Predicting Adsorption Coefficients at Air−Water Interfaces Using Universal Solvation and Surface Area Models

Casey P. Kelly, Christopher J. Cramer,* and Donald G. Truhlar*

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

Received: October 23, 2003; In Final Form: May 20, 2004

Vapor-phase molecules are adsorbed at air−water interfaces to a much greater extent than can be explained by air−water partition coefficients, indicating that interface adsorption can play an important role, and this can be very important for environmental phenomena. On the basis of a statistical thermodynamic analysis, we separate the observable free energy of adsorption into a dimensionality change and a coupling part so that the modeling effort is correctly focused on the coupling part. On the basis of this analysis, we present two kinds of models for predicting partitioning between the vapor phase and the macroscopic surface of liquid water. The first model, called SM5.0R-Surf, involves atomic surface tensions developed previously for bulk solvation in organic liquids and a set of four solvent descriptors that characterize the properties of the water layer at the air−water interface. The latter descriptors are treated as parameters that are determined empirically by optimization for a set of 85 solutes for which the air−water surface adsorption coefficients ($K_{i/a}$) are known experimentally. The resulting descriptors indicate that interfacial water has increased hydrogen-bond acidity and increased hydrogen-bond basicity as compared to bulk water. A second kind of model involves an empirical correlation of the interfacial-water partition coefficient $K_{i/w}$ with the calculated van der Waals surface area, and this kind of model can be based either on experimental data, yielding the semiempirical surface area (SESA) model, or on theoretical data, yielding the semitheoretical surface area (STSA) model. The SM5.0R-Surf and STSA models should be especially useful for environmental modeling because neither model requires any experimental data about the solute, other than its molecular structure. As an example, we use the above models to calculate air−water adsorption coefficients for 24 different pesticides, chlorinated arenes, and polyaromatic hydrocarbons (PAHs). We also show that several models in the literature can be used successfully even if we substitute calculated instead of experimental data for the solute parameters that they originally required. In related work reported here, the SM5.0R parametrization for predicting free energies of solvation in organic solvents is extended to include solutes containing phosphorus. This extension is based on the experimental free energies of 13 solutes in 9 organic solvents (37 data points). The SM5.0R model extended in this way and the new SM5.0R-Surf model can therefore be used to predict the free energy of solvation at air−water interfaces and in bulk organic liquids for any solute composed of H, C, N, O, F, S, Cl, Br, I, and/or P, whereas the STSA model does not contain parameters that depend on atomic number and can, in principle, be used for any molecule.

1. Introduction

Adsorption from the vapor phase onto a solid or liquid surface plays an important role in the transport of molecules in the environment. One particularly important process is the adsorption of solutes from the vapor phase onto a liquid water surface or onto thin surface films. For example, Valsaraj et al. suggested that unusually high concentrations of organic solutes found in fog droplets might be due to adsorption of these solutes at the air−water interface;[1] shortly thereafter, Goss also used adsorption at the air−water interface to explain measurements of organic solutes in fog droplets.[2] Brusseau and Costanza have pointed out that adsorption at the air−water interface of bubbles can play a large role in the efficiency of air stripping volatile organic compounds (VOCs) from wastewater.[3]

The equilibrium constant corresponding to the adsorption of a solute from the vapor phase (conventionally abbreviated by a for air) onto a bulk water surface (abbreviated i for interface) can be expressed as

$$K_{i/a} = \frac{\Gamma_i}{C_a} \tag{1}$$

where Γ_i is the concentration of the adsorbed solute at the air−water interface in mol/m^2 and C_a is the equilibrium vapor concentration in mol/m^3 of the solute. Several predictive models that relate various physicochemical parameters to the air−water interface adsorption coefficient have already been developed. As an example, we note one such model[4] that uses the solute's vapor pressure and hydrogen-bond basicity (modeled by Abraham's $\sum\beta_2$ parameter,[5,6] which we here call β) to predict its $K_{i/a}$ value. Recently, Roth et al.[7] measured air−water interface adsorption coefficients for a diverse set of 61 organic solutes and derived a general adsorption model based on the solute's air−hexadecane partition coefficient, the parameter β defined above, and the solute's hydrogen-bond acidity (modeled by Abraham's $\sum\alpha_2$ parameter,[5,6] which we here call α).

* To whom correspondence should be addressed. Tel.: (612) 624-0859. Fax: (612) 626-2006. E-mail: cramer@chem.umn.edu (C.J.C.), truhlar@chem.umn.edu (D.G.T.).

10.1021/jp037210t CCC: $27.50
Published on Web 07/29/2004

Other models have been developed for properties that can be directly related to the air−water interface adsorption coefficient, such as the interfacial-water partition coefficient $K_{i/w}$, where w denotes bulk water. Correlations between the interfacial-water partition coefficient and the aqueous solubility,[8] the octanol−water partition coefficient,[9] the liquid molar volume,[9] the hydrophobic molecular surface area,[9] and the first-order molecular connectivity index[10] have all been reported.

For the models described above, knowledge of one or more pieces of experimental data for the solute are required to predict its air−water interface adsorption coefficient; these data could be used to predict $K_{i/a}$ directly, or they could be used to convert $K_{i/w}$ to $K_{i/a}$. In the present article, we first report a model that can be used to predict a solute's air−water interface adsorption coefficient using only its three-dimensional geometry as input. This model has parameters, but they all come from a general training set, and no new data are required for each new solute. This model is based on the simplest member, SM5.0R,[11,12] of the SM5 family[11−22] of universal solvation models. We also present a second type of model for the air−water interface adsorption coefficient that is based on an empirical correlation of $K_{i/w}$ with the solute's calculated van der Waals surface area. This second model requires both the solute's three-dimensional geometry and its bulk water−air partition coefficient to predict its air−water interface adsorption coefficient. In addition to the new models presented here, we also test how well one can estimate the air−water interface adsorption coefficient using calculated instead of experimental values for the solute parameters required by several of the previously developed models mentioned above. As an example of an application of the models developed above, we end the paper by using these models to predict air−water adsorption coefficients for a test set of environmentally important solutes and then compare these values to those predicted using several previously developed models as well as to those reported in the literature.

Throughout the paper, a subscript S denotes transfer from air to an arbitrary bulk solvent; a subscript a/w denotes transfer from bulk water to air; a subscript org/a denotes transfer from air to a general organic solvent; a subscript i/a denotes transfer from air to an air−water interface; a subscript i/w denotes transfer from bulk water to an air−water interface; and subscripts o/w and h/a denote transfer from water to 1-octanol and from air to hexadecane, respectively.

