Prediction of Partitioning Properties for Benzaldehydes by Various Molecular Descriptors

L.-J. Jin, Z. Wei, J.-Y. Dai, P. Guo, L.-S. Wang

State Key Laboratory of Pollution Control and Resource Reuse, Department of
Environmental Science and Engineering, Nanjing University, Nanjing 210093,
Peoples' Republic of China

It is widely recognized that knowledge of the environmental fate of pollutants is a basic need for exposure assessment. The partitioning properties of the chemicals, such as water solubility and solvent (usually octanol)-water partition coefficients, would play an important role in their environmental transport and fate (Yalkowsky and Valvani, 1979; Abraham and Weathersby, 1994). Quantitative knowledge of these properties enables us to model the concentrations of a certain pollutant in different compartments of the environment (Lyman et al., 1982).

Substituted benzaldehyde compounds, used extensively as intermediates to synthesize pesticides (e.g. Monsanto's herbicide Lasso and other agricultural chemicals) and medicines (e.g. chloramphenicol, ephedrin, ampicillin, diphenylhydantoin and other products), are being introduced into the environment. However, few studies on their partitioning properties have been reported. Here, the water solubility and octaol/water partition coefficients of 15 substituted benzaldehydes with polar substituents are first reported. A series of equations are established to predict these properties.

MATERIALS AND METHODS

The tested chemicals were synthesized by College of Chemistry and Chemical Engineering of Nanjing University. The purity of each chemical was monitored by high pressure liquid chromatography (HPLC) to assure that no interference peak has occurred. The chemicals are listed in Table 1. All organic solvents and inorganic reagents used are analytical reagent grade.

The octanol/water partition coefficients and solubility were determined by shake flask method based on the OECD guidelines(l981) for testing of chemicals. All operations were conducted at 25 ± 0.5 °C The equilibrium concentrations for samples in the water phase were determined by UV/Visible spectrophotometry; in organic phase, the concentrations were calculated by the difference.

Regression analyses were performed using STATGRAPHICS program with a confidence limit of 95%. Model adequacy was measured as the square of

---

Correspondence to: L. Jin

correlation coefficient (R^2), the standard error of estimates (SE), the F value for analysis of variance (F) and the significance level (p). The reported R^2 values were adjusted for degree of freedom (d.f.).

Table 1. The compounds and the logarithm of observed K_{ow} and S_w (mol/L)

No.	Compounds	$\log K_{ow}$	$\log S_w$
1	2,5-dihydroxybenzaldehyde	0.54	-1.00
2	3,4-dihydroxybenzaldehyde	1.09	-1.34
3	4-hydroxy-3-methoxybenzaldehyde	1.17	-1.30
4	3,4-dimethoxybenzaldehyde	1.20	-1.42
5	3-hydroxybenzaldehyde	1.29	-1.23
6	4-hydroxybenzaldehyde	1.36	-1.16
7	3,4,5-trimethoxybenzaldehyde	1.39	-2.12
8	p-hydroxyacetphenone	1.44	-1.27
9	3-ethoxy-4-hydroxybenzaldehyde	1.61	-1.77
10	4-ethoxy-3-methoxybenzaldehyde	1.63	-2.19
11	3-bromo-4-hydroxybenzaldehyde	1.83	-2.18
12	isopropyl,4-formylphenoxyl acetate	1.88	-2.58
13	2,5-dimethoxybenzaldehyde	1.91	-2.32
14	5-bromo-3-hydroxybenzaldehyde	2.50	-3.25
15	2,5-dibromobenzaldehyde	3.35	-3.80

coefficient (R^2), the standard error of estimates (SE), the F value for analysis of variance (F) and the significance level (p). The reported R^2 values were adjusted for degree of freedom (d.f.).

RESULTS AND DISCUSSION

Collander (1950) has suggested that partition coefficients from different solvent pairs are linearly correlated. A series of equations between partition coefficients in octanol-water system (K_{ow}) and aqueous solubility(S_w) have been reported (Yalkowsky and Valvani, 1979; Mackay et al, 1980; Isnard and Lambert, 1989). Regression between $\log S_w$ and $\log K_{ow}$ (Table 1) of tested chemicals yields the following equation:

$$\log S_w = -0.0468 - 1.167 \ \log K_{ow} \tag{1}$$
$$R^2 = 0.8668 \ \ SE = 0.30 \ \ n = 15 \ \ F = 92.1 \ \ p = 0.0000$$

As shown, there is a good correlation between these two properties. It is similar to the theoretical equations obtained by other researchers (Hansch et al, 1968; Hansch and Leo, 1978, 1991; Brigs 1981; Miller et al, 1985). Based on this equation, $\log S_w$ of an unknown substituted bezaldehyde having similar structuremight be predicted from its $\log K_{ow}$.

