Whole-Molecule Calculation of Log P Based on Molar Volume, Hydrogen Bonds, and Simulated ^{13}C NMR Spectra

Laura K. Schnackenberg and Richard D. Beger*

Division of Systems Toxicology, National Center for Toxicological Research, Food and Drug Administration, Jefferson, Arkansas 72079

Received November 29, 2004

The prediction of Log P is usually accomplished using either substructure or whole-molecule approaches. However, these methods are complicated, and previous whole-molecule approaches have not been successful for the prediction of Log P in very complex molecules. The observed chemical shifts in nuclear magnetic resonance (NMR) spectroscopy are related to the electrostatics at the nucleus, which are influenced by solute−solvent interactions. The different solvation effects on a molecule by either water or methanol have a strong effect on the NMR chemical shift value. Therefore, the chemical shift values observed in an aqueous and organic solvent should correlate to Log P. This paper develops a rapid, objective model of Log P based on molar volume, hydrogen bonds, and differences in calculated ^{13}C NMR chemical shifts for a diverse set of compounds. A partial least squares (PLS) model of Log P built on the sum of carbon chemical shift differences in water and methanol, molar volume, number of hydrogen bond donors and acceptors in 162 diverse compounds gave an r^2 value of 0.88. The average r^2 for 10 training models of Log P made from 90% of the data was 0.87 ± 0.01. The average q^2 for 10 leave-10%-out cross-validation test sets was 0.87 ± 0.05.

INTRODUCTION

In the early stages of drug discovery, an important consideration is the partitioning of the drug between octanol and water. This partition coefficient is used as a descriptor of the potential pharmacokinetic behavior as well as interactions between a drug and its biological macromolecular counterpart. The partitioning between solvents is generally reported as a Log P value where P is the partition coefficient between octanol and water. Hansch and Fujita[1] first put forth the use of Log P as a descriptor of hydrophobicity. Since then, the value of Log P has become an important consideration in the design of new drugs. While Log P can be determined experimentally, this requires the prior synthesis of the compound. Therefore, numerous programs have been developed to allow the determination of Log P based on structural considerations including CLOG P,[2−6] KOWWIN,[7,8] AB/Log P[9] and SciLog P[10] to name a few. SciLog P is the only method that takes a whole-molecule approach toward the calculation of Log P, while the other methods are substructure approaches. Mannhold and Petrauskas[11] compared the whole-molecule approach to the substructure approaches for the calculation of Log P noting that the substructure methods performed better as a whole. Recently, Cheuman et al.[12] calculated Log P based on ab initio molecular orbital (MO) parameters and the accessible surface area of solute molecules.

Generally, the measurement of experimental Log P values has been accomplished using reversed-phase high-pressure liquid chromatography[13−15] or reversed-phase thin-layer chromatography.[16,17] The partition coefficient can be considered to be a quantitative descriptor of the combined solute and solvent enthalpic and entropic physical interactions. It has been stated that quantum mechanical methods operating on the molecule, while it is solvated by water in one case and by octanol in the other are required to calculate Log P accurately from structure alone.[6] One possible reason that the whole-molecule approaches have not been successful for complex structures is because of the complexity of quantum mechanical approaches and the fact that the solute−solvent interactions are described by a Boltzmann's distribution.[18] NMR chemical shifts represent the quantum mechanical energy of a nucleus in a magnetic field, and this energy is based on a Boltzmann's distribution of solute−solvent energy states. A quantum mechanical calculation of the molecule and solvent interaction will be based on electrostatic and kinetic energies. NMR chemical shifts are quantum mechanical energies that are composed of a diamagnetic term and a paramagnetic term.[19] The diamagnetic term is directly related to the electrostatic potential at the nucleus. The paramagnetic term is related to the electron orbitals surrounding the nucleus and usually does not change with respect to solvent−solute interactions. Therefore, changes in chemical shifts due to solvent can be directly related to the electrostatics at the nucleus and represent solvent−solute enthalpy changes in the calculation of Log P. The difference in measured chemical shifts from NMR spectroscopic analysis for a compound in an aqueous or organic environment should, therefore, allow for the prediction of Log P with values being highly influenced by the structure of the molecule as well as the electrostatic interaction environment including the solvation energies. In this study, we attempted to predict Log P values using the differences in NMR chemical shifts for a compound in water and in methanol. The modeling of Log

* Corresponding author phone: (870)543-7080; fax: (870)543-7686; e-mail: rbeger@nctr.fda.gov.

10.1021/ci049643e
Published on Web 03/01/2005

P by NMR spectra can be viewed as a quantitative spectroscopic data-property relationship (QSDPR). The QSDPR models are similar to previous quantitative spectroscopic data-activity relationship (QSDAR) models based on ^{13}C NMR data and enzyme binding data of steroids.[20,21] The major entropy contributions to Log P are related to the molar volume of the molecule, mixing of the solvent and the side-chain localization effects of hydrogen bonding between solvent and solute.