2. Methodology and Theory

2.1. Data Sets. Three data sets, which we call the SM5.0R phosphorus data set, the i/a data set, and the a/w data set, were used in various calculations described below. The SM5.0R phosphorus data set is a subset of the previously described SM5CR training set.[22] It consists of 37 experimental standard-state free energies of transfer from air to nine different organic solvents ($\Delta G^\circ_{org/a}$ values) for 13 neutral solutes containing phosphorus; these were taken from the SM5CR training set. A list of the 13 solutes in the SM5.0R phosphorus data set and their experimental $\Delta G^\circ_{org/a}$ values are given in the Supporting Information.

The remaining two data sets have not been described previously. The second data set (the i/a data set) contains 85 experimental air−water interface adsorption coefficient ($K_{i/a}$) values for 85 solutes at 298 K. The 85 solutes contained in this data set along with their experimental $K_{i/a}$ values are listed in Table 1. The experimental $K_{i/a}$ values[7,8,23−26] used to build this data set were originally determined at four different temperatures: 285.5,[23,24] 288,[7] 293,[25] and 298 K.[8,26] All of the $K_{i/a}$ values determined at 285.5 and 293 K were adjusted to 298 K using experimental values[24,25] for the air−water interface enthalpy of adsorption ($\Delta H_{i/a}$). For many of the solutes whose $K_{i/a}$ values were taken from ref 7, experimental $\Delta H_{i/a}$ values were not available, so for these solutes, calculated $\Delta H_{i/a}$ values were used instead. The regression equation developed by Goss[27] was used to calculate $\Delta H_{i/a}$ values for the aliphatic hydrocarbon solutes. The following equation was developed as part of this work and used for the remaining solutes

$$\Delta H_{i/a}\ (\text{kcal/mol}) = -1.53 \ln K_{i/a}\ (\text{m, 288 K}) - 28.7 \quad (2)$$

The number N of data points used to develop this regression was 13, and the square of the correlation coefficient is $r^2 = 0.979$. The data used to develop this equation can be found in the Supporting Information. A 62-member subset of the 85 solutes in the i/a data set is also in the SM5CR training set.

For the solutes in the i/a data set, the mean $K_{i/a}$ value was used in cases where more than one experimental value for a single solute was available. In these cases, all of the experimental data points were within 1.4 standard deviations of the mean.

All adsorption coefficients in this work employ a standard-state surface concentration of 1 mol/m^2 or the standard-state bulk concentration of 1 mol/m^3. These choices of standard state are discussed in detail below.

The third data set (the a/w data set) contains experimental $K_{a/w}$ values (which are also called Henry's law constants, or K_H values) for 78 of the 85 solutes in the i/a data set (experimental $K_{a/w}$ values were not available for the solutes 2,4-dimethylhexane, 2-methylheptane, 2,4-dimethylheptane, *Z*-2-octene, *E*-2-octene, *n*-perfluorohexane, and 4-fluorotoluene). The experimental $K_{a/w}$ values for 62 of the 78 solutes in this data set were taken from the SM5CR training set. The 17 additional $K_{a/w}$ values were taken from three different sources.[28−30]

2.2. SM5.0R Universal Solvation Model. The adsorption coefficient for a molecule transferring from the gas phase to a water−air interface is related to the standard-state free energy of transfer $\Delta G^\circ_{i/a}$ by

$$\ln K_{i/a} = -\frac{\Delta G^\circ_{i/a}}{RT} \quad (3)$$

where R is the universal gas constant and T is the temperature. Full details of the SM5.0R model for calculating free energies of transfer have been published previously.[11,12] What follows is a brief outline of those details critical to understanding the parametrizations of the present paper.

The SM5.0R model for the standard-state free energy of solvation ΔG°_S (where the 5 in the model name denotes that it is based on SM5 functional forms for atomic surface tensions; the 0 denotes that electrostatic contributions are treated implicitly; and the R stands for rigid, which denotes that its parameters were optimized using rigid, gas-phase geometries) predicts solvation free energies according to

$$\Delta G^\circ_S = \sum_k \sum_{k'} \sum_\delta A_k(\mathbf{R}) \tilde{\sigma}_{Z_k k' \delta} f_{Z_k k' \delta}(\{Z_k, \mathbf{R}\}) S_\delta \quad (4)$$

In this expression, $A_k(\mathbf{R})$ is the exposed van der Waals surface area (A_{vdW}) of atom k (which depends on the complete three-dimensional geometry $\mathbf{R}$ of the solute), $\tilde{\sigma}_{Z_k k' \delta}$ is an atomic surface tension coefficient that depends on the atomic number Z_k of atom k and the indices k' and δ, the function $f_{Z_k k' \delta}$ is a geometrical factor containing switching functions that depend on the atom (k) and the collection of all the atomic numbers Z_k