Meylan et al(1996) showed that addition of molecular weight (MW) could increase the estimation accuracy of $logS_w$ from $logK_{ow}$. In our study, when MW (Table 2) of substituted benzaldehydes is introduced, equation (1) becomes the following one:

$$logS_w = 0.812 - 0.706\ logK_{ow} - 0.00934\ MW \qquad (2)$$
$$R^2 = 0.9385\ \ SE = 0.20\ \ n = 15\ \ F = 107.9\ \ p = 0.0000$$

As seen from equation (2) correlation coefficient improves from 0.931 to 0.969, SE decreases from 0.30 to 0.20, and F increases from 92.1 to 107.9. This further verifies that molecular weight can increase the estimation accuracy of $logS_w$.

Table 2. Descriptors* for 15 substituted benzaldehydes

No.	MW	β	$^2X^v$	^2X	^0Y
1	138.12	1.02	1.861	4.064	0.352
2	138.12	1.02	1.866	4.042	0.352
3	152.15	1.01	2.049	4.233	0.314
4	166.18	1.00	2.232	4.424	0.275
5	122.12	0.79	1.713	3.546	0.222
6	122.12	0.79	1.709	3.534	0.222
7	196.20	1.22	2.577	5.154	0.367
8	136.15	0.49	2.103	4.264	0.222
9	166.18	1.01	2.278	4.613	0.314
10	180.20	1.00	2.461	4.804	0.276
11	201.02	0.71	2.682	4.042	0.222
12	222.23	1.23	3.514	6.607	0.367
13	166.18	1.00	2.226	4.424	0.275
14	201.02	0.71	2.735	4.064	0.222
15	263.92	0.47	3.551	4.064	0.092

* MW represented molecular weight;

β values were estimated by the method of Hickey and Passlno-Reader(1991);

^2X, $^2X^v$ were calculated according to the methods described by Kier and Hall (1976, 1981);

^0Y was a topological indice calculated from molecular connectivity indices (Bahnick and Doucette, 1988).

Beta (β) represents solute hydrogen-bond acceptor basicity, i.e., the ability to accept a proton in a solute-solvent hydrogen bond. β (Table 2) values of 15 substituted benzaldehydes were estimated by the method of Hickey and Passlno-Reader(1991). When β was introduced into the correlation equation of $logS_w$ vs. $logK_{ow}$, it can also improve the estimation accuracy of $logS_w$, just like the molecular weight. The yielded equation is the following:

$$\log S_w = 1.180 - 1.348 \log K_{ow} - 1.040 \ \beta \qquad (3)$$
$$R^2 = 0.9308 \quad SE = 0.22 \quad n = 15 \quad F = 95.2 \quad p = 0.0000$$

Though the relationship between $\log S_w$ and $\log K_{ow}$ could provide easy estimation of one from the other, the accuracy for estimation of one property would depend on the accuracy for determination of the other one. Moreover, this relationship can not provide a quantitative description for the governing factors on different partitioning properties of the compounds.

Molecular connectivity is a method of describing molecular structure based solely on bonding and branching patterns rather than physical or chemical characteristics. It has long been shown to be a useful structural parameter for describing diverse physicochemical properties and biological activity, such as water solubility (Nirmalakhandan and Speece, 1988) and octanol-water partition coefficient. In our work, the indices were introduced into the correlations by stepwise regression analysis to optimize the correlation satisfactorily. Molecular connectivity indices were calculated according to the methods described by Kier and Hall (1976, 1981). Only the parameters included in the regression equations are listed in Table 2. The equations are:

$$\log S_w = -0.034 - 1.582 \ ^2X^v + 0.422 \ ^2X \qquad (4)$$
$$R^2 = 0.8274 \quad SE = 0.34 \quad n = 15 \quad F = 34.6 \quad p = 0.0000$$

$$\log K_{ow} = 1.099 + 0.753 \ ^2X^v - 4.657 \ ^0Y \qquad (5)$$
$$R^2 = 0.8524 \quad SE = 0.25 \quad n = 15 \quad F = 41.4 \quad p = 0.0000$$

