbon to all kinds of organic compounds: if there is no heteroatom at a certain range, a carbon atom is a "hydrophobic carbon". Adopting the new concept, the correlation of the model was further improved, and the multivariate regression analysis generate a model with $r = 0.987$, $SD = 0.383$, and $F = 658.891$.

The introduction of the correction factor of "intramolecular hydrogen bond" can significantly improve the correlation of the model. But we found that not all possible intramolecular hydrogen bonds should be considered. Here, we defined that either the donor or the acceptor atom should be linked directly to a ring, and the ring serves to immobilize the orientations of the donor and the acceptor. Moreover, the two neighboring atoms in the ring should be sp^2 hybridized, which ensures that the intramolecular hydrogen bond could form a plane five- or six-membered ring. In our fitting process, we totally considered eight kinds of intramolecular hydrogen bonds (see Figures 2 and 5), but our calculated results show that only three kinds of intramolecular hydrogen bonds in Figure 2 should be included in linear fitting (see Table 3). The calculated results in Table 3 are not strange, because the three kinds of hydrogen bonds in Figure 2 are stable than those in Figure 5. The contribution of this correction factor is 0.688, which is very close to that reported by Leo, 0.63.[34]

Comparison with Other Methods. Only from the correlation between the experimental log *P* and the calculated values, our SLOGP model is very significant. But it is well-known that the actual prediction power may only be determined based on a list of compounds as the test set. Moreover, we want to know if our model can give comparative prediction with other log *P* calculation procedures.

The test compounds used here were cited from Mannhold et al.[35] The database of 138 test compounds comprises 90 simple organic structures and 48 chemically heterogeneous drug molecules (beta-blockers, class I antiarrhythmics, and neuroleptics). The reason that we selected this set of compounds is that Mannhold et al. compared 14 calculation procedures by comparing their predictions to these 138 compounds with experimental log *P* values from the literature. The methods compared by Mannhold et al. are well-established, commercially available procedures, which can be roughly grouped into three categories: atom-based, fragment-based, and conformation-dependent approaches.

The molecular models of the 138 compounds were built using Cerius2, stored in MACCS/sd format, and then subjected to calculation. The predicted log *P* values of these compounds are obtained by using SLOGP. The experimental and calculated log *P* values are listed in Table B in the Supporting Information. The calculated results for the comparison of 14 log *P* calculation procedures studied by Mannhold et al. were directed cited from ref 35, and the predicted power of our SLOGP method was judged according to the criteria used by Mannhold et al.: (1) The individual estimation errors are grouped using Mannhold's: errors less than ±0.50 are considered as acceptable; errors greater than ±0.50 and less than ±1.00 are considered as disputable; and errors exceeding ±1.00 are considered as unacceptable. The missing calculations are also counted. All these results are given as a percentage of the entire test set. (2) The experimental and the calculated log *P* values are correlated using linear regression analysis. The statistical results (i.e. *r*, *SD*, and *F*-value) are recorded. The mean squared deviations (MSD) are also calculated. All the results are summarized in Table 5.

Table 5 shows that the SLOGP model can give very good results to these compounds in test set. The correlation coefficient ($r = 0.974$) is higher those of the other approaches except the KOWWIN model. From the mean square devia-

Table 5. Comparison of 15 log *P* Procedures for the Prediction of the Test Set

method	acceptable[a]	disputable[b]	unacceptable[c]	uncalculated[d]	MSD[e]	*r*[f]	*SD*[g]	*F*[h]	ref
Atom-Based Methods									
MOLCAD	68.1	20.3	11.6	0.0	0.334	0.932	0.439	911	33
Tsar2.2	68.1	20.3	11.6	0.0	0.345	0.937	0.438	987	14
PROLOG_atom 5.1	76.8	14.5	7.2	1.4	0.262	0.947	0.431	1164	14
SMILOGP	49.3	24.6	18.8	7.2	0.551	0.917	0.588	660	18
SLOGP 1.0	90.6	7.2	2.2	0.0	0.107	0.974	0.327	2510	
Fragment-Based Method									
PROLOGP_cdr 5.1	76.8	16.7	5.1	1.4	0.199	0.957	0.448	1472	40
∑f-SYBYL	81.9	13.8	4.3	0.0	0.200	0.959	0.444	1583	41
SANALOG	79.7	15.2	3.6	1.4	0.167	0.967	0.402	1919	41
PROLOGP_comb 5.1	81.2	15.2	2.2	1.4	0.184	0.960	0.387	1582	14, 41
CLOGP 4.34	84.8	10.1	3.6	1.4	0.156	0.965	0.398	1849	1, 8
KLOGP	84.1	13.8	0.7	1.4	0.134	0.966	0.362	1859	9
KOWWIN	90.6	5.8	3.6	0.0	0.113	0.974	0.334	2517	10
CHEMICALC-2	68.8	17.4	13.8	0.0	0.418	0.926	0.535	827	11
Conformation-Dependent Methods									
HINT	68.1	15.9	13.8	2.2	0.454	0.912	0.682	665	18
ASCLOGP	55.1	28.3	15.2	1.4	0.583	0.873	0.771	431	19