While Log P values are determined for the partitioning of a compound between water and octanol, octanol is not a commonly utilized solvent for NMR. Thus, our solvent options for chemical shift prediction were somewhat limited. Methanol (MeOH-d4) was chosen as a substitute for octanol as it has the most closely related polarizability and miscibility of octanol from the organic NMR solvents available and was a solvent with a large NMR database for prediction of the ^{1}H and ^{13}C spectra. To simplify the analysis, we chose to use Log P values for model molecules that were primarily composed of carbon, proton, nitrogen and oxygen atoms. One hundred and sixty-two compounds were utilized for the development of this model. This initial study proposes the use of sum of chemical shift differences in water and methanol, molar volume, and the number of hydrogen bond acceptors and donors as an alternative method for the prediction of Log P.

MATERIALS AND METHODS

The experimental values for Log P of 162 compounds studied were taken from Mannhold and Petrauskas[11] and Chuman et al.[12] and are shown in Table 1. ACD/Labs 8.0 CNMR software (ACD/Labs Toronto, Canada) was utilized for the prediction of the ^{13}C NMR chemical shifts of each compound in methanol and water. ACD/Labs 8.0 HNMR software (ACD/Labs Toronto, Canada) was utilized for the prediction of the ^{1}H NMR chemical shifts of each compound in water and methanol. In the CNMR and HNMR ACD/Labs software, the NMR spectra are calculated by a substructure similarity technique called HOSE,[22] which correlates similar substructures within spherical shells with similar NMR chemical shifts. The ACD/Labs HNMR predictor has a database that contains 2580 spectra in D_2O and 3991 spectra in MeOH-d4. The ACD/Labs CNMR predictor has a database that contains 2942 spectra in D_2O and 4367 spectra in MeOH-d4. Even with these large NMR spectral databases, only 7 compounds out of the 162 well-known small molecules had experimental spectra in both water and methanol available in the ACD/Labs HNMR and CNMR spectral databases. Only 8 more compounds had more than two of the four experimental proton and carbon spectra in the ACD/Labs spectral NMR databases. The molar volume was calculated using ACD/Labs 8.0 software (ACD/Labs Toronto, Canada).

The difference in chemical shift (ΔH or ΔC) between the two solvents was determined by subtracting the calculated chemical shifts in water from the calculated chemical shifts in methanol for a compound. For methyl or methylene groups, the chemical shift values determined in both water and methanol were multiplied by the number of equivalent protons prior to determining the difference in proton chemical shift differences between the two solvents. Hydroxyl (OH), amine (NH), and amide (NH2) proton chemical shifts were not used in the calculation of ΔH because the HNMR software did not affect the prediction of the chemical shifts based on the local structural environment for the hydroxyl protons.

In the models, the total contribution of the sum of OH, NH, or NH2 groups to the change in proton chemical shift was accounted for with the parameter, number of hydrogen bond donors (NHD), in the compound. We also added the parameter, number of hydrogen bond acceptors (NHA), to the model. The value for NHA was calculated by summing the number of keto, NO_2, and oxygen and nitrogen atoms that did not have an attached proton in a compound. Any compound that was determined to be amphiprotic by Chuman et al.[12] had the potential to accept or donate a proton in a hydrogen bond and was counted as having at least one hydrogen bond acceptor and one hydrogen bond donor group. The set of 162 compounds contained subsets of 42 compounds that had no hydrogen bond donor or acceptor groups, 55 compounds that had hydrogen bond donor groups, and 120 compounds that had hydrogen bond acceptor groups.

Forward multiple linear regression (MLR) analysis was done using Statistica Version 6.0 (Statsoft, Tulsa, OK) to evaluate Log P based on the molar volume, NHA, NHD, ΔC, and ΔH. A parameter was added to the MLR model of Log P by forward regression where the *F-value* increased by greater than 1.0 with the addition of that parameter to the model and the p-level of the parameter was less than 0.05.

General linear partial least squares (PLS) analysis was done using Statistica version 6.0 (Statsoft, Tulsa, OK) to evaluate Log P based on the parameters molar volume, ΔC, NHD, and NHA that a MLR model of Log P showed had significant p-levels. Evaluation of the PLS Log P models was done using leave-multiple-out cross-validation procedures in which compounds were systematically excluded from the training set and the training set used to develop a model (less any contribution from the excluded compound-(s)).[23] The cross-validated r^2 (termed q_n^2) term results from the predictions of Log P obtained by using the training set model to predict the cross-validation test set of compounds excluded from the training set. PLS models of Log P were developed using 90% of the data with the remaining 10% of the data used in cross-validation test sets. This was done a total of 10 times to ensure all the molecules were left out of a training set once. In these "leave-multiple-samples-out" experiments, the results of the 10 corresponding training sets and cross-validations sets were averaged.

