Prediction of Soil Sorption Coefficients Using a Universal Solvation Model

ARTICLE *in* ENVIRONMENTAL SCIENCE AND TECHNOLOGY · NOVEMBER 2000

Impact Factor: 5.33 · DOI: 10.1021/es0009065

3 AUTHORS:

Paul Winget
Georgia Institute of Technology
41 PUBLICATIONS **1,444** CITATIONS
SEE PROFILE

Christopher J Cramer
University of Minnesota Twin Cities
534 PUBLICATIONS **23,570** CITATIONS
SEE PROFILE

Donald Truhlar
University of Minnesota Twin Cities
1,342 PUBLICATIONS **82,543** CITATIONS
SEE PROFILE

Environ. Sci. Technol. 2000, 34, 4733–4740

Prediction of Soil Sorption Coefficients Using a Universal Solvation Model

PAUL WINGET, CHRISTOPHER J. CRAMER, AND DONALD G. TRUHLAR*

Department of Chemistry and Supercomputer Institute, 207 Pleasant Street SE, Minneapolis, Minnesota 55455-0431

Using a database of 440 molecules, we develop a set of effective solvent descriptors that characterize the organic carbon component of soil and thereby allow quantum mechanical SM5 universal solvation models to be applied to partitioning of solutes between soil and air. Combining this set of effective solvent descriptors with solute atomic surface tension parameters already developed for water/air and organic solvent/air partitioning allows one to predict the partitioning of any solutes composed of H, C, N, O, F, P, S, Cl, Br, and I between soil and water. We also present linear correlations of soil/water partitioning with 1-octanol/water partition coefficients using the same database. The quantum mechanical calculations have the advantages that they require no experimental input and should be robust for a wide range of solute functionality. The quantitative effective solvent descriptors can be used for a better understanding (than with previously available models) of the sources of different partitioning phenomena in cases where the results exhibit significant fragment interactions. We anticipate that the model will be useful for understanding the partitioning of organic chemicals in the environment between water and soil or, more generally, between water and soil or sediments (geosorbents).

1. Introduction

The transport and fate of organic molecules in the environment depends in part on the degree to which they partition between different environmental phases (*1*–*4*). One particularly important process is sorption from water into soil. Because this often depends primarily on the soil's organic carbon content, measured values are usually normalized for the organic carbon (OC) content of soil, in which case the soil sorption equilibrium constant is expressed as (*5*, *6*)

$$K_{OC} = \frac{C_{soil}/C^{o}_{soil}}{C_{w}/C^{o}_{w}} \qquad (1)$$

where C_{soil} is the concentration of solute per gram of carbon in a standard soil and C_w is the concentration of solute per volume of aqueous solution. The standard state concentrations C^{o}_{soil} and C^{o}_{w} are typically chosen as 1 μg of solute/g of organic carbon for soil and 1 μg of solute/mL for aqueous solution. Although we call this the soil/water equilibrium constant for simplicity, one should note that it is actually a measure of partitioning between soil organic matter and water, normalized to organic carbon content.

The availability of measured values of K_{OC} for a large number of organic compounds has prompted many efforts to develop predictive models for this quantity. Such models are useful in providing estimates that may avoid the need for experimental measurements or assist in optimizing the methodology employed for those measurements as well as for environmental modeling. Two recent reviews are available (*6*, *7*).

Several models for estimating soil sorption coefficients take advantage of the correlation between K_{OC} and other experimental partition coefficients. For example, K_{OC} values have been estimated from experimental octanol/water partition coefficients ($P_{o/w}$) by

$$\log K_{OC} = m \log P_{o/w} + b \qquad (2)$$

where m and b are parameters developed from linear regression. Published values of m and b range from 0.5 to 1.1 and from −0.2 to 1.3, respectively, depending on the range of data employed in the individual regressions (*8*–*17*). The variation of these parameters means that when a model based on eq 2 is employed, care must be taken to ensure a reasonable similarity between the solute molecule of interest and the molecules used in the data set used for the particular regression upon which m and b are based, as has been noted previously (*16*). Predictive equations analogous to eq 2 have also appeared where $P_{o/w}$ is replaced by a chromatographic partition coefficient (*18*), by the aqueous solubility (*8*, *10*, *19*) (i.e., the partitioning of the solute between water and the pure solute phase), or by a bioconcentration factor (*10*) (i.e., the partitioning of the solute between water and biolipids).

Linear correlations between K_{OC} and other measured solute properties (and nonlinear correlations between K_{OC} and $P_{o/w}$ or other solute properties) have also been developed. The other solute properties that have been employed include parachor (*8*, *10*, *20*, *21*) (essentially molar volume times surface tension to the one-quarter power), the Hildebrand solubility parameter (*22*), and solvatochromic parameters (*23*).

All of the methods discussed thus far require that some experimental data be available in order to estimate K_{OC}. Several models have been developed that avoid this restriction. Some of these models require only topological descriptors, like molecular connectivity indices (*14*, *24*, *25*) or the so-called adsorbability index (*26*). Another such approach is based on fragment contributions; for example, Meylan et al. developed a model combining topological and fragment contributions (*27*), and Tao et al. (*28*) have described a model where the total partition constant is estimated as a sum over characteristic fragment contributions. In the spirit of correlating partitioning behavior with other physical properties, Reddy and Locke (*29*) developed a predictive model that is linear in certain molecular properties that are themselves *computed* from electronic structure calculations using semiempirical molecular orbital theory.

In this paper, we use two members, SM5.42R and SM5CR, of the SM5 family of quantum mechanical universal solvation models (*30*–*39*) to predict K_{OC}. First, we test how well we can predict K_{OC} by correlating it with values of $P_{o/w}$ that are computed using previously parametrized SM5CR (*39*) and SM5.42R (*37*) models for 1-octanol and water. Then, we develop new parameters for the SM5.42R model to directly compute soil/air partitioning free energies. A critical assumption in this treatment (the working hypothesis of the

* Corresponding author phone: (612)624-7555; fax: (612)626-9390; e-mail: truhlar@umn.edu.

10.1021/es0009065 CCC: $19.00
Published on Web 10/07/2000

model) is that soil may be modeled as a continuous medium with homogeneous, isotropic properties, whereas in reality it is a physically and chemically heterogeneous system. Continuum models have been very successful for predicting transfer free energies between fluids (*40*), but soil is most simply characterized as a colloid composed primarily of humic acids and minerals (*2*, *41*, *42*), and it has never previously been tested whether a continuum model is useful. The use of a continuum model in which the soil is modeled as homogeneous does not imply that soil is actually homogeneous; rather the effective solvent properties assigned to the soil represent an empirical average. Furthermore, the model does not imply a preference for the absorption model of soil/water partitioning over the adsorption model—both processes may be important. We agree with the statement that soil sorption results from "the balance of several possible interactions between the soil and solute molecule, the magnitude of which will be decided by the composition of the soil" (*43*). Our model includes the structural characteristics of the solute through its three-dimensional nuclear and electronic structure and the characteristics of soil through a set of effective solvent descriptors. The interaction of solute and soil is modeled by a sum of terms that include both microscopic interfacial area and bulk electrostatics in an effective dielectric medium. Our goal is to see how well such a model can predict K_{OC} data for a diverse set of nonionic solutes with a single set of parameters.