[a] Percentage of acceptable results (estimation error $< \pm 0.50$). [b] Percentage of disputable results (estimation error $> \pm 0.50$ and $< \pm 1.00$). [c] Percentage of unacceptable results (estimation error $> \pm 1.00$). [d] Percentage of uncalculated results. [e] Mean squared deviations. [f] Correlation coefficient between the experimental and calculated log *P* values. [g] Standard deviations. [h] Fisher values.

tion and standard deviation, the SLOGP model is better than the other models. From the calculated results, our method performs much better than the other four atom-based methods and gives comparative results with the fragment-based approach with the best performance. It should be noted that in 1995, Leo pointed out that the log *P* values for several compounds in the test sets might exist serious errors. We know that if the solute cannot be measured at a pH where it is essentially uncharged, then if an accurate pK_a is available, a correction can be made for the amount of neutral form present. For example, propafenone (compound **102** in Table B) has a pK_a of 9.62 and was partitioned at pH 5.0. So the hydrophobicity of the neutral propafenone should be greatly overstated. The CLOGP method estimates it at 3.55 rather than 4.63. Our SLOGP method give a predicted value of 3.59, which is quite close to the value predicted by CLOGP. Similar to propafenone, the hydrophobicity of disopyramide (compound **93**) should exist as a large error. It is interesting to find that the predicted value by SLOGP (3.74) also exists as a large difference with the experimental value (2.58). When the two drugs without reliable neutral values are removed from the linear regression, the predicted ability of SLOGP was improved further ($n = 136$, $r = 0.987$, $SD = 0.300$, $F = 2970.40$). Moreover, not considering these two possible outliers, the prediction to these 138 compounds also may not give a decisive rank of all these log *P* calculation procedures because the number of compounds in the test set is rather limited. But the comparison at least demonstrates that SLOGP gives the best results among all these methods and yielded acceptable estimations for the tested compounds.

In Mannhold's work, he concluded that the predictive power of the calculation procedures should be ranked as fragment-based > atom-based > conformation-dependent approaches according to the calculated results. From principle, the fragment-based approach can give better consideration of the electrostatic distribution than the atom-based ones. From this point of view, we also think that the fragment-based approach is more superior than the atom-based one. But we do not think that log *P* calculations will be completely dominated by the fragment-based methods because the atom-based methods have some characteristics, while the fragment-based methods do not have. First, for a fragment-based method, the classifications of the basic fragments are very difficult, and an additive method will not be able to do the calculation for any compound containing a "missing fragment". So, sometimes the fragment-based approach may not calculate the compounds with undefined fragment. But for an atom-based method, the description of an atom type is very simple. Second, the programming and extension of the atom-based approach is much simpler and easier than the fragment-based approach. For example, in our program, based on OElib, the definition and the determination of the atom types are very simple, and the whole program is very short and easy to be interpreted. Third, in some applications of hydrophobic parameters such as the molecular lipophilicity potential (MLP) approaches, the use of atom-centered parameters is preferred.

Last, it should be noted that although the training set used here is largely cited from Wang et al.,[16] the SLOGP model is quite different from the XLOGP model. First, in our method, we used a more effective way to define atom types. We defined two parameters files: one named def.dat saving atom-typing rules and the other named suf.dat saving hydrophobicity for each atom type. When users want to define new atom types and get new parameters for newly defined atom types, they only need to give little changes to these two files. While in the XLOGP method, the atom typing rules are hidden in the main program, and the original model is very difficult or even impossible to be modified or extended. Second, SLOGP only includes a single program written in C++, and all libraries used are employed from OElib. So our method can be simply embedded into other program. Third, because XLOGP adopted the atom typing rules in SYBYL, it can only support SYBYL/MOL2 format, while SLOGP supported many kinds of formats of molecular structures. The last, the SLOGP model, was obtained by correlating *SASA* of 100 atom types with the experimental log *P*, not *NA* of atom types used by XLOGP and other atom-based approaches.