RESULTS AND DISCUSSION

Methods that have been developed for the calculation of Log P use either substructure-fragment or whole-molecule approaches. In the substructure approach, a molecule is broken into single atoms or fragments with contributions to Log P calculated for each part and interactions between the atoms or fragments also considered in the determination of the overall Log P value. With the whole-molecule approach, the molecule is treated in its entirety. In both cases, however, the structure of the molecule is critical for the calculation of Log P. NMR is a spectroscopic technique in which the observed chemical shift values are directly related to the

Table 1. Compounds, Experimental Log P, Parameters Used in the Model of Log P, and Predicted Log P[a]

compound	Log P	molar vol (cm3)	ΔC (ppm)	NHD	NHA	pred Log P	compound	Log P	molar vol (cm3)	ΔC (ppm)	NHD	NHA	pred Log P
acetic acid	−0.17	56.1	−2.57	1	1	−0.04	isoxazole	0.08	65.4	−29.97	0	2	−0.45
imidazole	−0.08	60.9	−13.19	1	1	0.01	benzaldehyde	1.48	101.0	−5.32	0	1	1.57
propionic acid	0.33	72.6	−2.86	1	1	0.47	propionaldehyde	0.59	75.3	−6.76	0	1	0.77
aniline	0.90	91.7	−2.28	1	1	1.06	phenylacetonitrile	1.56	115.6	−3.09	0	1	2.04
benzyl alcohol	1.10	103.2	1.02	1	1	2.14	1-phenylacetone	1.44	135.8	−7.19	0	1	2.63
salicylamide	1.28	106.3	6.46	2	1	1.35	nitromethane	−0.35	57.8	−0.14	0	2	−0.41
4-hydroxybenzoic acid	1.56	100.3	0.91	2	1	1.11	2-phenylacetamide	0.45	123.0	−1.42	1	1	2.04
2,4-dihydroxybenzoic acid	1.64	98.8	6.27	3	1	0.86	phenylacetic acid	1.41	116.8	0.68	1	1	1.87
benzoic acid	1.87	101.9	−0.23	1	1	1.40	1-phenylmethanamine	1.09	109.4	−6.36	1	1	1.57
3-methylphenylacetic acid	1.95	133.1	−6.34	1	1	2.31	benzamide	0.64	108.1	−0.80	1	1	1.58
3-methoxybenzoic acid	2.02	125.9	−4.68	1	2	1.41	*N*-methyl benzamide	0.86	129.7	−0.22	1	1	2.26
2-naphthylamine	2.28	125.8	21.15	1	1	3.03	*N*-methylaniline	1.66	108.8	9.28	1	1	1.70
4-nitrotoluene	2.37	117.5	−1.87	1	1	1.87	phenylcyanamide	1.87	102.2	7.32	1	1	1.48
4-phenylbutyric acid	2.42	149.8	−7.28	1	1	2.81	*N*-phenylacetamide	1.16	122.5	−20.05	1	1	1.85
2-methylbenzoic acid	2.46	118.2	2.41	1	1	1.93	*N*-hydroxyaniline	0.79	89.8	10.16	2	1	0.87
1-naphthol	2.85	121.9	21.57	1	1	2.91	anilinobenzene	3.50	155.4	6.12	1	1	3.11
benzophenone	3.18	167.5	−6.98	0	1	3.61	phenol	1.46	87.8	9.24	1	1	1.05
naphthalene	3.30	123.5	8.60	0	0	3.09	ethanol	−0.31	59.0	1.14	1	1	0.09
4-butylbenzoic acid	3.97	167.7	−13.55	1	1	3.31	methanamine	−0.57	48.7	0.90	1	1	−0.24
pentamethylbenzene	4.56	170.8	−5.25	0	0	4.42	methanol	−0.77	42.5	−0.07	1	1	−0.44
bibenzyl	4.79	182.9	3.96	0	0	4.88	1-octanol	3.00	158.0	18.02	1	1	3.30
2-methylphenanthrene	4.86	173.9	15.67	0	0	4.71	ethanamine	−0.13	65.2	−2.03	1	1	0.25
pyrene	4.88	161.9	58.54	0	0	4.74	*N*-methylmethanamine	−0.38	70.3	−6.16	1	1	0.37
2-ethylanthracene	5.85	190.4	28.51	0	0	5.34	cyclohexanol	1.23	103.4	−2.07	1	1	1.43
benzene	2.13	89.4	−19.86	0	0	1.77	cyclohexamine	1.49	114.0	−13.53	1	1	1.65
toluene	2.73	105.7	1.38	0	0	2.47	1,2,3-triazole	−0.29	54.2	−7.74	1	2	−0.84
ethylbenzene	3.15	122.2	2.08	0	0	2.99	1,2,4-triazole	−0.58	54.2	−6.62	1	2	−0.83
styrene	2.95	115.3	6.43	0	0	2.82	4-nitroaniline	1.39	103.5	0.23	1	1	1.45
ethynylbenzene	2.53	107.4	−4.00	0	0	2.48	3-nitroaniline	1.37	103.5	−1.45	1	1	1.44
propylbenzene	3.72	138.7	5.53	0	0	3.53	2-nitroaniline	1.85	103.5	5.94	1	1	1.50
allylbenzene	3.23	133.8	0.24	0	0	3.33	4-hydroxyphenol	0.59	86.2	5.86	2	1	0.72
