

Prediction of aqueous solubility of PCBs based on molecular structure

Huang Guolan, Song Rengao and Dai Shugui

Department of Environmental Science, Nankai University, Tianjin 300071, China

Abstract — The thermodynamic relationships among aqueous solubility and molar volume (MV), total molecular surface area (TSA) and molecular connectivity index (MCI) for highly hydrophobic chemicals, PCBs are established and discussed, respectively. Good linear relationships exist among $\ln C_s$ and MV , TSA or MCI .

Keywords: aqueous solubility; molar volume; total molecular surface area; molecular connectivity index.

INTRODUCTION

One of the most important data items in assessing the environmental fate of a chemical is its aqueous solubility. However, our current knowledge of this parameter of a chemical is inadequate, and it is impossible to test experimentally for the hydrophobic property of all chemicals on the market. There is a growing interest in the development of predictive techniques to estimate solubility data. Availability of good predictive methods can also aid in devising optimal experimental conditions for laboratory measurement, in screening new chemicals for further scrutiny, and in estimating other related properties. The quantitative structure-activity relationship (QSAR) approach has been found to be a useful one in developing such predictive models and has become increasingly accurate and popular.

In this paper, independent variables such as molar volume (MV), total molecular surface area (TSA), and molecular connectivity index (MCI) which are all calculated purely without experimental inputs are used as descriptors of solute molecular structures to correlate and assess the solubility data of polychlorinated biphenyls (PCBs).

EXPERIMENTAL SECTION

Chemicals

The polychlorinated biphenyls were obtained from Aldrich and Foxboro/Analab, and some were kindly supplied by Dr. S. Safe. The chemicals received were of the

best available commercial quality and were used for analysis without further purification. Milli-Q water and retreated Milli-Q water using a Norganic system were used for all experiments.

Instruments

The instruments were Waters Associates Model ALC/GPC 244 liquid chromatograph consisting of a Model 6000 solvent delivery system, a Model 660 solvent flow programmer, a Model M45 solvent delivery system, a Model 440 UV absorbance detector with 254- and 280 nm kits and a Model 420-E fluorescence detector with excitation filter 340 nm. The analytical column was a 3.9 mm OD \times 300mm long μ Bondapak C₁₈ column. Integration of the peak area was done by a Hewlett-Packard Model 3380A integrator recorder.

Determination of aqueous solubility

The apparatus and experimental procedures used were the dynamic-coupled-column liquid chromatography method previous described by May *et al.* (May, 1978). Saturated aqueous solutions were prepared by passage of water through a generator column which was packed with glass beads coated with the interested PCB. The saturated solution was then pumped through an extractor column packed with 37–50 μ Bondapak/Corasil (Waters Associates). By switching of the eight-port Valco valve the absorbed chemical was extracted by the mobile phase and directly injected onto the analytical column.

THEORETICAL SECTION

Aqueous solubility

The solubility of a pure, non-ionizing, sparingly soluble liquid solute can be expressed by

$$f = P^s = x r_w P^s, \text{ thus } x = 1/r_w, \quad (1)$$

in which x is mole fraction, r_w is activity coefficient, P^s is vapour pressure, and f is the fugacity. For a solid solute, the equation takes the form

$$f = f_s = x \cdot r_w \cdot f_R, \quad (2)$$

where f_s is the fugacity of the pure solid, x is the aqueous solubility, r_w is the activity coefficient, and f_R is defined as the reference fugacity or the fugacity of pure subcooled liquid at the temperature of the solution and at its own saturation pressure. From Equation 2 the solubility is

$$x = f_s / r_w \cdot f_R. \quad (3)$$

The ratio (f_S/f_R) is termed the fugacity ratio, F . It is equal to unity at the triple point but it is less than unity below that point. Its importance lies in the fact that it causes a marked reduction in solubility of solid organic compounds compared to similar liquid compounds. To take the natural logarithms of both sides of Equation 3,

$$\ln x = \ln (f_S/f_R) - \ln r_w . \quad (4)$$

Prausnitz expressed the fugacity ratio F as (Huang, 1983)

$$\ln F = -\Delta H_f (1/T - 1/T_M) / R , \quad (5)$$

where ΔH_f is the enthalpy of fusion (J/mol), T and T_M are the system and solute melting point (K), respectively, and R is the gas constant (8.3 J/mol K). Since the entropy of fusion ΔS_f at the melting point is $\Delta H_f/T_M$ and is found to be fairly constant, Equation 5 can be rewritten as

$$\ln F = -\Delta S_f (T_M - T) / RT . \quad (6)$$

Yalkowsky (Yalkowsky, 1980) has shown that for many organic compounds ΔS_f is approximately 56.5 J/mol K, thus $\Delta S_f/R$ is approximately 6.8. Thus,

$$\ln F = -6.8 (T_M - T) / T \quad (7)$$

Then, Equation 4 becomes

$$\ln x = -6.8 (T_M - T) / T + \ln r_w \quad (8)$$

Therefore, the aqueous solubility of a slightly soluble solute depends on its activity coefficient and its triple point temperature which is approximately equal to its melting point, and the solubility can be estimated if data for r_w and T_M are available.