Comparison of Method Based on *NA* and That Based on *SASA*. In the above several sections, we discussed the model based on *SASA* according to eq 3. But in most atom-based additive methods, we only simply correlated *NA* of atom types with the experimental log *P* according to eq 1. Here, we also proposed a model based on eq 1 (see model 9 in Table 2). The model based on *NA* is worse than the best model based on *SASA*, but the difference is not very significant. It seems that both of the model based on *NA* and the model based on *SASA* can generate good results. Actually, for small organic molecules, these two kinds of models do not exist in large differences, because nearly all atoms in small compounds are exposed to solvent. In fact, even in the method based on *NA*, the atoms with the same atom types should have not only similar solvent accessible surfaces but also similar charge densities. But if we use the method based on *SASA*, we only need to guarantee that the atoms with the same atom types have similar charge densities. For example, in most methods for log *P* calculations, the aliphatic carbons atoms with different hydrogen atoms should be defined as different atom types, and the hydrophobicity of carbon atoms clearly decreases in the following order: −CH3, −CH2−, −CH<, >C<. In fact, the order of the hydrophobicity is mainly induced by the different exposure of an atom, which leads to different electrostatic and van der Waals interactions between solvent and solute. Here, in the method based on *SASA*, we only define one atom type for the aliphatic carbon, and the difference of the carbon atoms with different hydrogen atoms are considered by different *SASA*.

It should be noted that since the conformation of a molecule may have a considerable influence on its liphophilic nature,[38] the method based on *SASA* should be a more universal model especially for large molecules such as peptide. For the method based on *NA*, all atoms with the same atom types are considered equivalently, which means all atoms should be exposed to solvent. But we know if some atoms in a molecule are surrounded by other atoms and located in the interior of a molecule for example some atoms in peptide or protein, these atoms contribute little or even nothing to hydrophobicity. If we use the method based on *NA*, these interior atoms are also equivalently considered as the other atoms with the same atom types exposed to solvent. But if we use the method based on *SASA*, the contribution of the exposed atoms and the interior atoms can be separately effectively. In our previous work,[39] we compared the performance of the methods based on *NA* and *SASA* the solvation free energy. We found that for small organic molecules, these two methods did not exist in large differences if defined as suitable atom types, but for large molecules the predictions of these two methods were quite different; the calculated results by the method based on *SASA* could give good correlation with the method by PBSA, but those by the method based on *NA* did not show any regularity. Similarly, for the prediction of log *P*, the method based on *SASA* should be more universal than the method based on *NA*, especially for large molecules.

CONCLUSION

In the current work, the prediction model of log *P*, SLOGP, was proposed to estimate the partition coefficient of solutes in octanol/water, log *P*, automatically. The log *P* of a molecule can be calculated based on different atom types, corresponding *SASA*, and hydrophobicity parameters. In our work, the definition of atom types was based on SMARTS language in order to be unambiguous and allow for simple portability. The prediction using the model gives an average unsigned error of 0.261 and standard deviation of 0.368. In this model, 100 atom types and two correction factors were used to classify the various atoms in a molecule. Compared to other methods, our method give comparative or even better results.

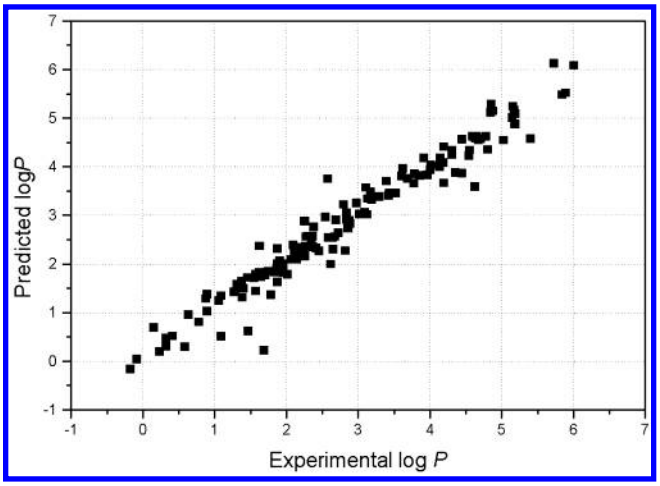

Figure 6. Correlation between experimental and calculated log *P* values of 138 compounds in the test set.