isopropylbenzene	3.66	139.5	14.79	0	0	3.64	3-hydroxyphenol	0.80	86.2	−1.52	2	1	0.65
n-butylbenzene	4.38	155.2	−6.12	0	0	3.93	2-hydroxyphenol	0.88	86.2	−22.16	2	1	0.46
tert-butylbenzene	4.11	156.1	6.45	0	0	4.08	4-nitrophenol	1.91	99.7	−6.75	1	1	1.27
cyclopropylbenzene	3.27	117.4	13.00	0	0	2.94	3-nitrophenol	2.00	99.7	2.16	1	1	1.35
biphenyl	4.01	154.7	−22.36	0	0	3.77	2-nitrophenol	1.79	99.7	4.45	1	1	1.37
benzylbenzene	4.14	168.8	21.68	0	0	4.61	4-nitrobenzoic acid	1.89	113.8	−9.61	1	2	0.99
ethane	1.81	61.5	−0.52	0	0	1.09	3-nitrobenzoic acid	1.83	113.8	−3.89	1	2	1.04
ethylene	1.13	58.1	5.20	0	0	1.04	2-nitrobenzoic acid	1.46	113.8	−5.78	1	2	1.02
acetylene	0.37	45.7	5.48	0	0	0.66	4-aminobenzoic acid	0.83	104.2	8.94	2	1	1.30
propane	2.36	78.0	2.00	0	0	1.62	4-aminophenol	0.04	90.1	4.84	2	1	0.83
propene	1.77	73.1	2.02	0	0	1.47	4-cyanobenzoic acid	1.56	111.0	−1.44	1	2	0.97
propyne	0.94	61.8	17.60	0	0	1.27	4-cyanophenol	1.60	97.2	1.27	1	1	1.27
n-butane	2.89	94.5	4.88	0	0	2.16	4-methoxybenzoic acid	1.96	125.9	−5.17	1	2	1.40
isobutane	2.76	94.9	14.50	0	0	2.26	2-methylpyridine	1.11	98.9	−3.75	0	1	1.52
butadiene	1.99	84.7	−6.84	0	0	1.75	2-cyanopyridine	0.40	92.8	−2.80	0	2	0.65
n-pentane	3.39	111.0	10.71	0	0	2.72	ethylpyridine-2-carboxylate	0.87	137.0	−0.33	0	3	1.34
neopentane	3.11	111.1	14.54	0	0	2.76	methylpyridine-2-carboxylate	0.43	120.5	−2.62	0	3	0.81
cyclopropane	1.72	53.2	3.39	0	0	0.87	*N*,*N*-dimethylpyridin-2-amine	1.65	120.6	−4.14	0	2	1.50
cyclopentane	3.00	88.7	3.35	0	0	1.97	2-ethoxypyridine	1.81	123.1	−6.61	0	2	1.55
cyclohexane	3.34	106.4	11.16	0	0	2.59	2-methoxypyridine	1.34	106.6	0.04	0	2	1.10
o-xylene	3.12	121.9	−2.78	0	0	2.94	2-acetylpyridine	0.83	114.1	6.03	0	2	1.39
m-xylene	3.20	121.9	−14.65	0	0	2.83	pyridin-2-ylmethyl acetate	0.39	135.4	0.27	0	3	1.30
p-xylene	3.15	121.9	1.34	0	0	2.97	2-ethylpyridine	1.60	115.4	0.78	0	1	2.07
hexamethylbenzene	4.61	187.0	0.00	0	0	4.97	2-phenylpyridine	2.65	147.9	−35.25	0	1	2.75
tetralin	3.49	136.2	13.52	0	0	3.53	pyrazine	−0.26	75.8	−23.44	0	2	−0.07
furan	1.34	72.2	8.88	0	0	1.51	2-methylpyrazine	0.21	92.1	−16.67	0	2	0.50
1-phenyl pyrrole	3.08	147.1	−4.15	0	0	3.70	2-cyanopyrazine	−0.01	85.7	−14.59	0	3	−0.37
anthracene	4.45	157.6	22.00	0	0	4.27	ethylpyrazine-2-carboxylate	0.28	130.2	−3.00	0	4	0.42
phenanthrene	4.46	157.6	15.80	0	0	4.21	methylpyrazine-2-carboxylate	−0.23	113.7	−5.88	0	4	−0.12
benzonitrile	1.56	99.9	−2.97	0	1	1.56	*N*,*N*-dimethylpyrazin-2-amine	0.93	113.8	−28.16	0	3	0.37
acetylbenzene	1.58	120.9	−10.89	0	1	2.14	2-ethoxypyrazine	1.28	116.3	−9.66	0	3	0.62
ethylbenzoate	2.64	143.8	−4.16	0	2	2.21	2-methoxypyrazine	0.73	99.8	−4.63	0	3	0.15
methylbenzoate	2.12	127.3	−4.06	0	2	1.70	2-acetylpyrazine	0.20	107.4	−5.35	0	3	0.38
N,*N*-dimethylaniline	2.31	127.4	13.08	0	1	2.56	2-ethylpyrazine	0.69	108.6	−11.01	0	2	1.06
nitrosobenzene	2.01	102.2	6.52	0	1	1.72	2-phenylpyrazine	2.06	141.1	−35.58	0	2	1.84
nitrobenzene	1.85	101.2	0.30	0	2	0.94	pyrimidine	−0.44	75.8	−12.69	0	2	0.03
ethoxybenzene	2.51	129.9	0.83	0	1	2.52	2-methylpyrimidine	−0.05	92.1	−14.30	0	2	0.52
methoxybenzene	2.11	113.4	3.85	0	1	2.04	2-cyanopyrimidine	0.08	85.7	−15.01	0	3	−0.38
acetonitrile	−0.34	54.9	−2.24	0	1	0.18	*N*,*N*-dimethylpyrimidin-2-amine	1.07	113.8	−14.70	0	3	0.49
acetone	−0.24	75.1	−6.71	0	1	0.76	2-ethoxypyrimidine	0.74	116.3	−18.32	0	3	0.54
dimethyl ether	0.10	67.9	−14.34	0	1	0.47	2-methoxypyrimidine	0.23	99.8	−14.07	0	3	0.07
cyclohexanone	0.81	102.9	1.81	0	1	1.70	4-methylpyrimidine	−0.07	92.1	−3.74	0	2	0.62