2. Methodology and Theory

2.1. Data Sets. Three data sets, which we will call K_{OC} data set 1, K_{OC} data set 2, and the $P_{o/w}$ data set, were employed in various calculations described below. The first K_{OC} data set is comprised of a 387-member subset of the 394-molecule merged data sets of Meylan et al. (*27*). The $P_{o/w}$ data set is a subset of K_{OC} data set 1 and is comprised of the 316 compounds for which at least one nonrejected (see below) experimental measurement of the 1-octanol/water partition coefficient is also available. All data are for 298 K.

Experimental values for K_{OC} data set 1 were taken from Meylan et al. (*27*), who collected data from various compilations and eliminated values from sources whose values were judged to be either consistently too high or too low.

Experimental values for $P_{o/w}$ were taken from the database of partition coefficients developed by Leo and Hansch (*44*). Only data measured within 1 log unit of neutral aqueous pH were considered. Values were rejected if they were measured outside the pH 6–8 range or outside the 20–30 °C temperature range, if they were marked unreliable by Leo and Hansch, if the aqueous phase was not pure aqueous, if the compound was ionized in one or both phases, or if ionization was supressed by using another compound. For compounds for which two acceptable data in the pH 6–8 range are presented in the database without explanation as to why one might be more reliable than another, the mean of the data was used. When three or more acceptable data were presented, values more than two standard deviations from the mean were discarded, and this process was repeated until all remaining experimental data were within two standard deviations of their mean.

The second K_{OC} data set was created by adding 53 molecules to set 1. These 53 molecules were selected as a representative cross section of additional molecules in the data set of Sablijc et al. (*16*) that are not in the data set of Meylan et al. (*27*). A list of all molecules in each data set and experimental and computed values for various free energies of transfer are provided as Supporting Information.

For convenience, all partition coefficients in this work employ standard state concentrations of 1 mol L^{-1}. This is a typical choice for $P_{o/w}$, but not, as noted above, for K_{OC}. The main advantage of using 1 mol L^{-1} as the standard state for all phases is that the concentration-dependent free-particle entropy of transfer vanishes when the number of particles per unit volume does not change during the transfer. This simplifies the theoretical treatment and allows the standard state free energy of transfer to be more directly related to solute–solvent interactions (*45*, *46*).

Conversion of conventional K_{OC} values to our preferred standard state requires multiplying them by the density of soil organic carbon expressed in units of grams per milliliter (g/mL). A representative value of this density measured for a variety of soil types from eastern North America is 0.11 (*47*), and this value is therefore taken as a standard value. Using it yields

$$P_{soil/w} = 0.11K_{OC} \tag{3}$$

where w denotes water. For a log K_{OC} value, the equivalent conversion requires subtraction of 0.96 log unit. (We use base-10 for logarithms throughout the paper.)

2.2. SM5 Universal Solvation Models. The partition coefficient $P_{A/B}$ for an individual solute transferring from phase B to phase A is related to the difference between the standard-state free energies of transfer of the solute into the two phases from the gas-phase according to

$$\log P_{A/B} = -\frac{\Delta G^\circ_A - \Delta G^\circ_B}{2.303RT} \tag{4}$$

where ΔG°_X is the standard-state free energy of transfer from the gas phase ("air") to the condensed phase X (a solvation free energy if into solution), R is the universal gas constant, and T is the temperature. The SM5 universal solvation models have been developed to compute such free energies of transfer (*30–39*). Full details of the SM5.42R (*34–38*) and SM5CR (*39*) models for calculating free energies of transfer from the gas phase to a condensed phase have been published previously. We recapitulate here only those details critical to understanding the parametrizations of the present paper

The standard-state free energies of solvation (i.e., free energies of transfer from air to solution) for the two SM5 models employed here are written as

$$\Delta G^\circ_X(\text{SM5CR or SM5.42R}) = \Delta G_{EP} + G_{CDS} \tag{5}$$

where ΔG_{EP} is the change in quantum mechanical solute electronic energy and solvent electric polarization free energy when the solute is immersed in continuous medium X described by dielectric constant ϵ (also called the relative permittivity), and G_{CDS} is a cavitation–dispersion–solvent–structure term depending on solvent descriptors.

The SM5.42R and SM5CR methods differ in the way that ΔG_{EP} is estimated. In both models, the solute geometric relaxation is not included explicitly, i.e., for computational purposes, the geometry of the solute in liquid solution is modeled using a gas-phase geometry. The R in the model name denotes that parametrization is for "rigid" gas-phase geometries and that geometries are not reoptimized in the liquid phase. In the SM5.42R model (*34–38*), the electrostatic polarization energy ΔG_{EP} is calculated by the generalized Born model (*48–50*) with class IV partial atomic charges (*51*) computed from an electronic wave function. In the SM5CR (*39*) model, ΔG_{EP} is calculated by the conductor-like screening model (*52*) from the density matrix computed from an electronic wave function. Two quantum mechanical methods were used to compute the electronic wave functions and geometries: (i) Semiempirical calculations at the neglect of diatomic differential overlap molecular orbital level of theory with the Austin Model 1 (*53–56*) (AM1) parametrization was used for both the SM5.42R and the SM5CR models. (2) Ab

initio Hartree–Fock (*57*) (HF) calculations with the MIDI! basis set (*58, 59*) were employed with the SM5.42R model.

The G_{CDS} term in its most general form is written as a sum over contributions from individual atoms k of the solute as follows:

$$G_{CDS} = \sum_k \sum_i \sum_j \sum_\delta A_k(\mathbf{R}, r_i) \tilde{\sigma}_{Z_k ij\delta} f_{k_j}(\{Z_{K'}, R_{KK'}\}) S_\delta \quad (6)$$

In this expression, the sum over i involves one or two terms, corresponding to effective solvent radii r_i, and $A_k(\mathbf{R}, r_i)$ is the solvent-accessible surface area of atom k that depends on the complete three-dimensional geometry $\mathbf{R}$ of the solute and on the effective solvent radius; $\tilde{\sigma}_{Z_k ij\delta}$ is an atomic surface tension coefficient that depends on the atomic number Z_k of atom k (Z_k = 1, 6, 7, ... for H, C, N, ...) and on the term indices i, j, and δ; f_{k_j} is a geometrical factor that depends on the atom (k) and the collection of all the atomic numbers $Z_{K'}$ and interatomic distances $R_{KK'}$ in the molecule; and S_δ is a solvent descriptor. We use one solvent descriptor (arbitrarily set equal to unity) for aqueous solvent and six for organic solvents (although there are seven terms in the sum over δ since one of the descriptors, β, appears both as β for S_3 and as β^2 for S_5). The six descriptors are as follows: n, refractive index at the wavelength of the Na D line; α, Abraham's (*60, 61*) hydrogen bond acidity parameters $\Sigma\alpha_2$; β, Abraham's (*60, 61*) hydrogen bond basicity parameter $\Sigma\beta_2$; γ, macroscopic molecular surface tension in units of cal mol^{-1} Å^{-2}; ϕ^2, square of the fraction ϕ of nonhydrogenic solvent atoms that are aromatic carbon atoms (aromaticity); and ψ^2, square of the fraction ψ of nonhydrogenic solvent atoms that are F, Cl, or Br (electronegative halogenicity).