^0Y is a topological indice calculated from molecular connectivity indices (Bahnick and Doucette, 1988), which reflects the ability of the non-disperse force of the molecular. It is seen from the above correlation equations that the leading terms are $^2X^v$. Increasing $^2X^v$ leads to decreasing S_w and increasing K_{ow}. Additionally, increasing ^2X leads to increasing S_w, while increasing ^0Y leads to decreasing K_{ow}. The reason is that increasing ^0Y means increasing polarity of solute, and favors partition into water.

Additionally, we find that the non-disperse force factor improves greatly the correlation of $\log S_w$ vs. $\log K_{ow}$ (equation (1)):

$$\log S_w = 1.543 - 1.478 \log K_{ow} - 3.984 \ _0Y \qquad (6)$$
$$R^2 = 0.9391 \quad SE = 0.20 \quad n = 15 \quad F = 109 \quad p = 0.0000$$

As shown in the equation above, non-disperse force is better than β and molecular weight to improve the estimation accuracy of $\log S_w$, because R^2 and F values are a little higher. The tested chemicals all have polar substituents, which more favor partition into the polar water, so non-disperse force factor can greatly improve the estimation accuracy of $\log S_w$ from $\log K_{ow}$ (equation (1)).
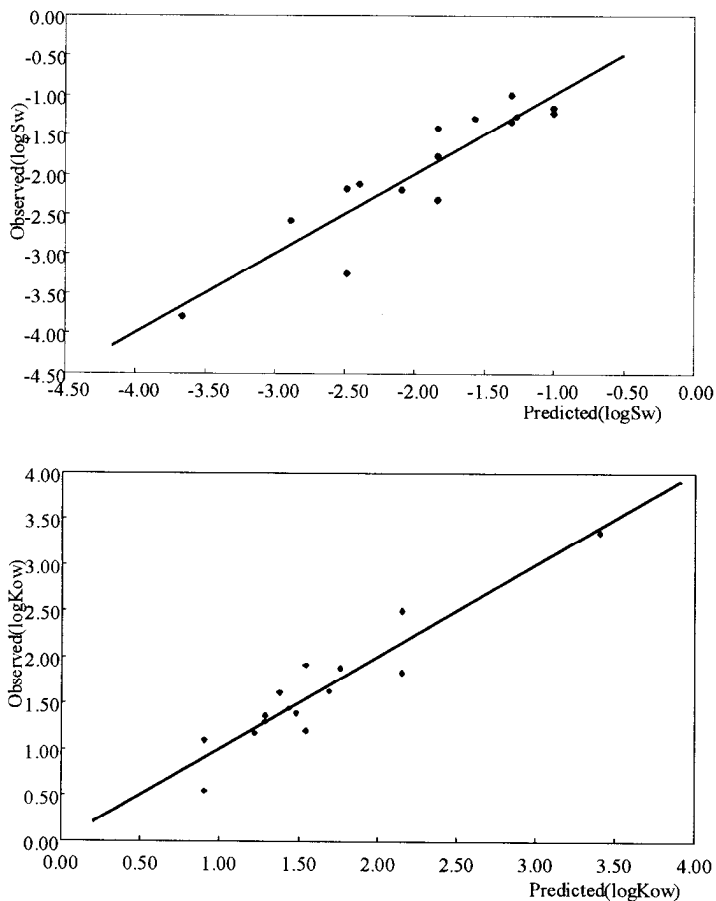
Figure 1. Plot of observed properties vs. predicted properties (calculated with equation (7) and (8) respectively)

Stepwise regression of $\log S_w$ and $\log K_{ow}$ with molecular weight(MW), molecular connectivity indices, LSER (linear solvation energy relationship) variables calculated by Hickey(1991) and the descriptors derived from quantum computation (Chen et al, 1996) yields the following equations:

$$\log S_w = 1.289 - 0.0188 \; MM \tag{7}$$
$$R^2 = 0.8250 \quad SE = 0.34 \quad n = 15 \quad F = 67.0 \quad p = 0.0000$$

$$\log K_{ow} = 0.890 + 0.011 \; MW - 4.282 \; {}^0Y \tag{8}$$
$$R^2 = 0.8583 \quad SE = 0.25 \quad n = 15 \quad F = 43.4 \quad p = 0.0000$$

From the above correlation equations, we find that the most effective coefficient of $\log S_w$ and $\log K_{ow}$ is the molecular weight(MW), which can explain 83 percent of

the variance in equation (7). Increasing MW leads to increasing K_{ow} and decreasing S_w. The reason is that high molecular weight means large volume of solute molecule, and large molecule favors partition into the less cohesive solvent, such as octanol in equation (8).