Table 1 (Continued)

compound	Log P	molar vol (cm3)	ΔC (ppm)	NHD	NHA	pred Log P	compound	Log P	molar vol (cm3)	ΔC (ppm)	NHD	NHA	pred Log P
ethylene epoxide	0.30	44.2	2.44	0	1	−0.11	4-ethoxypyrimidine	0.97	116.3	−12.78	0	3	0.59
pyridine	0.65	82.6	2.68	0	1	1.08	5-methylpyrimidine	0.01	92.1	−4.65	0	2	0.61
quinoline	2.03	116.7	6.61	0	1	2.17	5-cyanopyrimidine	−0.42	85.7	−11.24	0	3	−0.34
methyl acetate	0.18	81.5	−2.09	0	2	0.31	ethylpyrimidin-5-carboxylate	0.52	130.2	−5.54	0	4	0.39
ethylpropionate	1.21	114.5	−1.95	0	2	1.33	methylpyrimidin-5-carboxylate	0.03	113.7	−7.92	0	4	−0.14
pentan-2-one	0.91	108.1	−0.43	0	1	1.84	*N*,*N*-dimethylpyrimidin-5-amine	0.46	113.8	11.28	0	3	0.73
propionitrile	0.16	71.4	−8.98	0	1	0.62	5-ethoxypyrimidine	0.56	116.3	−18.27	0	3	0.54
nitroethane	0.18	74.3	−7.39	0	1	0.73	5-methoxypyrimidine	0.07	99.8	−15.17	0	3	0.06

[a] NHD stands for number of hydrogen bond donors, NHA stands for number of hydrogen bond acceptors, Δ*C* stands for summed differences in ^{13}C NMR chemical shifts and ppm stands for parts for million.

Table 2. Regression Summary of MLR Model of Log P

	B	Std. Error	p-level
molar volume	0.0309	0.0013	0.000000
NHA	−0.693	0.042	0.000000
NHD	−0.249	0.064	0.00016
Δ*C*	0.0092	0.0038	0.018
intercept	−0.8056	0.1635	0.000002

diamagnetic shielding of each proton and carbon nucleus in a molecule. The diamagnetic shielding term has contributions from electrostatic properties of the solvent and solute including electrostatic contributions from solvent−solute hydrogen bonding.[24−26] NMR chemical shifts, therefore, take into account the different electrostatic interactions that may occur between the solute and solvent. Based on this, it should be possible to utilize the differences in all the NMR chemical shifts for a molecule as an alternate means of calculating Log P.