A particular parametrization of an SM5 solvation model is defined by its atomic surface tension coefficients and by the atomic radii in the ΔG_{EP} term (*37*). There are separate sets of coefficients for each choice of model and molecular orbital method. The aqueous and general organic surface tension coefficients have been determined semiempirically by regression on a large number of solvation free energies (more than 2000). Thus, not only do they account for short-range cavitation, dispersion, and solvent–structure interactions, such as hydrogen bonding and the hydrophobic effect, but they also make up semiempirically for any systematic deficiencies in the modeling of long-range electrostatic interactions. The solute surface tension coefficients and atomic radii developed previously for general organic solvents are used in this paper without change for soil.

2.3. Soil Descriptors. Application of the general organic parametrizations to a new medium requires specification of the dielectric constant of the condensed phase and of the six solvent descriptors treated as variables for the computation of G_{CDS}. For a medium such as soil, it is not obvious what values some of these descriptors should have. However, since G_{CDS} is linear in n, α, β, γ, ϕ^2, ψ^2, and β^2, it is possible to obtain parameter values for solvent descriptors by linear regression on experimental data; this minimizes the error in predicted transfer free energies for a particular choice of ϵ. Then we can optimize ϵ by a one-dimensional search over such fits. In such a minimization, however, not all six of the solvent descriptors are independent because γ, β^2, ϕ^2, and ψ^2 are all multiplied by coefficients independent of Z_k and solute geometry. Thus, we may choose to formally set ϕ^2 and ψ^2 to zero and only optimize n, α, β, and γ, which is what we do.

The soil/air partitioning free energies needed for optimizing the soil descriptors are not directly available (*62*). Since no experimental water/air partition coefficient is available for many of the compounds in the K_{OC} data set, a calculated water/air free energy of transfer is used to calculate the soil/air partition coefficient. In particular, we derive soil/air transfer free energies from experimental K_{OC} values and aqueous free energies of solvation computed at the SM5.42R level according to

$$\Delta G^\circ_{soil} = -2.303RT \log P_{soil/w} + \Delta G^\circ_w \quad (7)$$

where the geometry and the aqueous solvation free energy are calculated using the same level of electronic structure theory (in particular AM1 or HF/MIDI!) as that being used to parametrize the soil model.

Many of the molecules in the training set are sufficiently complex so that several conformational minima are available to them. Rigorously, the free energy of solvation is the difference between the statistically weighted free energies of the gas phase and the solvated conformational populations. Prediction of these populations is difficult, and it requires fairly high levels of electronic structure theory to compute conformational energy differences reliably. Therefore, we applied previously described (*30, 32, 38*) rules for choosing which conformation to use and computed the free energy of solvation for only this single conformation. One goal of our study is to learn whether useful results can be obtained this way.

2.4. Software. All SM5.42R/AM1 calculations were carried out with the AMSOL version 6.6 computer program (*63*); all SM5CR/AM1 calculations were carried out with the AMPAC 5.4m2 computer program (*64*). All calculations at the ab initio HF/MIDI! level were carried out with the Gaussian 98 computer program (*65*) augmented with MN-GSM version g1 (*66*).

2.5. Geometries. SM5.42R/AM1 and SM5CR/AM1 calculations are based on gas-phase AM1 molecular geometries, and SM5.42R/HF/MIDI! cacluations are based on gas-phase HF/MIDI! geometries. The Supporting Information includes the HF/MIDI! Cartesian coordinates for all molecules.

3. Results

We optimized new solvent parameters for two methods: SM5.42R/AM1 and SM5.42R/HF/MIDI!. First we optimized soil descriptors for both methods using K_{OC} data set 1. As stated in section 2.2, since solvation models are nonlinear in ϵ, we examined this parameter by a one-dimensional search, optimizing n, α, β, and γ by multilinear regression for each choice of ϵ. In the case of SM5.42R/AM1, the mean unsigned error in K_{OC} falls from 1.17 log units when using a dielectric constant of 7 to 1.10 log units with ϵ = 9.87 (the value for 1-octanol) and to 1.06 log units with ϵ = 15. Further increasing the dielectric constant to 25 reduces the MUE by an additional 0.04 log unit. Since so high a dielectric constant for soil seems unphysical (see below), we settled on ϵ = 15. A choice of ϵ = 15 for SM5.42R/HF/MIDI! followed from a similar analysis. Table 1 lists the condensed-phase parameters optimized for the two cases. The mean unsigned error in log K_{OC} corresponds to an error in the differential free energy of solvation (at 298 K) of 1.4 kcal/mol for SM5.42R/AM1 and of 1.3 kcal/mol for SM5.42R/HF/MIDI!.

To test the robustness of our parametrization, we then added 53 more data to the K_{OC} data set. These consisted of eight carbamates (five N–CO–O and three N–CO–S); seven halogenated hydrocarbons; seven organophosphorus compounds (four P=S and three P=O); four triazoles; three each of acids and nitro compounds; two each of phenols, benzamides, acetanilides, and ureas; one each of aliphatic alcohol, ester, triazine, and sulfone; and finally the molecules endosulfan, imazalil, metalaxyl, metamitron, and oryzalin. We re-optimized n, α, β, and γ with the expanded data set for the SM5.42R/AM1 method and obtained the new values given in Table 1. Table 2 compares statistics for the two parametrizations. First, we see that the RMS error for the new compounds with the original parameters is only 19%

TABLE 1. Effective ϵ, n, α, β, and γ for Various Models

model	data set	m	b	ϵ	n	α	β	γ	MUE
exptl log $P_{o/w}$	$P_{o/w}$	0.63	0.82						0.42[a]
SM5.42R/HF/MIDI!	K_{OC} 1			15	1.379	0.61	0.60	46.0	0.97[b]
SM5.42R/AM1	K_{OC} 1			15	1.504[c]	0.38[c]	0.31[c]	61.4[c]	1.06[b]
SM5.42R/AM1	K_{OC} 2			15	1.541	0.36	0.34	63.3	1.10[d]

[a] For 316 compounds for which experimental $P_{o/w}$ data are available. [b] MUE, mean unsigned error in log K_{OC} over K_{OC} set 1 of 387 compounds. [c] Not final. Final values are those obtained with data set 2. [d] MUE, mean unsigned error in log K_{OC} over K_{OC} set 2 of 440 compounds.