TABLE 1: Data for Solutes in the i/a Data Set

solute	$\Delta G_{coup(a\rightarrow i)}$	$\Delta G^{\circ}_{3D\rightarrow 2D}$	$\Delta G^{\circ}_{i/a}$	log $K_{i/a}$ (SM5.0R-Surf)	A_{vdW} (Å^2)[a]	log $K_{i/a}$ (SESA)[b]	log $K_{i/a}$ expt (298 K)
n-pentane	−4.72	14.89	10.17	−7.46	136.32	−8.15	−7.22
n-hexane	−5.65	14.95	9.30	−6.82	158.26	−7.61	−6.90
n-heptane	−6.57	14.99	8.42	−6.18	180.20	−7.00	−6.60
n-octane	−7.49	15.03	7.54	−5.53	202.13	−6.47	−6.23
n-nonane	−8.54	15.06	6.52	−4.78	224.07	−5.99	−5.91
n-decane	−9.34	15.10	5.76	−4.22	246.01	−5.32	−5.58
n-undecane	−10.26	15.12	4.86	−3.57	267.96	−5.16	−5.20
2,2,4-trimethylpentane	−6.19	15.03	8.84	−6.48	193.42	−6.79	−6.47
2,4-dimethylhexane	−6.59	15.03	8.44	−6.19	195.40		−6.44
2-methylheptane	−7.11	15.03	7.92	−5.81	199.87		−6.33
2,4-dimethylheptane	−7.50	15.06	7.56	−5.55	217.24		−6.41
cyclohexane	−5.69	14.94	9.25	−6.78	134.89	−7.39	−6.97
cycloheptane	−6.46	14.99	8.52	−6.25	153.62	−6.52	−6.54
cyclooctane	−7.07	15.03	7.96	−5.84	168.08	−6.13	−6.25
Z-2-octene	−7.26	15.03	7.76	−5.69	190.30		−5.96
E-2-octene	−7.35	15.03	7.67	−5.63	192.69		−5.96
1-nonene	−8.11	15.06	6.95	−5.10	215.66	−5.47	−5.72
1-heptyne	−6.13	14.98	8.85	−6.49	164.14	−6.03	−5.80
1-octyne	−7.06	15.02	7.96	−5.84	186.07	−5.42	−5.49
benzene	−5.55	14.92	9.37	−6.87	109.48	−6.70	−6.32
toluene	−6.22	14.97	8.75	−6.42	131.25	−6.09	−5.93
ethylbenzene	−7.09	15.01	7.92	−5.81	153.14	−5.51	−5.58
p-xylene	−6.89	15.01	8.12	−5.96	153.07	−5.43	−5.57
styrene	−6.75	15.00	8.26	−6.05	142.74	−5.36	−5.56
indane	−8.04	15.04	7.00	−5.14	155.43	−4.85	−5.27
1,2,4-trimethylbenzene	−7.56	15.05	7.48	−5.49	173.07	−4.73	−5.13
isopropylbenzene	−7.61	15.05	7.43	−5.45	173.40	−5.12	−5.39
naphthalene	−8.65	15.06	6.42	−4.71	156.48	−4.15	−4.69
biphenyl	−10.29	15.12	4.83	−3.54	189.49	−2.91	−4.10
phenanthrene	−11.63	15.16	3.53	−2.59	200.50	−1.70	−1.00
ethanol	−9.07	14.76	5.69	−4.17	81.67	−4.44	−4.15
1-propanol	−10.02	14.84	4.82	−3.53	103.60	−3.95	−3.83
2-propanol	−9.83	14.84	5.01	−3.67	103.27	−4.13	−3.85
2-methyl-2-propanol	−10.19	14.90	4.72	−3.46	123.71	−3.58	−3.57
2-methyl-1-propanol	−10.36	14.90	4.54	−3.33	123.36	−4.01	−3.61
diethyl ether	−6.40	14.90	8.50	−6.23	123.70	−5.56	−5.30
tetrahydrofuran	−6.83	14.89	8.07	−5.92	106.66	−4.87	−4.85
1,4-dioxane	−8.89	14.95	6.06	−4.45	115.32	−3.46	−3.88
tert-butyl methyl ether	−6.40	14.95	8.56	−6.27	142.94	−4.65	−4.86
diisopropyl ether	−7.72	15.00	7.28	−5.34	165.70	−4.52	−4.83
di-*n*-propyl ether	−7.80	15.00	7.20	−5.28	167.28	−4.45	−4.81
methylphenyl ether	−8.57	15.01	6.44	−4.72	140.54	−4.59	−4.89
pentanal	−9.25	14.95	5.70	−4.18	138.29	−4.19	−4.68
benzaldehyde	−10.74	15.01	4.26	−3.13	132.92	−3.68	−4.13
acetone	−7.69	14.83	7.14	−5.23	95.24	−4.95	−4.78
2-butanone	−8.41	14.89	6.49	−4.76	116.17	−4.47	−4.60
3-methyl-2-butanone	−8.95	14.95	6.00	−4.40	136.89	−4.08	−4.42
cyclopentanone	−9.73	14.94	5.21	−3.82	119.14	−3.64	−3.77
methyl formate	−5.93	14.84	8.91	−6.54	83.65	−6.09	−5.55
ethyl formate	−6.95	14.90	7.95	−5.83	104.73	−5.68	−6.08
methyl acetate	−7.13	14.90	7.77	−5.70	105.26	−5.13	−5.06
ethyl acetate	−8.19	14.95	6.76	−4.96	127.06	−4.53	−4.59
isobutyl acetate	−9.55	15.04	5.48	−4.02	166.72	−3.81	−4.16
nitrobenzene	−9.99	15.05	5.06	−3.71	136.17	−3.54	−3.99
2-nitrotoluene	−10.45	15.08	4.63	−3.40	154.34	−3.35	−3.96
ethanethiol	−5.57	14.85	9.28	−6.80	93.54	−6.97	−6.62
1-propanethiol	−6.53	14.91	8.38	−6.14	115.43	−6.36	−6.25
thiophenol	−8.93	15.02	6.09	−4.46	131.52	−4.83	−5.27
thiophene	−6.93	14.94	8.01	−5.88	98.97	−6.78	−6.43
fluorobenzene	−5.37	14.98	9.61	−7.05	115.60	−6.60	−6.25
n-perfluorohexane	−4.61	15.39	10.79	−7.91	249.30		−7.14
4-fluorotoluene	−6.04	15.02	8.98	−6.59	137.39		−5.90
tetrachloroethene	−6.18	15.14	8.96	−6.57	126.15	−6.82	−6.49
Z-1,2-dichloroethene	−4.07	14.98	10.91	−8.00	95.06	−7.35	−6.56
1,1,2,2-tetrachloroethane	−7.62	15.14	7.52	−5.52	135.21	−4.84	−5.22
1,1,1-trichloroethane	−5.92	15.08	9.16	−6.71	119.71	−6.94	−6.54
1,1,2-trichloroethane	−6.43	15.08	8.64	−6.34	120.51	−5.77	−6.58
1,2-dichloroethane	−5.10	14.99	9.89	−7.25	104.42	−6.25	−6.08
1-chlorobutane	−5.38	14.97	9.59	−7.03	131.37	−6.49	−6.35
1,2,4-trichlorobenzene	−8.63	15.17	6.53	−4.79	157.95	−4.98	−5.13
1,2-dichlorobenzene	−7.58	15.11	7.53	−5.52	141.36	−5.32	−5.41
1,3-dichlorobenzene	−7.66	15.11	7.45	−5.46	142.72	−5.42	−5.57
1,4-dichlorobenzene	−7.66	15.11	7.45	−5.46	142.78	−5.37	−5.57

TABLE 1 (Continued)

solute	$\Delta G_{coup(a\rightarrow i)}$	$\Delta G^{\circ}_{3D\rightarrow 2D}$	$\Delta G^{\circ}_{i/a}$	log $K_{i/a}$ (SM5.0R-Surf)	A_{vdW} (Å^2)[a]	log $K_{i/a}$ (SESA)[b]	log $K_{i/a}$ expt (298 K)
chlorobenzene	−6.60	15.03	8.42	−6.18	126.14	−6.02	−5.90
dichloromethane	−4.01	14.94	10.93	−8.02	82.74	−7.27	−6.63
trichloromethane	−5.36	15.04	9.68	−7.10	99.14	−6.98	−6.40
tetrachloromethane	−6.37	15.12	8.75	−6.42	114.81	−7.37	−6.71
1-bromobutane	−6.50	15.08	8.58	−6.29	135.91	−6.15	−6.07
1-bromopentane	−7.42	15.11	7.69	−5.64	157.84	−5.81	−5.74
bromobenzene	−7.75	15.12	7.37	−5.41	130.62	−5.80	−5.73
1-iodopropane	−5.94	15.15	9.21	−6.75	120.72	−6.71	−6.24
iodobenzene	−8.33	15.20	6.87	−5.04	137.31	−5.35	−4.71
1,1,2-trichloro-1,2,2-trifluoroethane	−5.74	15.18	9.44	−6.92	135.60	−7.98	−6.95
2,2,2-trifluoroethanol	−8.95	14.99	6.04	−4.43	98.91	−4.59	−4.28
2-chloroaniline	−10.71	15.06	4.36	−3.19	139.23	−2.86	−3.68