In this study, Log P was predicted for a data set of 162 compounds based on the calculated differences in the ^{1}H and ^{13}C NMR chemical shift between methanol and water. Compounds were selected that contained primarily carbon, hydrogen, nitrogen and oxygen atoms in order to limit the number of structural factors and put the emphasis on the prediction of Log P using the NMR chemical shifts that could be predicted with the ACD HNMR and CNMR predictor software. Further, the compounds were chosen to cover a range of Log P values. Table 1 presents the values for experimental Log P, Δ*C*, molar volume, NHD, and NHA for the 162 compounds investigated.

The forward MLR model of Log P of the 162 compounds had an r^2 of 0.88. The *F-value* was 281.87, and the standard error of estimate was 0.494. The MLR equation for Log P is given below.

$$\text{Log P} = 0.0309*\text{MV} + 0.0092*\Delta C + 0.249*\text{NHD} - 0.694*\text{NHA} - 0.806 \quad (1)$$

The forward multiple linear regression analysis was developed using only significant variables. Table 2 shows the B coefficients, standard errors of the B coefficients and the p-level of each parameter selected for the MLR model of Log P. The Δ*H* term was not selected by forward MLR and was not used for the development of the PLS models of Log P. The fact that Δ*H* was not significant may be due to the fact that the predicted NMR chemical shifts were based on water and methanol and not water and octanol. Another reason that the Δ*H* term was not significant is that some of the compounds were not predicted using water or methanol

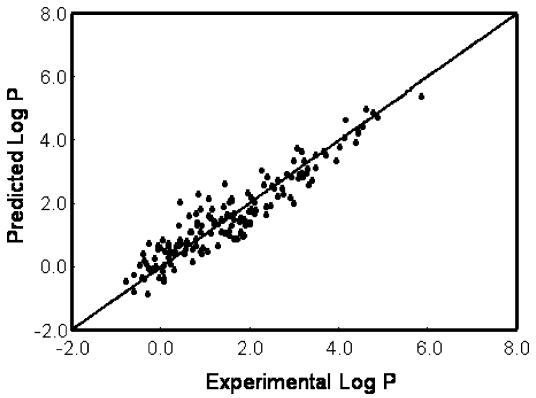

Figure 1. Plot of experimental Log P versus predicted Log P.

data alone. The Δ*C* term had a significant contribution ($p = 0.018$) to the MLR model of Log P. Carbon chemical shifts are not as dependent on the solvent interactions as proton chemical shifts because they have less direct solute−solvent interactions and therefore can be calculated more accurately than proton NMR chemical shifts. The parameters, NHD and NHA, are similar to the hydrogen bond acidity and hydrogen bond basicity parameters that were used by Abraham.[27] Unlike Abraham,[27] we did not use nonlinear combinations of the hydrogen bonding terms in building the models. When we added nonlinear combinations to the model of Log P, we were able to increase r^2 of the model; however, this results in a loss of physical understanding of our model of Log P. One could understand hydrogen bond nonlinearity by noting that a compound with three potential hydrogen bond donor and/or acceptor sites will not have three hydrogen bonds with the solute at the same time but will be more likely to have a hydrogen bond with the solute than a compound with only one hydrogen bond donor or acceptor site. Our model of Log P uses molar volume as applied by Abraham.[27] Theoretically molar volume can be related to the entropy of the solvent that results from lost available volume for dispersion due to the volume occupied by the solute.

A linear PLS model of Log P was developed using parameters that the MLR model of Log P found to be statistically significant. Columns 3−6 in Table 1 show the values for the parameters molar volume, Δ*C*, NHA, and NHD. Figure 1 is a plot of the experimental Log P versus predicted Log P from the PLS model for the 162 compounds. The PLS model of Log P for 162 compounds resulted in an r^2 value of 0.88. The 162 compounds were then used to generate 10 training sets and 10 external cross-validated sets.

The 10 training set PLS models of Log P resulted in an average r^2 of 0.87 ± 0.01. The 10 cross-validated test sets had an average $q_{10\%}{}^2$ of 0.87 ± 0.05. These results are similar to Chuman et al.[12] who reported an r^2 of 0.94 for 208 compounds using ab initio MO calculations, accessible surface area, a binary hydrogen bond acceptor indicator, and a binary solvent indicator parameter in linear regression equations.