TABLE 2. Comparison of Two Parameterizations for SM5.42R/AM1

method	data[a]	MSE[b]	MUE[c]	RMSE[d]
original[e]	387 original[f]	0.06	1.06	1.359
original[e]	53 new	0.07	1.34	1.615
original[e]	all 440[g]	0.06	1.09	1.392
final[h]	all 440[g]	0.04	1.10	1.389

[a] Data used to compute mean errors. [b] Mean signed error in log K_{OC}. [c] Mean unsigned error in log K_{OC}. [d] Root-mean-squared error in log K_{OC}. [e] Optimized over original data. [f] K_{OC} data set 1. [g] K_{OC} data set 2. [h] Optimized over K_{OC} data set 2.

larger than the RMS error for compounds in the original training set. Second, we see that re-optimizing four solvent descriptors lowers the RMS error by only 0.2%. We conclude from this that data set 1 is sufficiently large and diverse so that adding additional data is not warranted. We will retain the second parametrization for SM5.42R/AM1, but we did not re-optimize SM5.42R/HF/MIDI! with the larger data set.

Table 3 provides, for selected compounds, soil/air and water/air transfer free energies and log K_{OC} values derived therefrom (eq 4) computed with one of the two new SM5.42R parametrizations. The total air/soil transfer free energies are decomposed into bulk electrostatic (ΔG_{EP}) and short-range (G_{CDS}) terms as discussed in section 2.2.

We also carried out fits using eq 2. This was done in two different ways. In the "experimentally based" regressions, both K_{OC} and $P_{o/w}$ were obtained from experiment, and the fit was carried out for the 316 molecules in the $P_{o/w}$ data set. In the "semitheoretical" fits, K_{OC} is from experiment, and $P_{o/w}$ is calculated by an SM5 solvation model. In these fits, the full K_{OC} data set 1 (387 molecules) could be (and was) used. For comparison purposes though, we carried out such fits for the $P_{o/w}$ subset as well the larger set. The resulting slopes and intercepts are given in Table 4.

4. Discussion

4.1. Octanol/Water Partitioning. Of the 316 compounds in the $P_{o/w}$ data set, only 41 are contained in the parametrization sets (*34, 37*) used for the SM5.42R universal solvation model. An additional 12 of the $P_{o/w}$ molecules were included in the augmented parametrization set (*39*) used to develop the SM5CR model. We first examined the accuracy of the SM5 models over the complete data. Table 5 lists the mean signed error (MSE), mean unsigned error (MUE), and root mean square error (RMSE) in predicted log $P_{o/w}$ values over the $P_{o/w}$ data set for each of the three solvation methods employed in this work. Table 5 shows that the mean unsigned error (MUE) in predicted 1-octanol/water partition coefficients over the $P_{o/w}$ data set is slightly less than 0.9 log unit for each of the three parametrizations considered here—this corresponds to an error of about 1.2 kcal/mol in free energies of transfer. This error is larger than the error in the free energy of transfer observed for the 41 molecules found in the SM5 solvation model training set (about 0.5 kcal/mol) but is still reasonable given (i) the highly functionalized character of many of the molecules in the $P_{o/w}$ data set and (ii) the approximation of choosing a single geometry to represent molecules that may exist at equilibrium as a large number of different conformers.

TABLE 3. Data for Selected Molecules at the SM5.42R/HF/MIDI! Level

	water	soil			log K_{OC}	
solute	ΔG^{o}_{w}	ΔG^{o}_{soil}	ΔG_{EP}	G_{CDS}	model	expt
benzene	−1.3	−3.8	−1.0	−2.7	2.7	1.8
naphthalene	−2.7	−6.3	−1.6	−4.7	3.6	3.1
anthracene	−3.9	−8.7	−2.1	−6.6	4.4	4.3
tetracene	−5.2	−11.1	−2.6	−8.5	5.3	5.8
pyrene	−5.1	−10.2	−2.5	−7.7	4.7	4.9
benzo[*a*]pyrene	−6.3	−12.5	−2.9	−9.5	5.5	6.0
benz[*a*]anthracene	−5.2	−11.0	−2.5	−8.5	5.2	5.3
dibenz[*a,h*]anthracene	−6.5	−13.4	−3.0	−10.4	6.0	6.3
1,1,1-trichloroethane	−0.7	−3.7	−1.2	−2.6	3.2	2.0
1,1,2-trichloroethane	−1.3	−4.3	−1.7	−2.7	3.2	1.9
1,1,2,2,-tetrachloroethane	−1.5	−5.1	−1.6	−3.5	3.6	1.9
trichloroethene	−0.2	−2.7	−0.7	−2.1	2.8	2.0
tetrachloroethene	0.3	−2.9	0.0	−2.9	3.2	2.4
biphenyl	−2.7	−7.0	−1.6	−5.5	4.2	3.3
2-chlorobiphenyl	−3.0	−7.7	−1.8	−6.0	4.5	3.5
2,2′-dichlorobiphenyl	−3.5	−8.7	−2.2	−6.5	4.8	3.9
2,4′-dichlorobiphenyl	−3.1	−8.4	−1.8	−6.6	4.9	4.1
2,2′,4-PCB[a]	−3.4	−9.1	−2.0	−7.1	5.2	4.8
2,2′,5-PCB	−3.4	−9.1	−2.0	−7.1	5.2	4.6
2,4,4′-PCB	−3.1	−8.9	−1.7	−7.2	5.2	4.6
2,5,2′,5′-PCB	−3.3	−9.5	−1.9	−7.6	5.5	4.9
2,4,5,2′,5′-PCB	−3.2	−9.9	−1.7	−8.2	5.9	4.6
2,3,4,2′,5′-PCB	−3.6	−10.2	−2.1	−8.1	5.8	4.5
2,4,5,2′,4′,5′-PCB	−3.1	−10.2	−1.5	−8.7	6.2	5.6
2,3,4,2′,3′,4′-PCB	−4.0	−10.9	−2.4	−8.5	6.0	5.1
2,3,4,5,6,2′,5′-PCB	−2.7	−10.1	−1.2	−9.0	6.4	6.0
α-benzene hexachloride	−2.3	−8.5	−2.3	−6.2	5.5	3.3
β-benzene hexachloride	−3.0	−9.2	−2.9	−6.3	5.5	3.5
γ-benzene hexachloride	−2.7	−8.9	−2.6	−6.2	5.4	3.3
aldrin[b]	−2.5	−10.0	−1.6	−8.4	6.5	4.7
endrin[c]	−6.1	−11.7	−3.0	−8.6	5.0	4.1
dieldrin[d]	−6.6	−12.0	−3.3	−8.7	4.9	4.1

[a] PCB, polychlorobiphenyl. [b] 1,2,3,4,10,10-Hexachloro-1,4,4α,5,8,8α-hexahydro-*exo*-1,4-*endo*-5,8-dimethanonaphthalene. [c] 1,2,3,4,10,10-Hexachloro-6,7-*epoxy*-1,4,4α,5,6,7,8,8α-octahydro-*endo*-1,4-*endo*-5,8-dimethanonaphthalene. [d] 1,2,3,4,10,10-Hexachloro-6,7-*epoxy*-1,4,4α,5,6,7,8,8α-octahydro-*exo*-1,4-*endo*-5,8-dimethanonaphthalene.