Converting to mol/m³ units using the water molar volume V_w (18×10^{-6} m³/mol) gives $C_s = X/V_w$ and $C_s = \frac{f_S/f_R \cdot x}{V_w}$, similarly converting to mg/L or g/m³ units using molecular weight M gives $C_s = MX/V_w$, where C_s is the aqueous solubility of a solute, r_w is the activity coefficient of the solute in water.

Correlation

In recent years, a number of workers have established correlations between the aqueous solubility and other physicochemical parameters based on molecular structure. D. Mackay and W. Y. Shiu or Leinonen *et al.* (Mackay, 1977) have used carbon number or molar volume to correlate r_w .

$$\ln r_w = a (MV) + b . \quad (9)$$

Correlations by Yalkowsky and Valvani (Yalkowsky, 1979) have shown that total molecular surface area is an excellent parameter for correlating r_w .

$$\ln r_w = a (TSA) + b . \quad (10)$$

Because molecular connectivity index is a parameter for the quantitation of molecular structure that encodes information about size, branching, cyclization, and so on, and good correlation exists between TSA and MCI , it should be found a good correlation between MCI and r_w .

$$\ln r_w = a (MCI) + b . \quad (11)$$

For Equations 9, 10 and 11, a and b are constants. Since the solubility of liquid x is inversely proportional to r_w , a similar correlation exists for $\ln x$, except that the b changes and the sign of a changes. For solid PCBs, the overall correlation is

$$\ln x = -6.8 (T_M - T) / T - a (MV) - b , \quad (12)$$

$$\ln x = -6.8 (T_M - T) / T - a (TSA) - b , \quad (13)$$

$$\ln x = -6.8 (T_M - T) / T - a (MCI) - b , \quad (14)$$

TSA , MCI , and MV calculation.

The procedure for the calculation of TSA is the same as method C of Valvani *et al.* (Valvani, 1973). The following values for interatomic distances in a molecule were used: aromatic C-C, 1.40 Å; aromatic C-H, 1.08 Å; aromatic C-Cl, 1.70 Å. The Van der Waal's radii used were aromatic C, 1.70 Å; aromatic H, 1.20 Å; aromatic Cl, 1.80 Å.

For the compounds studied, aromatic C-C-C angles and H-C-C angles used were 120°. The TSA is the sum of individual atoms or group surface area contributions. Only the simple first valence MCI s used here will be presented. A more detailed discussion is provided by Kier and Hall (Kier, 1976).

$${}^1X = \sum (\delta_i \cdot \delta_j)^{-1/2} . \quad (15)$$

$${}^1X' = \sum (\delta_i' \cdot \delta_j')^{-1/2} . \quad (16)$$

It is evident that solubility difference of toluene and chlorobenzene is caused by the only difference between methyl group and chlorine atom. Therefore, the difference between them can be approximately eliminated or at least reduced by subtracting the

value of a particular index for the interested molecule such as chlorobenzene (CB) from the same index for a molecule formed by replacing the heteroatom by carbon.

$$\Delta X = X_{\text{CB}}^v - X_{\text{TOLUENE}}^v$$

ΔX is called polar correction factor. Molar volume of a molecule is the ratio of molecular weight and its density and it can be calculated by Le Bas method (Reid, 1977).

RESULT AND DISCUSSION

Table 1 is a compilation of molar volume, melting point, total molecular surface area, molecular connectivity index, and solubility data for 17 PCBs.

Table 1 Physico-chemical properties of PCBs at 25 °C

Compounds	mp, K	MV, cm ³ /mol	TSA, Å ²	'X	'X ^v	ΔX	Solubility, mol/m ³
biphenyl	342.6	178.3	192.5	5.966	4.071	0	0.0435
2-	305.3	205.5	208.4	6.377	4.586	0.097	0.0268
2, 5-	295.0	226.4	227.6	6.771	5.094	0.196	0.0087
2, 6-	307.9	226.4	224.4	6.788	5.100	0.196	0.00623
2, 4, 5-	349.5	247.3	241.6	7.181	5.608	0.293	6.32 × 10 ⁻⁴
2, 4, 6-	334.3	247.3	239.8	7.181	5.608	0.293	8.76 × 10 ⁻⁴
2, 3, 4, 5-	363.9	268.2	255.2	7.609	6.212	0.390	7.17 × 10 ⁻⁵
2, 2', 4, 5-	339.1	268.2	257.4	7.592	6.122	0.390	5.63 × 10 ⁻⁵
2, 3, 4, 5, 6-	347.6	287.1	269.2	8.037	6.732	0.488	1.68 × 10 ⁻⁵
2, 2', 4, 5, 5'-	350.1	287.1	275.2	7.985	6.631	0.488	5.92 × 10 ⁻⁵
2, 2', 3, 3', 6, 6'-	385.2	310	285.2	8.430	7.157	0.586	1.67 × 10 ⁻⁵
2, 2', 3, 3', 4, 4'-	424.9	310	286.8	8.430	7.157	0.586	7.84 × 10 ⁻⁷
2, 2', 4, 4', 6, 6'-	386.7	310	291.5	8.396	7.145	0.586	1.13 × 10 ⁻⁶
2, 2', 3, 3', 4, 4', 6-	395.4	330.9	294.8	8.841	7.671	0.683	5.49 × 10 ⁻⁶
2, 2', 3, 3', 5, 5', 6, 6'-	433.8	351.8	318.7	9.252	8.186	0.781	9.15 × 10 ⁻⁷
2, 2', 3, 3', 4, 5, 5', 6, 6'-	455.8	372.7	337.0	9.680	8.706	0.878	3.88 × 10 ⁻⁸
deca-	578.9	393.6	345.9	10.11	9.226	0.976	1.49 × 10 ⁻⁸