The prediction of Log P from the sum of calculated ^{13}C NMR chemical shift differences allows the enthalpy contribution from all the carbon atoms in a compound to be accounted for. With fragment-based methods, some atoms are potentially accounted for more than once in the determination of Log P. This allows a fragment-based model to depend on the same structural factors more than once, which cannot be interpreted physically. The whole-molecule approach to calculating Log P limits the number of factors used in the model. The results from this initial study indicate that the summed differences of calculated NMR chemical shifts have the potential to increase the accuracy of Log P predictions. While the chemical shift values were calculated using prediction programs in this study, it is possible to verify the chemical shifts experimentally. The prediction of Log P could also be improved if the chemical shifts could be calculated in octanol. At this time, models of Log P are limited to only carbon chemical shifts. The ability to improve proton prediction and predict nitrogen, phosphorus, fluorine, and oxygen NMR chemical shifts may improve the models for Log P.

The predicted chemical shifts of many compounds are based on water and methanol data, but for some compounds, the chemical shift predictions are based on all solvents when there was not sufficient data available about the same or similar structures in the ACD/Labs HNMR and CNMR prediction software. Increasing the accuracy of the NMR spectra in both water and methanol or octanol NMR solvents should increase the accuracy of the Log P models based on chemical shift deviations.

While the partition coefficient can be measured using a variety of spectroscopic[28] or chromatographic techniques,[12−16] this requires the synthesis of the parent compound. Log P can also be calculated using computer-based substructure or whole molecule methods. These methods have been improved throughout the years to include the many types of solute−solvent interactions that may occur in order to more accurately estimate Log P. The results from this initial study using a small sample set indicates that summed differences in calculated carbon NMR chemical shifts can offer an additional parameter to the structure-based computer programs for the estimation of Log P.

CONCLUSIONS

We have presented a novel approach to the calculation of Log P. Differences in carbon NMR chemical shifts between water and methanol were used along with molar volume and number of hydrogen bond acceptors and donors to produce a model of Log P. Both the PLS and MLR models of Log P had an r^2 of 0.88. The r^2 of the PLS model was approximately the same as the average r^2 obtained when from 10 training sets and 10 leave-10%-out cross-validation sets. While 162 compounds were used to build the models, the results show that the sum of changes in calculated NMR chemical shifts can be employed with previously used parameters in the determination of Log P values. The sum of differences in predicted carbon chemical shifts had a significant contribution to the model of Log P, whereas the sum of differences in predicted proton chemical shifts did not have a significant contribution to the model of Log P. This could be due to the fact that the prediction of carbon NMR chemical shifts tends to be more accurate than the prediction of proton chemical shifts calculated in water and methanol. Since protons have a larger interaction with the solvent than carbon atoms, the electrostatic effects of the solvent will have a greater influence on proton chemical shifts than carbon chemical shifts. We believe that the use of experimental NMR spectra or improvement in the prediction of NMR chemical shifts in individual solvents such as water and methanol or eventually octanol should improve the accuracy of the contribution of the sum of chemical shift differences to the model Log P, especially the sum of differences in the proton chemical shifts. The current ACD/Labs HNMR and CNMR database only had the experimental proton and carbon NMR spectra of 7 compounds in both water and methanol solvents, which is not a large enough data set to justify the development of a model of Log P or to justify the findings from such a model. Therefore, a model based on experimental spectra will have to be developed in-house. Further, the ability to predict nitrogen, phosphorus, fluorine, and oxygen chemical shifts in different solvents should allow for a more robust model especially when considering potential drug compounds. The sum of the NMR chemical shift differences between a molecule in water and methanol or octanol represents a simple, computer based, whole-molecule approach to model Log P that is fairly accurate. This computer simulated whole-molecule approach can be easily applied to an even larger and diverse set of compounds.

REFERENCES AND NOTES

(1) Hansch, C.; Fujita, T. Correlation of biochemical activity of phenoxyacetic acids with Hammett substituent constants and partition coefficients. *Nature* **1962**, *194*, 178−180.

(2) Hansch, C.; Leo, A. *Substituent constants for correlation analysis in chemistry and biology*; Wiley: New York, 1979.

(3) *Exploring QSAR; hydrophobic, electronic and steric constants*; Hansch, C., Leo, A., Hoekman D., Eds.; ACS Professional Reference Book, 1995.

(4) Leo, J. Hydrophobic parameter: measurement and calculation. *Methods Enzym.* **1991**, *202*, 544−591.

(5) Leo, J. Calculating log Poct from structures. *Chem. Rev.* **1993**, *93*, 1281−1306.

(6) Leo, J.; Hoekman, D. Calculating log P(oct) with no missing fragments; The problem of estimating new interaction parameters. *Perspect. Drug Discovery* **2000**, *18*, 19−38.