[a] van der Waals surface area calculated using OMNISOL, version 1.1, with an *m*PW1PW91/MIDI!-optimized geometry. [b] Experimental $K_{a/w}$ value used in eq 20.

and the geometry of the molecule, and S_δ is a solvent descriptor. The forms of the $f_{Z_k k' \delta}$ functions for the SM5.0R model have been published previously.[11,12] For interpretation purposes, it is useful to point out that A_{vdW} is a special case of the solvent-accessible surface area[31] (SASA) in which the solvent radius is negligible compared to the solute atomic radii. We note that the van der Waals radii used in SM5.0R are taken from Bondi.[32]

Models[11,13,16−22] developed to predict only aqueous free energies of solvation do not require solvent descriptors, whereas the universal models[12,14−22] developed for predicting free energies of solvation in organic solvents use six solvent descriptors. In the universal models, there are seven terms in the sum over *j* because one of the descriptors, β, appears both as β for S_3 and as β^2 for S_6. The six solvent descriptors are as follows: *n*, the refractive index at the wavelength of the Na D line; α, the solvent's hydrogen-bond acidity parameter[5,6] $\sum\alpha_2$; β, the solvent's hydrogen-bond basicity parameter[5,6] $\sum\beta_2$; γ, the macroscopic molecular surface tension in units of cal/mol Å^{-2}; ϕ^2, the square of the fraction ϕ of nonhydrogenic solvent atoms that are aromatic carbon atoms; and ψ^2, the square of the fraction ψ of nonhydrogenic solvent atoms that are F, Cl, or Br.

A particular parametrization of the SM5.0R solvation model is defined by its atomic surface tension coefficients. These coefficients differ for the SM5.0R aqueous solvent model and for the SM5.0R universal solvent model. For the SM5.0R universal model, the surface tension coefficients were parametrized against a training set of experimental $\Delta G^{\circ}_{org/a}$ values for 227 neutral solutes in 90 organic solvents (1836 data points).[12] By basing the surface tension coefficients on such a large body of experimental data, the SM5.0R model accounts for a number of solvent effects, such as short-range cavitation, dispersion, and solvent-structure interactions such as hydrogen bonding and the hydrophobic effect. The solute surface tension coefficients for general organic solvents have been published previously[12] and are used in this paper without change for surface water.

2.3. Surface Water Descriptors. To use the SM5.0R universal solvent model, the solvent's descriptors must be known. Several groups have developed methods that can be used to estimate these descriptors for liquid solvents where experimental data are not readily available.[33−35] We use a different approach here and define "surface water" as a unique solvent whose solvent descriptors we determine empirically. Because ΔG°_S is linear in *n*, α, β, γ, ϕ^2, ψ^2, and β^2, it is possible to obtain values for the solvent descriptors by a regression on experimental data. For surface water, we set ϕ and ψ equal to zero and optimize *n*, α, β, and γ.

In SM5.0R, ΔG°_S is the free energy associated with coupling the solute molecule to molecules in bulk liquid solution. Therefore, it is not entirely analogous to the standard-state adsorption free energy ($\Delta G^{\circ}_{i/a}$) given by eq 3 as the latter also includes the cost in free energy associated with transferring a solute from a three-dimensional volume ("air") to a two-dimensional interface, which is an effect that would not vanish in the limit of negligible solute−solvent coupling. As a consequence, the free energy used in this paper for optimizing the surface water descriptors is given by

$$\Delta G_{coup(a\rightarrow i)} = \Delta G^{\circ}_{i/a} - \Delta G^{\circ}_{3D\rightarrow 2D} \tag{5}$$

where $\Delta G_{coup(a\rightarrow i)}$ is defined as the free energy associated with coupling the solute molecule to the air−water interface (as modeled by eq 4) and $\Delta G^{\circ}_{3D\rightarrow 2D}$ is defined as the standard-state dimensionality-change free energy associated with moving the solute molecule from a three-dimensional volume to a two-dimensional surface for a pair of standard-state definitions.

2.4. Dimensionality-Change Free Energy. The standard-state free energies of coupling against which the surface water solvent descriptors will be optimized are different from experimental $\Delta G^{\circ}_{i/a}$ values by a term that accounts for the free energy change associated with a change in the spatial dimensionality of the solute, which we call $\Delta G^{\circ}_{3D\rightarrow 2D}$. We now present a statistical mechanical derivation of this dimensionality-change free energy. This derivation is similar in spirit to Ben-Naim's identification[36] of the liberation free energy as the key to defining free energies of solvation in bulk water.

A particle of mass *m* at temperature *T* has a de Broglie wavelength[37] equal to

$$\Lambda = \left(\frac{2\pi mkT}{h^2}\right)^{-1/2} \tag{6}$$

where *k* is Boltzmann's constant and *h* is Planck's constant. For an ideal gas molecule in *M* dimensions, the molecular translational partition function can be written as[37]

$$q = \left(\frac{L^{\circ}}{\Lambda}\right)^{M} \tag{7}$$

where L° is the standard-state unit of length. The molar translational partition function is then

$$Z = \frac{1}{N_A!}\left(\frac{L^{\circ}}{\Lambda}\right)^{MN_A} \tag{8}$$

where N_A is Avogadro's number. The molar internal energy is defined as[37]

$$U^\circ = kT^2\left(\frac{\partial \ln Z}{\partial T}\right)_{N,V} \quad (9)$$

where N is the number of particles and V is the volume. Employing eq 8 for Z gives

$$U^\circ = \frac{M}{2}RT \quad (10)$$

The molar enthalpy is defined as[37]

$$H^\circ = U^\circ + PV \quad (11)$$

and the molar entropy as[37]

$$S^\circ = k \ln Z + kT\left(\frac{\partial \ln Z}{\partial T}\right)_{N,V} \quad (12)$$