Correlations between the aqueous solubility and descriptor of molecular structures are shown in Table 2 .

As a result, a good linear relationship exists between $\ln C_s$ and every descriptor of molecular structure ($R > 0.9750$). PCBs with higher values of *MV*, *TSA* or *MCI* show weaker solubility in water, on the other hand, stronger partition into biota, sediment and soil

Table 2 Correlations between $\ln C_s$ and TSA, MV, and MCI values

Independent variables	C1	C2	B	R	S	C3
-0.0228 T_M , MV	0.3270	-0.0663	11.56	0.9753	1.058	
-0.0228 T_M , TSA	0.3305	-0.0910	17.14	0.9794	0.9686	
-0.0228 T_M , 'X	0.2397	-3.493	19.67	0.9751	1.062	
-0.0228 T_M , 'X'	0.2708	-2.780	10.42	0.9765	1.033	
-0.0228 T_M , 'X', ΔX	0.4439	15.02	-89.17	0.9970	1.062	-77.40
-0.0228 T_M , 'X', ΔX	0.2853	-7.071	28.00	0.9770	1.062	22.73

C1 and C2: coefficient of the first two independent variables, and C3 is the coefficient of ΔX ;
 B: intercept; R: correlation coefficient; S: standard deviation; C_s : in unit of mol/m³

with organic carbon. Therefore, pollutants with higher values of above parameters are much lipophilic and of environmental significance. In terms of the above models, one can calculate relatively accurate aqueous solubilities of PCBs, without measuring them. When the molecular structure of PCB is known, the MV, TSA or MCI are simply and rapidly calculated. Finally, the aqueous solubility is obtained after those values are introduced into the established models, respectively.

During the model establishment, it was found that the correlation coefficient R was about 2% higher when C_s in unit of mol/m³ compared to that when C_s in unit of g/m³, mg/L or ppm. The difference of correlation coefficients was caused by the presence of a magnifying factor M, molecular weight, when mol/m³ is converted to ppm.

$$(\text{Total error})^2 = (\text{Method error})^2 + (\text{Propagated error})^2$$

Propagated error in the second term was M times of the former term. For this reason, it is suggested that C_s in the unit of mol/m³ is used to establish models if the observed C_s is in unit of mol/m³. Of course, one can convert mol/m³ to ppm in the light of actual conditions. However, this can be done only after the C_s (mol/m³) is obtained according to the related equations.

For C_s -MCI models, it was also found that the R values were a little higher after the polar correction factor was introduced into the models. However, in this case, F values decreased and S values increased. That is the correlations between $\ln C_s$ and MCI was not improved. This is probably due to the much weak polarity of PCBs. Therefore, it is not encouraged to use those models for PCBs.

The effect of temperature on solubility is as follows, assuming that the enthalpy of solution is constant.

$$\ln (C_{s1} / C_{s2}) = H/R (1/T_2 - 1/T_1) ,$$

in which the range between T_1 and T_2 does not include the melting point and is not very large.

CONCLUSIONS

MV , TSA and MCI are all good descriptors of PCB molecular structures for predicting aqueous solubility data.

This method may be applicable to other series of sparingly soluble organic contaminants.

It should be clear that the use of MV , TSA and MCI models involve some acceptance of a few untested assumptions or hypothesis with independent and dependent variables.

The accuracy of measured C_s data are often poor in comparison with the computed MV , TSA and MCI values, so the propagated error is mainly from the aqueous solubility error input.

REFERENCES

- Huang Guolan, Dissertation, University of Toronto, 1983
Kier, L. B., Hall, L.H., Molecular connectivity in chemistry and drug research, New York: Academic Press, 1976
May, W. E., Wasik, S. P., Freeman, D.H., Anal. Chem. , 1978, 50:997
Mackay, D., J. of Chem. and Eng. Data, 1977, 22:399
Reid, R. C., Prausnitz, J. M., The properties of gases and liquids, 3rd ed., New York: McGraw-Hill, 1977
Valvani, S.C., Yalkowsky, S.H., J. Phys. Chem., 1973, 80:829
Yalkowsky, S.H. and Valvani, S. C., J. Pharm. Sci., 1980, 69:912
Yalkowsky, S.H. and Valvani, S. C., J. Chem. Eng. Data, 1979, 24:127

(Received May 22, 1991)