For small organic molecules, our model based on *SASA* gives better results than the model based on *NA*. Moreover, the model based on *SASA* is a more universal model especially for large molecules. The methods proposed here and all the parameters for calculations on log *P* have been incorporated into a computer program called SLOGP. The SLOGP computer code can be obtained by contacting the authors. In SLOGP, two sets of hydrophobicity parameters are afforded: surf.parm and atom.parm. surf.parm is used for log *P* using the method based on *SASA*, and atom.parm is used for log *P* using the method based on *NA*. The SLOGP program has been tested on IRIX and Linux operation systems.

Supporting Information Available: The experimental and calculated log *P* values for molecules of the training set (Table A), the experimental and calculated log *P* values for molecules in the test set (Table B), and the structures of the training databases and the test set are saved in MACCS/sd database format (the sd database files include the experimental log *P* values of all compounds). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Hansh, C.; Leo, A. J. *Substituent constants for correlation analysis in chemistry*; Wiley: New York, 1979.
(2) Pliska, V.; Testa, B.; Waterbeemd, H. *Lipophilicity in drug action and toxicology*; Pliska, V., Testa, B., Waterbeemd, H., Eds.; VCH Publishers: New York, 1996.
(3) Leo, A.; Hansch, C.; Elkins, D. Partition Coefficients and Their Uses. *Chem. Rev.* **1971**, *71*, 525−616.
(4) Fujita, T.; Iwasa, J.; Hansch, C. *J. Am. Chem. Soc.* **1964**, *86*, 5175.
(5) Nys, G. G.; Rekker, R. F. *Chim. Ther.* **1973**, *8*, 521.
(6) Nys, G. G.; Rekker, R. F. *Eur. J. Med. Chem.* **1974**, *9*, 361.
(7) Hansch, C.; Leo, A. *Substituent Constants for Correlation Analysis in Chemistry and Biology*; Wiley: New York, 1979.