TABLE 4. Linear Regression of Log K_{OC} with Log $P_{o/w}$ with Various Parameterized Models

source of $P_{o/w}$	data[a]	m	b	MUE[b]	MUE[c]
experiment	316	0.63	0.82	0.42	
SM5CR/AM1	316	0.46	1.47	0.66	0.67
SM5CR/AM1	387	0.44	1.52	0.66	0.67
SM5.42R/AM1	316	0.42	1.66	0.57	0.59
SM5.42R/AM1	387	0.41	1.70	0.57	0.58
SM5.42R/HF/MIDI!	316	0.43	1.63	0.58	0.60
SM5.42R/HF/MIDI!	387	0.42	1.68	0.58	0.59

[a] Data used for the regression to determine the slope and intercept. [b] Mean unsigned error in log K_{OC} for 316 compounds for which there is an experimental log $P_{o/w}$ available. [c] Mean unsigned error in log K_{OC} for all 387 compounds.

TABLE 5. Error in log $P_{o/w}$ for Various Models for 316 Compounds

model	MSE[a]	MUE[b]	RMSE[c]
SM5CR/AM1	−0.35	0.84	1.08
SM5.42R/AM1	−0.55	0.89	1.25
SM5.42R/HF/MIDI!	−0.55	0.86	1.21

[a] MSE, mean signed error. [b] MUE, mean unsigned error. [c] RMSE, root-mean-squared error.

One potential source of error in the model is the use of gas-phase geometries for the condensed-phase calculations (i.e., not permitting the geometry to relax in the condensed-phase environment). Calculations at the SM5.4/AM1 level (*30*–*32*), where explicit geometric relaxation in the condensed phase *is* part of the model, were also carried out to assess this possibility. These results are not shown because no significant systematic improvement in accuracy was evident, and we will not discuss this issue further.

We note that the SM5CR model was developed (*39*) using a training set that, in addition to the free energy of solvation data used to train all other SM5 models, included 76 solvent/solvent partition coefficients. These partition coefficients were chosen in an attempt to provide a more robust model for nitrogen-containing heterocycles, amides, ureas, and compounds containing sulfur and phosphorus. The SM5CR model does appear to be more robust for these kinds of molecules than the SM5.42R models in correlating octanol partitioning—for example, over a set of 13 ureas the MUE for SM5CR in log $P_{o/w}$ is 1.17 log unit, while for SM5.42R/AM1 the MUE is 1.85 log unit, and SM5.42R/HF/MIDI! has an MUE of 1.58. Overall, SM5CR has the smallest error values. All three models, however, appear to be useful for estimating $P_{o/w}$.

4.2. Predicting K_{OC} from $P_{o/w}$. Although our $P_{o/w}$ data set is larger than most previously tested data sets, there is a good correlation between experimental $P_{o/w}$ and K_{OC} values (first line of Table 3), as noted many times previously (*8*–*12*, *17*). The constants we obtain for the slope and intercept in eq 2 are each in the middle of the range of previously reported values (*8*–*12*, *17*), and we obtain an MUE of 0.42 log unit for the fit.

When we carry out our regressions on *computed* $P_{o/w}$ values for compounds in the $P_{o/w}$ data set, we find somewhat smaller values for the slopes and large values of the intercepts of the regression lines than those found for the experimental data, and the MUE of the fits is larger by about 0.2 log unit. (Note that the value of b depends on standard state choices; e.g., if we fit $P_{soil/w}$ instead of K_{OC}, all b values would be 0.96 smaller. Values of b are relative to one another independent of standard state.)

A key strength of the computational models is that they are not limited by the availability of experimental data. Thus, we may compute $P_{o/w}$ values for molecules in the *full* K_{OC} data set 1 and use the slopes and intercepts determined from the $P_{o/w}$ data set to evaluate the robustness of the various regressions. For the SM5CR/AM1, SM5.42R/AM1, and SM5.42R/HF/MIDI! parametrizations, the MUEs over the K_{OC} set are larger than those for the $P_{o/w}$ training set by 0.01, 0.02, and 0.02 log unit, respectively. If we determine optimal values of m and b over the entire K_{OC} set, the MUEs are comparable to those computed for the smaller $P_{o/w}$ set. The MUEs in predicted K_{OC} values are actually better than those found for computing $P_{o/w}$ itself, suggesting that a portion (~35%) of the systematic errors are corrected by the linear regression.

We note that the quality of the different predictive models does not vary much with choice of electronic wave function and geometry, AM1 vs HF/MIDI! (the error in log $P_{o/w}$ by electronic structure calculations with SM5.42R/AM1 at HF/MIDI! rather than AM1 geometries resulted in a reduction in the MUE of less than 0.04 log unit). Thus, given its significantly greater speed, the SM5.42R/AM1 regression provides a particularly efficient model for predicting K_{OC} values (a typical calculation on a molecule having 20 heavy atoms would require, depending on solute conformational flexibility, 1–5 CPU min on a modern workstation).

4.3. Solvent Parameters Characterizing Soil. Next, we focus on the parameters listed in Table 1 and what they tell us about the nature of soil as compared to other condensed phases. The dielectric constant for both parametrizations is 15 (see section 3); this may seem somewhat high as compared to various solvents into which dissolution of a solute has been used to model the sorption process, e.g., benzene and 1-octanol with dielectric constants of 2.27 and 9.87, respectively (*67*). That being said, it must be noted that the polarization free energy depends on the dielectric constant through its proportionality to $(1 - 1/\epsilon)$, and in this context, the difference between 15 for soil and 9.87 for 1-octanol is not particularly large. (0.933 vs 0.899, a difference of 3.8%)

One may also compare the value used here for ϵ with recent measurements for various soils. However, the dielectric constant of soil depends on the particular soil under consideration and its water content, and it is not clear what value is most appropriate for modeling K_{OC}. Orr and Wraith (*68*) measured values ranging (at 298 K) from 4 to 25, depending on the soil and water content, and our value is satisfactorily close to the average of these extremes. Robinson et al. (*69*) measured values ranging from 3 to 25, depending on soil and water content; a value of 15 would correspond to a volumetric water content of about 0.2–0.25.