(7) Meylan, W. M.; Howard, P. H. Atom-fragment contribution method for estimating octanol−water partition coefficients. *J. Pharm. Sci.* **1995**, *84*, 83−92.

(8) Meylan W. M.; Howard, P. H. Estimating log P with atom-fragments and water solubility with log P. *Perspect. Drug Discovery Des.* **2000**, *19*, 67−84.

(9) Japertas, P.; Didzipetris, R.; Petrauskas, A. Fragmental methods in the design of new compounds. Application of the advanced algorithm builder. *Quant. Struct-Act. Relat.* **2002**, *21*, 23−37.

(10) Votano, J. R. *2nd LogP Symposium, Lipophilicity in drug disposition*; Lausanne, March 2000.

(11) Mannhold, R.; Petrauskas, A. Substructure versus whole molecule approaches for calculating log P. *QSAR Comb. Sci.* **2003**, *22*, 466−475.

(12) Chuman, H.; Mori, A.; Tanaka, H.; Yamagami, C.; Fujita, T. Analyses of the partition coefficient, log P, using Ab Initio MO parameter and accessible surface area of solute molecules. *J. Pharm. Sci.* **2004**, *93*, 2681−2697.
(13) Araki, Y. K.; Ohnishi, K.; Hanasato, K.; Inaba, H.; Aono, M.; Ohta, A. Measurement and prediction of hydrophobicity parameters for highly lipophilic compounds: application of the HPLC column-switching technique to measurement of log P of diarylpyrazines. *J. Pharm. Sci*. **1999**, *88*, 1299−1304.
(14) Yamagami, C.; Tachikawa, H. Hydrophobicity parameters determined by reversed-phase liquid chromatography. XVI: A new hydrogen-accepting parameter for monosubstituted thiophenes and furans for correlating retention factors and octanol−water partition coefficients. *Chem. Pharm. Bull*. **2003**, *51*, 1196−1200.
(15) Daniel-Mwambeti, K.; Torrado, S.; Cuesta-Bandera, C.; Ponce-Gordo, F.; Torrado, J. J. The effect of solubilization on the oral bioavailability of three benzimidazole carbamate drugs. *Int. J. Pharm*. **2004**, *272*, 29−36.
(16) Taylor, P. J.; Cruickshank, J. M. Distribution coefficients of atenolol and sotalol. *J. Pharm. Pharmacol*. **1985**, *37*, 143−144.
(17) Dross, K.; Sonntag, C.; Mannhold, R. Determination of the hydrophobicity parameter RMw by reversed-phase thin-layer chromatography. *J. Chromatogr. A* **1994**, *671*, 113−124.
(18) DeBolt S.; Kollman, P. 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.
(19) Emsley, J. W.; Feeney, J.; Sutcliffe, L. H. *High-Resolution Nuclear Magnetic Resonance*; Pergamon Press Ltd.: Oxford, 1965; Volume I, pp 1−287.
(20) Beger, R. D.; Wilkes, J. G. Developing 13C NMR quantitative spectrometric data-activity relationship (QSDAR) models of steroid binding to the corticosteroid binding globulin. *J. Comput.-Aided Mol. Des.* **2001**, *15*, 659−669.
(21) Beger, R. D.; Buzatu, D. A.; Wilkes, J. G. 13C NMR quantitative spectrometric data-activity relationship (QSDAR) models of steroids binding the aromatase enzyme. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1360−1366.
(22) Bremser, W. HOSE − a Novel substructure Code. *Anal. Chim. Acta* **1978**, *103*, 355−365.
(23) Cramer, R. D.; Bunce, J. D.; Patterson, D. E. Cross-validation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies. *Quant. Struct.-Act. Relat.* **1988**, *7*, 18−25.
(24) Wallen, S. L.; Palmer, B. J.; Garrett, B. C.; Yonker, C. R. Density and temperature effects on the hydrogen bond structure of liquid methanol. *J. Phys. Chem.* **1996**, *100*, 3959−3964.
(25) Laszlo, P. Solvent effects and NMR. *Prog. Nucl. Magn. Reson. Spectrosc.* **1967**, *3*, 231−402.
(26) Fukui, H. Theory and calculation of nuclear shielding constants. *Prog. Nucl. Magn. Reson. Spectrosc.* **1997**, *31*, 317−342.
(27) Abraham, M. H.; Le, J. The correlation and prediction of the solubility of compounds in water using an amended solvation energy relationship. *J. Pharm. Sci.* **1999**, *88*, 868−880.
(28) Santos, N. C.; Prieto, M.; Castanho, M. A. R. B. Quantifying molecular partition into model systems of biomembranes: An emphasis on optical spectroscopic methods. *Biochim. Biophys. Acta* **2003**, *1612*, 123−135.

CI049643E