Using eq 8 for Z and using Stirling's approximation for ln(N!) yields

$$S^\circ = R \ln\left[\frac{e^{(M+2)/2}}{N}\left(\frac{L^\circ}{\Lambda}\right)^M\right] \quad (13)$$

The standard-state free energy change associated with a change in spatial dimensionality can now be calculated. For the specific case of $M = 3$ dimensions transferring to $M' = 2$ dimensions, we have

$$\Delta G^\circ_{3D\rightarrow 2D} = \Delta H^\circ_{3D\rightarrow 2D} - T\Delta S^\circ_{3D\rightarrow 2D} \quad (14)$$

Using eqs 10, 11, 13, and 14, we arrive at

$$\Delta G^\circ_{3D\rightarrow 2D} = \Delta(PV)_{3D\rightarrow 2D} - RT \ln\left(\frac{\Lambda}{L^\circ}\right) \quad (15)$$

The PV terms can be calculated from the microcanonical relationship[37]

$$dS = \frac{1}{T}dU + \frac{P}{T}d[(L^\circ)^M] - \frac{\mu}{T}dN \quad (16)$$

where μ is the chemical potential. Equation 16 can be expressed as

$$\frac{P}{T} = \left\{\frac{\partial S}{\partial[(L^\circ)^M]}\right\}_{M,U} \quad (17)$$

Inspection of eq 13 for S° shows that the derivative on the right-hand side of eq 17 is equal to $R(L^\circ)^{-M}$. The ideal gas equation of state

$$PV = RT \quad (18)$$

thus holds in all dimensions if we take the "volume" V of any dimension to be $(L^\circ)^M$, so that $\Delta(PV) = 0$ in eq 15; we then have

$$\Delta G^\circ_{3D\rightarrow 2D} = -RT \ln\left(\frac{\Lambda}{L^\circ}\right) \quad (19)$$

It should be pointed out that standard states of adsorption have been derived[38,39] that attempt to account for the change in spatial dimensionality of the adsorbing solute; however, unlike the dimensionality-change free energy presented above (which is based on a statistical thermodynamic analysis of the free energy change accompanying adsorption), previous standard states of adsorption use ad hoc and model-dependent corrections to the free energy that are based on the distance between molecules in the vapor and adsorbed phases or on the thickness of the adsorbing surface.

2.5. Empirical van der Waals Surface Area Models. Next, we consider a model for the interfacial water equilibrium constant $K_{i/w}$ based solely on the solute's total van der Waals surface area (A_{vdW}). The surface area is a convenient choice for a molecular descriptor given that it is well defined for any possible solute; the values used here are calculated by the OMNISOL, version 1.1, computer program and are the sums over all atoms of a solute of the values used in eq 4, which involves the exposed van der Waals surface areas of the individual atoms in the solute. The sum of these atomic exposed surface areas A_{vdW} is the total van der Waals surface area of the solute.

We start with the water-to-air equilibrium constant $K_{a/w}$, and we calculate the interfacial water equilibrium constant $K_{i/w}$ by

$$K_{i/w} = K_{i/a}K_{a/w} \quad (20)$$

We then convert the result to a standard-state free energy change by

$$\Delta G^\circ_{i/w} = -RT \ln K_{i/w} \quad (21)$$

and remove the dimensionality-change component by writing

$$\Delta G_{coup(w\rightarrow i)} = \Delta G^\circ_{i/w} - \Delta G^\circ_{3D\rightarrow 2D} \quad (22)$$

where

$$\Delta G_{coup(w\rightarrow i)} \equiv \Delta G_{coup(a\rightarrow i)} - \Delta G_{coup(a\rightarrow w)} \quad (23)$$

Then, the coupling part of the free energy is correlated with A_{vdW} by

$$\Delta G_{coup(w\rightarrow i)} = \lambda A_{vdW} \quad (24)$$

where λ is a parameter to be determined.

2.6. Software. All SM5.0R calculations were carried out with the OMNISOL, version 1.1, computer program[40] that is freely available on our Web site.[41] For calculating van der Waals surface areas with nonzero values for the solvent radius and for calculations at the SM5.42R/AM1 level, we used AMSOL, version 7.0.[42] This program is available on our Web site.[43] For all other SM5.42R and SM5.43R calculations, we used the MN-GSM, version 3.1, module.[44] All calculations at the ab initio *m*PW1PW91/MIDI! level were carried out with the Gaussian 03[45] electronic structure package. The AMSOL and OMNISOL computer programs use the ASA algorithm[46] to calculate surface areas.

2.7. Geometries. All SM5.0R calculations are based on gas-phase *m*PW1PW91/MIDI! molecular geometries. (The *m*PW1PW91 method[47] and the MIDI! basis set[48] are explained elsewhere.[49]) For the solutes in the i/a data set, the molecular geometry was represented by a single molecular structure corresponding to the lowest-energy conformer.

3. Results

3.1. Extension of the SM5.0R Universal Model to Phosphorus-Containing Molecules. To enhance its utility for modeling pesticides, we extended the SM5.0R universal model for predicting $\Delta G^\circ_{org/a}$ values to include phosphorus-containing solutes. To this end, we fixed all the parameters for all atoms

other than phosphorus to the values in the SM5.0R universal model paper.[12] For phosphorus ($Z = 15$) there is only one $\tilde{\sigma}^{(n)}_{15k'\delta}$ and it equals $\tilde{\sigma}^{(n)}_{P}$ with $S_\delta = n$ and $f_{15k'\delta} = 1$. The phosphorus radius was set to Bondi's value of 1.80 Å, and the single surface tension coefficient for phosphorus ($\tilde{\sigma}^{(n)}_{P}$) was optimized to minimize the sum of the squares of the error between the 37 predicted and experimental[50] $\Delta G^\circ_{org/a}$ values in the SM5.0R phosphorus data set. The surface tension coefficient $\hat{\sigma}^{(n)}_{P}$ that we obtained is 250.7 cal/mol Å^2.