The values of the remaining descriptors for the "effective solvent" representing soil are particularly interesting in comparison to organic solvents. The SM5 models were parametrized for 90 different organic solvents, and the same models have been shown to effectively predict vapor pressures (self-solvation free energies) for 156 different organic solvents defined by these descriptors (*70*), but soil is a heterogeneous system and not a real solvent; therefore, it is interesting to question whether the soil parameters will provide some insight into the physical processes governing sorption. Indeed, if the soil descriptors were to fall far outside the range of values found for organic solvents, one might worry about the physicality of the model, but this is not the case. The index of refraction optimized for the SM5.42R/AM1 soil model is 1.541, which is very close to the value for benzene, 1.501 (*67*). This is reasonable since linear correlations between benzene/water and soil/water partition coefficients have been noted previously (*7*). The index of refraction for the SM5.42R/HF/MIDI! soil model is 1.379, which is in the range of normal alkanes. Clearly, neither value is counterintuitive.

The optimized γ value for SM5.42R/HF/MIDI! is 46.03 cal $mol^{-1}Å^{-2}$, which is rather close to the value for benzene, 40.62 (*70*). However, this comparison is complicated by the nonzero ϕ^2 for benzene. The γ value for SM5.42R/AM1 is larger.

The values for α and β are significant in both soil models and are in the 0.36–0.60 range. For comparison, we note that carboxylic acids typically have α and β values of about 0.6 and 0.45, respectively, while alcohols have typical values of 0.37 and 0.51. Since the values of α and β in multifunctional solvents generally satisfy approximate group additivity schemes, the soil values appear quite reasonable given the collection of humic acids (*71*) that make up a large portion of the organic content of soil.

The physically reasonable magnitude of the optimized soil descriptors leads to confidence in the robustness of the model, and this situation is reminiscent in this respect of an

analogous development of descriptors for a phospholipid bilayer (*72*).

4.4. Analysis of K_{OC} Values for Particular Solutes. We now proceed with an analysis of K_{OC} values for several sets of compounds found in Table 2. For the sake of simplicity, we restrict our discussion to the slightly more accurate SM5.42R/HF/MIDI! calculations. We emphasize that the compound sets to be discussed were not chosen because the model does particularly well in the computation of these particular absolute K_{OC} values, but rather because they lend themselves to a physical discussion of trends in the data for related compounds.

Polycyclic aromatic hydrocarbons (PAHs) display large negative free energies of transfer into the soil, with the bulk of these energies deriving from short-range (G_{CDS}) effects. There *are* increasingly favorable contributions from ΔG_{EP} as the size of the PAH increases, which is to be expected from increased polarizability, but the major contributor to *variation* in K_{OC} derives from G_{CDS}, and this in turn derives mainly from contributions associated with the *n* descriptor. Thus, from benzene to dibenz[*a*,*h*]anthracene, ΔG_{EP} ranges from −1.0 to −3.0 kcal/mol but G_{CDS} ranges from −2.7 to −10.4 kcal/mol. Although the SM5.42R/HF/MIDI! parametrization tends to somewhat overestimate K_{OC} for the smaller PAHs and underestimate K_{OC} for the larger ones, it does well for predicting trends in different arrangements of the benzene rings. For instance, tetracene and pyrene are each composed of four aromatic rings but differ in their connectivity. Experimentally, tetracene is found to have a larger K_{OC} by 0.9 log unit. The SM5.42R/HF/MIDI! model reproduces the order and quantitatively predicts a difference of 0.6 log unit. Note that the aqueous free energies of solvation for these two molecules are nearly the same, and so are the soil ΔG_{EP} values. The difference derives almost entirely from differences of G_{CDS} in soil. Another interesting observation is that the similarity in measured K_{OC} values for tetracene and the larger benzo[*a*]pyrene derives from balancing increases in the magnitudes of the negative water/air and soil/air transfer free energies.

Rates for environmental reductive dehalogenation of chlorinated hydrocarbons depend on soil/water partitioning because the sediment-bound state tends to be unreactive (*4*). Table 2 includes data for chlorinated ethanes and ethylenes and several different polychlorinated biphenyls (PCBs) and cyclic and bicyclic chloroalkanes.

It is found experimentally that an additional chlorine atom causes the measured K_{OC} value for tetrachloroethylene to be 0.4 log unit greater than that for trichloroethylene. This difference is precisely reproduced by theory, and analysis of the contributing terms suggests that the greater hydrophobicity of tetrachloroethylene is primarily responsible for the increase in K_{OC}. The free energy of aqueous solvation changes by 0.5 kcal/mol while the soil/air transfer free energy changes by only −0.2 kcal/mol. In the chlorinated ethane series, theory agrees with experiment in predicting little difference in K_{OC} values for 1,1,1- vs 1,1,2-trichloroethane. However, the addition of a chlorine atom to make 1,1,2,2-tetrachloroethane is predicted to increase K_{OC} by 0.4 log unit—the same amount as found in the ethylene case—which fails to agree with the experimental situation, where no effect on K_{OC} is reported. In this instance, given trends in other systems on increasing chlorination, it is tempting to speculate that the experimental value may be in error. One might speculate whether the experimental value is in error or whether the trend is a result of significant natural variability in the log K_{OC} data rather than a true "solute" trend. One of the virtues of theoretical models, if they are sufficiently robust, is that they can identify possible experimental errors through analysis of trends or outliers.

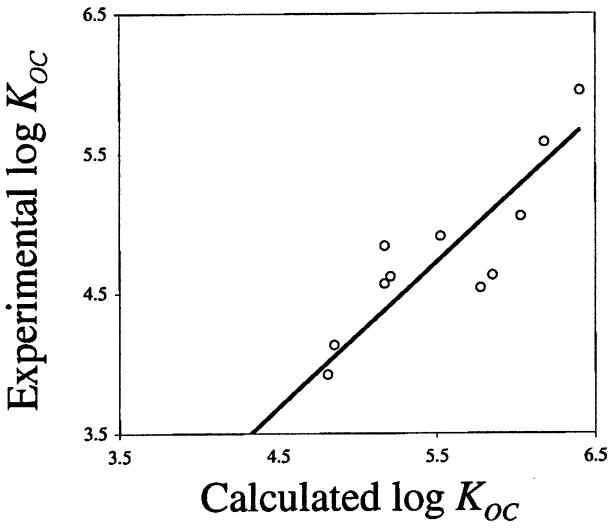


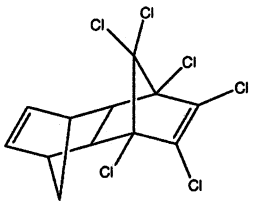
FIGURE 1. Correlation between calculated (SM5.42R/HF/MIDI!) and experimental log K_{OC} values for biphenyl and several polychlorinated biphenyls.