Additionally, increasing ^0Y(non-disperse force factor) leads to decreasing K_{ow}. This is in agreement with the fact that high polarity of tested benzaldehydes favors partition into water, which also has large dipolarity/polarizability. These established correlation equations could be used to predict S_w and K_{ow} of some other similar substituted benzaldehydes. Figure 1 shows the correlationship between observed and estimated (calculated with equation (7) and (8) respectively) values of $logS_w$ and $logK_{ow}$ respectively.

In summary, molecular weight, β and non-disperse force factor can increase the estimation accuracy of S_w from K_{ow} of substituted benzaldehydes. The most effective coefficient of $logS_w$ and $logK_{ow}$ is the molecular weight. The next effective coefficient of $logK_{ow}$ is the non-disperse force factor.

REFERENCES

Abraham MH, Weathersby PK (1994) Hydrogen bonding. 30. Solubility of gases and vapors in biological liquids and tissues. J Pharm Sci 83: 1450-1456.

Bahnick DA Doucette WJ (1988) Use of molecular connectivity indices to estimate soil sorption coefficients for organic chemicals. Chemosphere 17: 1703-1715.

Brigs GG (1981) Theoretical and experimental relationships between soil adsorption, octanol-water partition coefficients, water solubility, bioconcentration factors, and the parachor. J Agric Food Chem. 29: 1050-1059.

Chen J, Feng L, Liao Y, Han S, Wang L and Hu H (1996) Using AM1 Hamiltonian in quantitative structure-properties relationship studies of alkyl (l-phenylsulfonyl)cycloalkane-carboxylates. Chemosphere 33:537-546.

Collander R (1950) The distribution of organic compounds between iso-butanol and water. Acta Chem Scand 4:1085-1098.

Hansch C, Leo A (1979) Substituent constants for correlation analysis in chemistry and biology. John Wiley & Sons. New York.

Hansch C, Quinlan JE and Lawrence GL (1968) The linear free-energy relationship between partition coefficients and aqueous solubility of organic liquids. J Org Chem 33:347-350.

Hickey JP, Passlno-Reader DR (1991) Linear solvation energy relationships: "rules of thumb" for estimation of variable values. Environ Sci Technol 25:1753-1760.

Isnard P and Lambert S (1989) Aqueous solubility and n-octanol-water partition coefficient correlation. Chemosphere 18: 1837-1853.

Kier LB, Hall LH (1976) Molecular connectivity in chemistry and drug design. Academic Press. New York.

Kier LB, Hall LH (1981) Derivation and significance of valence molecular connectivity. J. Pharm. Sci. 70:583-589.

Lyman WJ, Reehl WF, Rosenblatt DH (1982) Handbook of chemical property estimation methods. McGraw-Hill Book Company. New York.

Mackay D, Bobra A and Shiu WY (1980) Relationship between aqueous solubility and octanol/water partition coefficients. Chemosphere 9:701-711.

Meylan WM, Howard PH, Boethling RS (1996) Improved method for estimating water solubility from octanol/water partition coefficient. Environ Toxicol Chem 15:100-106.

Miller MM, Wasik SP, Huang GL, Shiu WY, Mackay D (1985) Relationships between octanol-water partition coefficient and aqueous solubility. Environ Sci Technol 19:522-529.

Nirmalakhandan NN, Speece RE (1988) Prediction of aqueous solubility of organic chemicals based on molecular structure. Environ Sci Technol 22:328-338.

OECD (The Organization for Economic Cooperation and Development) Guideline for testing of chemicals. Paris. 105. 1981.

Yalkowsky SH, Valvani SC (1979) Solubilities and partitioning 2. Relationships between aqueous solubilities, partition coefficients and molecular surface areas of rigid aromatic hydrocarbons. J Chem Eng Data 24: 127-129.