3.2. SM5.0R-Surf Model. Using the solute surface tension coefficients previously optimized for the bulk liquid phase (ref 12 and section 3.1 above), we optimized a set of solvent descriptors characterizing the air−water interface by minimizing the sum of the squares of the error between the calculated air−water interface coupling free energy ($\Delta G_{coup(a\to i)}$) and the experimental air−water interface coupling free energy [$\Delta G^\circ_{i/a}$ (exp) − $\Delta G^\circ_{3D\to 2D}$] for the 85 solutes in the i/a data set. (Because of eqs 3 and 5, this is equivalent to minimizing the sum of the squares of the error in ln $K_{i/a}$.) We discovered throughout the optimization process that assigning the solvent descriptor n a value equal to its experimental bulk water value (1.342) instead of using its optimized value for surface water (1.398) increased the mean unsigned error in the logarithm (MUEL) of $K_{i/a}$ by less than 0.01. Because this indicates negligible effect of optimizing n on the quality of the model and because using the experimental bulk water value reduces the number of effective parameters required by the model, we chose to use the experimental bulk water value for n instead of the optimized value. With n fixed at 1.342, the optimized values of the other parameters are $\alpha = 1.11$, $\beta = 0.59$, and $\gamma = -144.6$. We will denote the use of these descriptors optimized above, the previously optimized SM5.0R atomic surface tensions,[12] and the atomic surface tension for phosphorus obtained in the previous section by SM5.0R-Surf. This distinguishes the present model, which is designed specifically for the prediction of air−water interface adsorption coefficients, from earlier solvent models developed within this group[11−22] for the prediction of air−bulk solvent partition coefficients.

It is useful to reemphasize here that the above model contains two types of adjustable parameters: the atomic surface tension coefficients and the solvent descriptors. The atomic surface tension coefficients for the atoms H, C, N, O, F, S, Cl, Br, and I were optimized in an earlier paper[12] against a training set of 1836 experimental $\Delta G^\circ_{org/a}$ values and are used in this work without change. The atomic surface tension coefficient for the atom P was optimized as part of this work against a training set of 37 experimental $\Delta G^\circ_{org/a}$ values for 13 phosphorus-containing solutes in 9 different solvents. The remaining adjustable parameters in the SM5.0R model, the solvent descriptors defined above, were also optimized as part of this work using the procedure described above. Thus, all parameters necessary for predicting air−water interface adsorption coefficients have been defined, and the SM5.0R-Surf model can be used to predict the air−water interface adsorption coefficient of any solute containing the atoms H, C, N, O, F, S, Cl, Br, I, and/or P using only the solute's three-dimensional geometry as input.

3.3. Semiempirical Surface Area Model (SESA). Figure 1 shows a plot of $\Delta G_{coup(w\to i)}$ versus the computed A_{vdW} values for the 78 solutes in the a/w data set. The intercept in eq 24 is equal to zero, and this properly accounts for the fact that $K_{i/w}$ vanishes in the limit of negligible solute−solvent coupling. To determine whether optimizing the intercept in eq 24 would empirically improve the correlation between $\Delta G_{coup(w\to i)}$ and

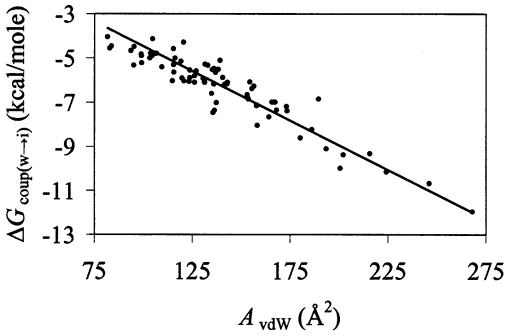

Figure 1. Correlation between $\Delta G_{coup(w\to i)}$ and computed van der Waals surface area A_{vdW} for the 78 solutes in the a/w data set.

A_{vdW}, we also fit the 78 solutes in the a/w data set to an equation with a nonzero value for the intercept. This resulted in a regression equation with an intercept of −0.63 but led to only a slightly higher correlation between $\Delta G_{coup(w\to i)}$ and the computed A_{vdW} ($r^2 = 0.908$ vs 0.864 with zero intercept). Therefore, we use the theoretical value of zero for the intercept in eq 24 throughout the remainder of this work.

The van der Waals surface area appearing in eqs 4 and 24 is a special case of the solvent-accessible surface area (SASA), originally proposed by Lee and Richards,[31] in which the solvent radius is set equal to zero. To investigate whether increasing the solvent radius would strengthen the correlation between A_{vdW} and $\Delta G_{coup(w\to i)}$, we used AMSOL, version 7.0, to compute SASAs (with nonzero solvent radii) for the 78 solutes used above; then we used these areas to develop correlations between $\Delta G_{coup(w\to i)}$ and SASA. (We note here that AMSOL, version 7.0, was used to compute A_{vdW} corresponding to nonzero solvent radii only because OMNISOL, version 1.1, does not allow the user easily to change the solvent radius from its default value of zero, not because using the latter software program leads to erroneous SASAs.) We found that the correlation between $\Delta G_{coup(w\to i)}$ and SASA slowly and systematically decreased when the solvent radius was increased by 0.5 Å or more. Because increasing the solvent radius is equivalent to increasing all solute radii by the same value and because we pointed out above that the correlation between $\Delta G_{coup(w\to i)}$ and SASA decreased with increasing solvent radius, we also attempted to correlate $\Delta G_{coup(w\to i)}$ with surface areas computed using values for the van der Waals radii that were decreased by 0.25 Å. We found that the correlation between $\Delta G_{coup(w\to i)}$ and surface areas also decreased when we tried this. Therefore, we continue this work using the original A_{vdW} values computed by OMNISOL, version 1.1, which yields the following regression equation

$$\Delta G_{coup(w\to i)}\ (\text{kcal mol}^{-1}) = -0.0448\, A_{vdW}\ (\text{Å}^2) \qquad (N = 78,\ r^2 = 0.864) \quad (25)$$

Using eq 25 along with A_{vdW} values calculated by OMNISOL, version 1.1, we recalculated $\Delta G_{coup(w\to i)}$ values for all 78 solutes in the a/w data set. Then, using the $\Delta G^\circ_{3D\to 2D}$ values given in Table 1 along with the $\Delta G_{coup(w\to i)}$ values calculated using eq 25, we used eqs 21 and 22 to calculate $K_{i/w}$ values for all 78 solutes in the a/w data set. Finally, we used experimental $K_{a/w}$ values in eq 20 to convert $K_{i/w}$ values to $K_{i/a}$ values. Throughout the remainder of this work, we refer to this procedure for calculating $K_{i/a}$ as the semiempirical surface area (SESA) model, because it uses experimental $K_{a/w}$ values in eq 20.

For many solutes, reliable experimental $K_{a/w}$ values are not readily available, and as an example, we note the seven solutes in the i/a data set that were excluded from the a/w data set.