Incremental effects from increasing levels of chlorination are also seen in the polychlorinated biphenyls (PCBs). The SM5.42R model predicts an increase of 0.3 log unit per chlorine atom on going from biphenyl to 2-chlorobiphenyl to 2,2′-dichlorobiphenyl, which is in reasonable agreement with experimental increases of 0.2 and 0.4 log unit, respectively. The large number of data for the PCBs allows us to obtain specific predictions more accurately when attention is focused on a subclass of compounds. For example, if we correlate the SM5.42R predictions with the experimental values for a subset of 13 PCBs, we obtain a regression line having a slope of 1.05 and an intercept of −0.99 log units (Figure 1). This indicates that, while the SM5.42R model overestimates PCB K_{OC} values by about 1 log unit, trends in the data are reproduced on average to within about 5%. Using 1.05 and −0.99 for *m* and *b*, the linear regression corrects the predicted K_{OC} values and reduces the MUE over these compounds to 0.2 log unit.

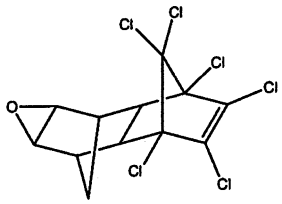
Theory is in good agreement with experiment that there is relatively little difference in K_{OC} values for α-, β-, and γ-benzenehexachloride. (Note that benzene hexachloride is another name for 1,2,3,4,5,6-hexachlorocyclohexane, where β denotes that all chlorines are equatorial, α denotes that the 1,2 chlorines are axial, and γ denotes that the 1,2,3 chlorines are axial.) Analysis of the different contributors to K_{OC} indicates that this insensitivity does not reflect similar free energies of transfer into both phases but rather that changes in water/air transfer free energies are almost exactly balanced by changes in soil/air transfer free energies.

Finally, we consider three related pesticides, aldrin, endrin, and dieldrin (Chart 1). In good agreement with the experimental situation, the predicted K_{OC} values for endrin and dieldrin, which are isomeric, are very similar. Dieldrin derives from epoxidation of aldrin. Experimentally, this causes K_{OC} to drop by 0.6 log unit. The SM5.42R model predicts the effect of epoxidation to be larger, but in the right direction. We note that the overall change in K_{OC} on epoxidation does *not* derive from a less negative soil/air transfer free energy. Instead, the soil/air transfer free energy becomes more negative by 2 kcal/mol. However, the water/air transfer free energy becomes more negative by a larger margin, 4.1 kcal/mol, and this accounts for the drop in K_{OC}. Despite the successes in predicting some of the trends in the last two paragraphs, it is disquieting that some of the absolute values disagree with experiment by 1.7−2.2 log unit, e.g., for aldrin, benzenehexachlorides, and tetrachloroethane.

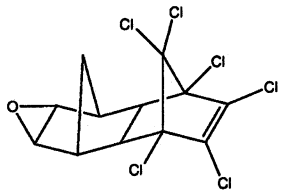
CHART 1



aldrin



dieldrin



endrin

4.5. Significance. We have presented effective solvent descriptors that allow the SM5.42R solvation model to be used to predict soil/water partition coefficients. The mean errors in predicting experimental data are greater than the errors generated using some other methods for predicting selected partitioning coefficients, but the SM5.42R solvation model has several advantages in comparison to previous methods. First, the present models do not use different parameter sets for different classes of solutes, and thus they are applicable to totally new classes of molecules. Second, no experimental data are needed for a new compound once the molecular structure is known, and assigning functional types to compounds to fit them into separate regression schemes is not necessary; this is relevant when comparing to models that require experimental input. Additionally, since the resulting solvent descriptors have reasonable values, it is possible to understand the sources of different partitioning phenomena in cases where the results exhibit significant fragment interactions. Finally, note that the physicality of the predicted numbers makes them useful for identifying possible errors in experimental data, which is important because soil/water partition coefficients are hard to measure accurately.

Acknowledgments

We are grateful to Bill Meylan and Phil Howard of Syracuse Research Corporation for sharing data files and helpful discussions. We acknowledge the participation of Brent Fisher in the early stages of this work. This work was supported by the Alfred P. Sloan Foundation, the Environmental Protection Agency, and the National Science Foundation.

Supporting Information Available

Tables of molecules in each training set, experimental and computed values for various free energies of transfer, and HF/MIDI! geometries (218 pages). This material is available free of charge via the Internet at http://pubs.acs.org.