(8) Leo, A. *CLOGP*, version 3.63; Daylight Chemical Information Systems: Irvine, CA, 1991.
(9) Klopman, G.; Li, J.-Y.; Wang, S.; Dimayuga, M., Computer automated logP calculations based on extended group contribution approach. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 752−781.
(10) Meylan, W. M.; Howard, P. H. atom fragment contribution method for estimating octanol−water partition-coefficients. *J. Pharm. Sci.* **1995**, *84*, 83−92.
(11) Suzuki, T.; Kudo, Y. Automatic log *P* estimation based on combined additive modeling methods. *J. Comput.-Aid. Mol. Des.* **1990**, *4*, 155−198.
(12) Ghose, A. K.; Crippen, G. M. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure−activity relatioinships. I. partition coefficients as a measure of hydrophobicity. *J. Comput. Chem.* **1986**, *7*, 565−577.
(13) Ghose, A. K.; Crippen, G. M. Atomic physicochemical parameters for three-dimensional- structure-directed quantitative structure−activity relationships. 2. Modeling dispersive and hydrophobic interactions. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 21−35.
(14) Ghose, A. K.; Pritchett, A.; Crippen, G. M. Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure−activity relatioinships. III. Modeling hydrophobic interaction. *J. Comput. Chem.* **1988**, *9*, 80−90.
(15) Convard, T.; Dubost, J.-P.; Le Solleu, H. *Quant. Struct.-Act. Relat.* **1994**, *13*, 34.
(16) Wang, R. X.; Fu, Y.; Lai, L. H. A new atom-additive method for calculating partition coefficient. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 615−621.
(17) Gombar, V. K.; Enslein, K. Assessment of *n*-octanol/water partition coefficient: When is the assessment reliable. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1127−1134.
(18) Kellogg, G. E.; Semus, S. F.; Abraham, D. J. HINT: a new method of empirical hydrophobic field calculation for CoMFA. *J. Comput.-Aid. Mol. Des.* **1991**, *5*, 545−552.
(19) Abraham, D. J.; Kellogg, G. E. The effect of physical organic properties on hydrophobic fields. *J. Comput.-Aid. Mol. Des.* **1994**, *8*, 41−49.
(20) Debolt, S. E.; Kollman, P. A. Investigation of structure, dynamics, and solvation in 1-octanol and its water-saturated solution − molecular-dynamics and free-energy perturbation studies. *J. Am. Chem. Soc.* **1995**, *117*, 5316−5340.
(21) Eksterowicz, J. E.; Miller, J. L.; Kollman, P. A. Calculation of chloroform/water partition coefficients for the *N*-methylated nucleic acid bases. *J. Phys. Chem.* **1997**, *101*, 10971−10975.
(22) Tafazzoli, M.; Jalili, S. Absolute partition coefficients for organic solutes by using Monte Carlo simulations in chloroform/water system. *Chem. Phys. Lett.* **2000**, *331*, 235−242.
(23) Hou, T. J.; Xu, X. J. ADME Evaluation in Drug Discovery. 1. Applications of Genetic Algorithms on the Prediction of Blood-brain Partitioning of a Large Set Drugs. *J. Mol. Model.* **2002**, *8*, 337−349.
(24) *Cerius2 4.5;* Molecular Simulation Inc.: San Deigo, CA, U.S.A., 2001.
(25) Hou, T. J.; Qiao, X. B.; Zhang, W.; Xu, X. J. Empirical Aqueous Solvation Models Based on Accessible Surface Areas with Implicit Electrostatics. *J. Phys. Chem. B* **2002**, *106*, 11295−11304.
(26) Hou, T. J.; Xu, X. J. Aqueous solvation models based on accessible surface area calculations. *Acta Phys. Chim. Sin.* **2002**, *18*, 1052−1056.
(27) Hansch, C.; Leo, A.; Hoekman, D. *Exploring QSAR: hydrophobic, electronic, and steric constants*; American Chemistry Society: Washington, DC, 1995; Vol. 2.
(28) Halgren, T. A. Merck molecular force field .1. Basis, form, scope, parametrization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490−519.
(29) James, C. A.; Weininger, D.; Delany, J. Daylight Theory Manual Daylight 4.62; Daylight Chemical Information Systems, Inc.: Los Altos, 2001.
(30) Sanner, M. F.; Olson, A. J.; Spehner *J. Biopolymers* **1996**, *38*, 305−320.
(31) Leo, A. J. personal communication, **2003**.
(32) Wildman, S. A.; Crippen, G. M. Prediction of physicochemical parameters by atomic contribution. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868−873.
(33) Broto, P.; Moreau, G.; Vandycke, C. Molecular structures: perception, autocorrelation descriptor and SAR studies. *Eur. J. Med. Chem.* **1984**, *19*, 71−78.
(34) Leo, A. Calculating $\log P_{oct}$ from structures. *Chem. Rev.* **1993**, *93*, 1281−1306.
(35) Mannhold, R.; Dross, K. Calculation procedures for molecular lipophilicity: a comparative study.*Quant. Struct.-Act. Relat.* **1996**, *15*, 403−409.
(36) Furet, P.; Sele, A.; Cohen, N. C. 3D molecular lipophilicity potential profiles: a new tool in molecular modeling. *J. Mol. Graph.* **1988**, *6*, 182−189.
(37) Gaillard, P.; Carrupt, P. A.; Testa, B.; Boudon, A. Molecular lipophilicity potential, a tool in 3D QSAR: method and applications. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 83−96.
(38) Hopfinger, A. J.; Battershell, R. D. Application of SCAP to drug design. 1. Prediction of octanol−water partition coefficients using solvent-dependent conformational analyses. *J. Med. Chem.* **1976**, *19*, 569−573.
(39) Hou, T. J.; Zhang, W.; Qiao, X. B.; Huang, Q.; Xu, X. J. An extended aqueous solvation models based on atom-weighted solvent accessible surface areas, SAWSA v2.0. *J. Phys. Chem. B* Submitted for publication.
(40) Rekker, R. F. *The Hydrophobic Fragment Constant*; Elsevier: New York, 1977.
(41) Rekker, R. F.; Mannhold, R. *Calculation of Drug Lipophilicity*; VCH: Weinheim, 1992.
(42) Leo, A. J. *Chem. Pharm. Bull.* **1995**, *43*, 512−513.

CI034007M