TABLE 2: Errors in the Logarithm of $K_{i/a}$ Values Calculated Using the SM5.0R-Surf and SESA Models by Solute Class

		SM5.0R-Surf model			SESA model		
	solute class	no. of solutes	MSEL	MUEL	no. of solutes	MSEL	MUEL
1	unbranched alkanes	7	0.72	0.79	7	−0.29	0.38
2	branched alkanes	4	0.40	0.41	1	−0.32	0.32
3	cycloalkanes	3	0.29	0.29	3	−0.09	0.19
4	alkenes	3	0.40	0.40	1	0.25	0.25
5	alkynes	2	−0.52	0.52	2	−0.07	0.15
6	arenes	11	−0.32	0.45	11	0.18	0.41
7	alcohols	5	0.22	0.22	5	−0.20	0.20
8	ethers	7	−0.69	0.73	7	0.19	0.27
9	aldehydes	2	0.76	0.76	2	0.47	0.47
10	ketones	4	−0.16	0.17	4	0.11	0.19
11	esters	5	−0.32	0.47	5	0.04	0.28
12	nitriles	2	0.41	0.41	2	0.53	0.53
13	thiols	3	0.24	0.37	3	−0.01	0.30
14	sulfides	1	0.56	0.56	1	−0.35	0.35
15	fluorohydrocarbons	3	−0.76	0.76	1	−0.35	0.35
16	chlorohydrocarbons	15	−0.35	0.50	15	−0.15	0.37
17	bromohydrocarbons	3	0.06	0.21	3	−0.07	0.07
18	iodohydrocarbons	2	−0.06	0.16	2	−0.55	0.55
19	multifunctional halogen compounds	3	0.13	0.22	3	−0.17	0.72
	all solutes	85	−0.07	0.47	78	−0.03	0.34

Later in the paper, we use calculated instead of experimental $K_{a/w}$ values in eq 20 along with the procedure outlined above. We refer to this procedure as the semitheoretical surface area (STSA) model, given that it uses theoretical instead of experimental $K_{a/w}$ values in eq 20. It is important to point out that the STSA model is more general than the universal SM5.0R, SM5.0R-Surf, and SESA models, as it would be possible to use a number of methods to calculate the theoretical $K_{a/w}$ values that it requires.

Table 1 provides the $K_{i/a}$ values obtained using the SM5.0R-Surf and SESA models described above. (We note that, in Table 1 and throughout the paper, log denotes logarithm to the base 10, whereas ln denotes the natural logarithm, in accord with standard mathematical conventions.) For the values calculated using the SM5.0R-Surf model, Table 1 decomposes the $\Delta G^\circ_{i/a}$ values into $\Delta G_{coup(a\rightarrow i)}$ and $\Delta G^\circ_{3D\rightarrow 2D}$ terms.

4. Discussion

4.1. Performance of the SM5.0R-Surf Model. The SM5.0R-Surf model gives a MUEL of $K_{i/a}$ of 0.47 for the 85 solutes in the i/a data set. Table 2 provides the MUELs of $K_{i/a}$ arranged by solute class. We see first that the i/a data set is diverse and contains at least one solute from 19 different solute classes, with the most solutes coming from the arene and chlorohydrocarbon solute classes. The MUELs of $K_{i/a}$ for these two solute classes are 0.45 and 0.50, respectively. Much of the MUEL of $K_{i/a}$ for the chlorohydrocarbons comes from *Z*-1,2-dichloroethene, 1,2-dichloroethane, and dichloromethane, for which the SM5.0R-Surf model underestimates $K_{i/a}$ by 1.44, 1.18, and 1.39 log units, respectively. Much of the MUEL of $K_{i/a}$ for the arenes comes from phenanthrene, for which the SM5.0R-Surf model underestimates $K_{i/a}$ by 1.60 log units. The solute classes with the largest MUELs of $K_{i/a}$ are the unbranched alkanes, fluorohydrocarbons, and aldehydes. The SM5.0R-Surf model systematically overestimates $K_{i/a}$ by 0.72 and 0.76 log units for the unbranched alkanes and the aldehydes, respectively, whereas the $K_{i/a}$ values for the fluorohydrocarbons are underestimated by 0.76 log units. This overestimation error for the unbranched alkanes increases with increasing chain length. Despite this error, the SM5.0R-Surf model does correctly predict that the $K_{i/a}$ values for *n*-hexane, *n*-heptane, and *n*-octane are approximately equal to those of their alicyclic analogues cyclohexane, cycloheptane, and cyclooctane, respectively.

To test the robustness of the i/a data set, we also reoptimized the surface water descriptors after removing 25 $K_{i/a}$ values from the i/a data set. The procedure we used for this is as follows: First, we averaged the experimental $K_{i/a}$ values within each solute class containing three or more solutes. Then, we removed approximately 33% of the $K_{i/a}$ values from each solute class, starting with those $K_{i/a}$ values nearest to the average $K_{i/a}$ value. Finally, we reoptimized the solvent descriptors n, α, β, and γ by minimizing the sum of the squares of the error between the predicted air−surface water coupling free energy ($\Delta G_{coup(a\rightarrow i)}$) and the experimental air−water interface coupling free energy [$\Delta G^\circ_{i/a}$ (exp) − $\Delta G^\circ_{3D\rightarrow 2D}$] for the remaining 60 solutes in the i/a data set. We found that assigning the solvent descriptor n a value equal to its experimental bulk water value instead of using its optimized value for surface water increased the MUEL of $K_{i/a}$ by less than 0.01, which is the same result as obtained for the full i/a data set. The final solvent descriptors obtained using the 60-solute i/a data subset described above are $n = 1.342$, $\alpha = 1.08$, $\beta = 0.57$, and $\gamma = -142.5$, which are nearly identical to the values obtained above using the full i/a data set. We continue the remainder of this work using the original set of surface water descriptors optimized against the full i/a data set, although the above result is encouraging.

4.2. Performance of the SESA Model. The SESA model (which requires the solute's experimental $K_{a/w}$ value) gives a MUEL of $K_{i/a}$ of 0.34 for the 78 solutes in the a/w data set. Table 2 provides the MUELs of $K_{i/a}$ arranged by solute class. The solute classes with the largest MUELs of $K_{i/a}$ are the multifunctional halogen compounds, iodohydrocarbons, and nitriles. The SESA model gives a MUEL of $K_{i/a}$ of 0.72 for the multifunctional halogen compounds in the a/w data set and underestimates $K_{i/a}$ by 0.55 and 0.53 log units for the nitriles and iodohydrocarbons, respectively. The SESA model gives a lower MUEL of $K_{i/a}$ for the *n*-alkanes (0.38) than does the SM5.0R-Surf model (0.79); although the SESA model systematically underestimates $K_{i/a}$ for these solutes, this underestimation error decreasing with increasing chain length (recall that the SM5.0R-Surf model systematically overestimated $K_{i/a}$ for the *n*-alkanes and that this overestimation error increased with increasing chain length). Unlike the SM5.0R-Surf model, the SESA model incorrectly predicts that the $K_{i/a}$ value for an *n*-alkane should be greater than the $K_{i/a}$ value for its alicyclic analogue. This result is obtained because the calculated A_{vdW} values for all of the cycloalkanes are significantly smaller than those for their corresponding *n*-alkanes. Using the SESA model instead of the SM5.0R-Surf model to predict $K_{i/a}$ values for the ethers and the alkynes lowers the MUELs of $K_{i/a}$ by 0.46 and 0.37 log units, respectively.