Literature Cited

(1) Baily, G. W.; White, J. L. *Residue Rev.* **1970**, *32*, 29–92.
(2) Pionke, H. B.; Chesters, G. *J. Environ. Qual.* **1973**, *2*, 29–45.
(3) Karickhoff, S. W. *J. Hydraul. Eng.* **1984**, *110*, 707–735.
(4) Larson, R. A.; Weber, E. J. *Reaction Mechanisms in Environmental Organic Chemistry*; Lewis Publishers: Boca Raton, FL, 1994; pp 10ff.
(5) Hamaker, J. W. In *Environmental Dynamics of Pesticides*; Hague, R., Freed, V. H., Eds.; Plenum: New York, 1975; pp 115–133.
(6) Lyman, W. J. In *Handbook of Chemical Property Estimation Methods*; Lyman, W. J., Reehl, W. F., Rosenblatt, D. H., Eds.; American Chemical Soiciety: Washington, DC, 1990; Chapter 4.
(7) Gawlik, B. M.; Sotiriou, N.; Feicht, E. A.;Schulte-Hostede, S.; Kettrup, A. *Chemosphere* **1997**, *34*, 2525–2551.
(8) Briggs, G. G. *Nature* **1969**, *223*, 1288.
(9) Karickhoff, S. W.; Brown, D. S.; Scott, T. A. *Water Res.* **1979**, *13*, 241–248.
(10) Briggs, G. R. *J. Agric. Food Chem.* **1981**, *29*, 1050–1059.
(11) Karickhoff, S. W. *Chemosphere* **1981**, *10*, 833–846.
(12) Brown, D. S.; Flagg, E. W. *J. Environ. Qual* **1981**, *10*, 382–386.
(13) Schellenberg, K.; Lauenberger, C.; Schwarzenbach, R. P. *Environ. Sci. Techol.* **1984**, *18*, 652–657.
(14) Sabljic, A. *Environ. Sci. Technol.* **1987**, *21*, 358–366.
(15) Lagas, P. *Chemosphere* **1988**, *17*, 205–216.
(16) Sabljic, A.; Güsten, H.; Verhaar, H.; Hermens, J. *Chemosphere.* **1995**, *31*, 4489–4514.
(17) Seth, R.; Mackay, D.; Muncke, J. *Environ. Sci. Technol.* **1999**, *33*, 2390–2394.
(18) Hance, R. J. *Nature* **1967**, *214*, 630–631.
(19) Chiou, C. T.; Peters, L. J.; Freed, V. H. *Science* **1979**, *206*, 831–832.
(20) Lambert, S. M. *J. Agric. Food Chem.* **1967**, *15*, 572–576.
(21) Hance, R. J. *J. Agric. Food Chem.* **1969**, *17*, 667–668.
(22) Pussemier, L.; Borger, R. D.; Cloos, P.; Bladel, R. V. *Chemosphere* **1989**, *18*, 1871–1882.
(23) Park, J. H.; Lee, H. J. *Chemosphere* **1993**, *26*, 1905–1916.
(24) Sabljic, A. *J. Agric. Food Chem.* **1984**, *32*, 243–246.
(25) Bahnick, D. A.; Doucette, W. J. *Chemosphere* **1988**, *17*, 1703–1715.
(26) Okouchi, S.; Saegusa, H. *Bull. Chem. Soc. Jpn.* **1989**, *62*, 922–924.
(27) Meylan, W.; Howard, P. Personal communication of data sets described in Meylan, W.; Howard, P.; Boethling, R. S. *Envrion. Sci. Technol.* **1992**, *26*, 1560–1567.
(28) Tao, S.; Piao, H.; Dawson, R.; Lu, X.; Hu, H. *Environ. Sci. Technol.* **1999**, *33*, 2719–2725.
(29) Reddy, K. N.; Locke, M. A. *Weed Sci.* **1994**, *42*, 453–461.
(30) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385–16398.
(31) Giesen, D. J.; Gu, M. Z.; Cramer, C. J.; Truhlar, D. G. *J. Org. Chem.* **1996**, *61*, 8720–8721.
(32) Giesen, D. J.; Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1997**, *98*, 85–109.
(33) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1998**, *102*, 3257–3271.
(34) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*, 9117–9133; **1999**, *111*, 56241(E).
(35) Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *288*, 293–298.
(36) Hawkins, G. D.; Zhu, T.; Li, J.; Chambers, C. C.; Giesen, D. J.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. Universal Solvation Models. In *Combined Quantum Mechanical and Molecular Mechanical Methods*; Gao, J., Thompson, M. A., Eds.; American Chemical Society: Washington, DC, 1998; Vol. 712, pp 201–219.
(37) Li, J.; Zhu, T.; Hawkins, G. D.; Winget, P. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1999**, *103*, 9–63.
(38) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 2178–2182.
(39) Dolney, D. M.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **2000**, *21*, 340–366.
(40) Cramer, C. J.; Truhlar, D. G. *Chem. Rev.* **1999**, *99*, 2161–2200.
(41) Grathwohl, P. *Environ. Health Perspect.* **1990**, *24*, 1687–1693.
(42) Luthy, R. G.; Aiken, G. R.; Brusseau, M. L.; Cunningham, S. D.; Gschwned, P. M.; Pignatello, J. J.; Reinhard, M.; Traina, S. J.; Weber, W. J., Jr.; Westall, J. C. *Environ. Sci. Technol.* **1997**, *31*, 3341–3347.
(43) Sabljic, A. *Environ. Health Perspect.* **1989**, *83*, 179–190.
(44) Leo, A. J. *Masterfile database.* MedChem Software, BioByte Corp.: P.O. Box 517, Claremont, CA 9711; 1994.
(45) Ben-Naim, A. *Statistical Thermodynamics for Chemists and Biochemists*; Plenum: New York, 1992.

(46) Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1994**, *98*, 4141−4147.
(47) Federer, C. A.; Turcotte, D. E.; Smith, C. T. *Can. J. For. Res.* **1993**, *23*, 1026−1032.
(48) Tapia, O. In *Quantum Theory of Chemical Reactions*; Daudel, R., Pullman, A., Salem, L., Veilland, A., Eds.; Reidel: Dordrecht, 1980; Vol. 2, pp 25−72.
(49) Tucker, S. C.; Truhlar, D. G. *Chem. Phys. Lett.* **1989**, *157*, 164−170.
(50) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(51) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998**, *102*, 1820−1831.
(52) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799−805.
(53) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902−3909.
(54) Dewar, M. J. S.; Zoebisch, E. G. *J. Mol. Struct. (THEOCHEM)* **1988**, *180*, 1−21.
(55) Dewar, M. J. S.; Jie, C. *J. Mol. Struct. (THEOCHEM)* **1989**, *187*, 1−13.
(56) Dewar, M. J. S.; Yate-Ching, Y. *Inorg. Chem.* **1990**, *29*, 3881−3890.
(57) Roothaan, C. C. J. *Rev. Mod. Phys.* **1951**, *23*, 69−89.
(58) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chim. Acta* **1996**, *93*, 281−301.
(59) Li, J.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1998**, *99*, 192−196.
(60) Abraham, M. H. *Chem. Soc. Rev.* **1993**, 73−83.
(61) Abraham, M. H. In *Quantitative Treatments of Solute/Solvent Interactions*; Politzer, P., Murray, J. S., Eds.; Elsevier: Amsterdam, 1994; pp 83−134.
(62) Hippelein, M.; McLachlan, M. S. *Environ. Sci. Technol.* **1998**, *32*, 310−316.
(63) Hawkins, G. D.; Giesen, D. J.; Lynch, G. C.; Chambers, C. C.; Rossi, I.; Storer, J. W.; Li, J.; Winget, P.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *AMSOL version 6.6*; University of Minnesota: Minneapolis, 1999.
(64) Hawkins, G. D.; Dolney, D. M.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *AMPAC version 5.4m2*; Univeristy of Minnestota: Minneapolis, 1999.
(65) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T. A.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Peng, C. Y.; Ayala, P. A.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 98*; Gaussian, Inc.: Pittsburgh, PA, 1998.
(66) Li, J.; Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G.; Frisch, M. J. *Minnesota Gaussian Solvation Module MN-GSM version g1*; University of Minnesota: Minneapolis, 1999.
(67) *CRC Handbook of Chemistry and Physics*, 75th ed.; Lide, D. R., Ed.; CRC Press: Boca Raton, FL, 1995.
(68) Or, D.; Wraith, J. M. *Water Resour. Res.* **1999**, *35*, 371−383.
(69) Robinson, D. A.; Gardner, C. M. K.; Cooper, J. D. *J. Hydrol.* **1999**, *223*, 198−211.
(70) Winget, P.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **2000**, *104*, 4726−4734.
(71) *Organic Chemicals in the Soil Environment*; Goring, C. A. I., Hamaker, J. W., Eds.; Marcel Dekker: New York, 1972; Vol. 1.
(72) Chambers, C. C.; Giesen, D. J.; Hawkins, G. D.; Vaes, W. H. J.; Cramer, C. J.; Truhlar, D. G. In *Rational Drug Design*; Truhlar, D. G., Howe, W. J., Hopfinger, A. J., Blaney, J. M., Dammkoehler, R. A., Eds.; Springer: New York, 1999; pp 51−72.

Received for review January 20, 2000. Revised manuscript received August 7, 2000. Accepted August 7, 2000.

ES0009065