We also wanted to investigate how well the above models perform for the larger solutes considered here, especially given that many solutes of environmental interest (pesticides, polychlorinated biphenyls, polyaromatic hydrocarbons, etc.) are large in size. Table 3 provides the MUEL of $K_{i/a}$ arranged by solute size for both the SESA and SM5.0R-Surf models. The total van der Waals surface areas for most of the solutes considered here (68 of the 85 solutes in the i/a data set and 63 of the 78 solutes in the a/w data set) are between 100 and 200 Å^2. For these solutes, the SM5.0R-Surf and SESA models give MUELs of $K_{i/a}$ of 0.37 and 0.33 log units, respectively. The SESA model performs quite well for the six solutes in the a/w data set with an $A_{vdW} > 200$ Å (MUEL of $K_{i/a}$ of 0.26 log units), whereas the SM5.0R-Surf model performs worse for these same six

TABLE 3: Errors in the Logarithm of $K_{i/a}$ Values Calculated Using the SM5.0R-Surf and SESA Models by Solute Size

	SM5.0R-Surf model			SESA model		
A_{vdW}[a] (Å^2)	no. of solutes	MSEL	MUEL	no. of solutes	MSEL	MUEL
<100	9	−0.52	0.65	9	−0.44	0.44
100−150	44	−0.15	0.41	43	−0.02	0.33
151−200	24	0.05	0.31	20	0.17	0.34
>200	8	0.49	1.08	6	−0.07	0.26

[a] van der Waals surface area calculated using OMNISOL, version 1.1, with an *m*PW1PW91/MIDI!-optimized geometry.

TABLE 4: Surface Water Descriptors

method	$n_{surface}$	$\alpha_{surface}$	$\beta_{surface}$	$\gamma_{surface}$	MUEL[a]
1	1.342[b]	0.82[b]	0.35[b]	71.2[b]	3.16
2	1.398	1.12	0.57	−134.9	0.47
3[c]	1.342[b]	1.11	0.59	−144.6	0.47

[a] Mean unsigned error in the logarithm of $K_{i/a}$ over 85 solutes. [b] Not optimized; experimental bulk water value. [c] This row is the final SM5.0R-Surf model.

solutes (MUEL of $K_{i/a}$ of 1.17 log units). For the two solutes with A_{vdW} > 200 Å that are not in the a/w data set but that are in the i/a data set (2,4-dimethylheptane and *n*-perfluorohexane), the SM5.0R-Surf model gives a MUEL of $K_{i/a}$ of 0.82 log units.

Encouraged by the success of the SESA model, we also tried to develop a similar surface-area-only type of model for $K_{i/a}$ by correlating $\Delta G_{coup(a\rightarrow i)}$ values to the same A_{vdW} values as were used above. This model performed quite poorly, giving a MUEL of $K_{i/a}$ of 1.29 for the 78 solutes in the a/w data set. This suggests that accounting for specific solute−solvent interactions, such as hydrogen-bonding and dispersion interactions, is more critical for modeling adsorption from the vapor phase to the air−water interface (i.e., calculating $K_{i/a}$) than it is for modeling adsorption from the bulk water phase to the air−water interface (i.e., calculating $K_{i/w}$).

To determine whether accounting for specific solute−solvent interactions would improve the performance of the SESA model, we developed another type of model that is similar to the SM5.0R-Surf model described above. For this model, we used the same values for the solute surface tension coefficients as were used above and optimized a new set of solvent descriptors against $\Delta G_{coup(w\rightarrow i)}$ values instead of $\Delta G_{coup(a\rightarrow i)}$ values for the 78 solutes in the a/w data set. This model yielded a set of optimized descriptors for which n, α, and β were all nearly equal to zero. Using these optimized solvent descriptors to calculate $K_{i/a}$ values for the 78 solutes in the a/w data set gave a MUEL of $K_{i/a}$ only 0.02 log units lower than the original SESA model. The results obtained from this model lend further support to the notion that accounting for specific solute−solvent interactions is much more important for predicting $K_{i/a}$ than it is for $K_{i/w}$. This also supports the success of some of the previously developed models in the literature[8−10] (as well as the SESA and STSA models developed here) that found it advantageous to use empirical correlations between various physicochemical parameters and $K_{i/w}$ but not $K_{i/a}$.

4.3. Solvent Descriptors for the Air−Water Interface. We originally optimized the surface water solvent descriptors required by the SM5.0R-Surf model using three different methods. These results are summarized in Table 4. For the first method, we calculated $\Delta G_{coup(a\rightarrow i)}$ values for the 85 solutes in the i/a data set using experimental[28,51] bulk water solvent descriptors. For the second method, we optimized all four solvent descriptors. This reduced the MUEL of $K_{i/a}$ from 3.16 to 0.47. This suggests that the physical and chemical properties of the air−water interface are much different than those of the bulk water phase. For the third method, we fixed n at its experimental bulk water value and then optimized the remaining three solvent descriptors. This increased the MUEL of $K_{i/a}$ by less than 0.01. The solvent descriptors found using this third method yielded the solvent descriptors that we have used to define the SM5.0R-Surf model.

We now focus on the insight that the solvent descriptors in Table 4 can provide into the differences between the structure of bulk water and that of the air−water interface. The value of n that minimizes the MUEL of $K_{i/a}$ over all 85 solutes in the i/a data set is 1.398, although when the experimental bulk water value is used instead, it has little effect on the overall performance of the model. We conclude from this result that, for the solutes examined in this work, either the solvent descriptor n plays the smallest role in the prediction of air−surface water adsorption or it requires the least empirical change from its nominal model value.

We next focus on the remaining three solvent descriptors and what they tell us about differences (if any) between bulk water and the air−water interface. The optimized α and β values for surface water are both higher than their respective bulk water[28] values. This suggests that the air−water interface is more acidic and more basic than bulk water, which is consistent with previous experimental and theoretical work. For example, Gragson and Richmond, using resonant vibrational sum frequency spectroscopy, suggested that there is a lower degree of hydrogen-bond order at the air−water interface than in bulk solution and that the air−water interface is characterized by "dangling" −OH groups.[52] In two earlier papers, Shen et al. used a similar spectroscopic technique to probe the air−water interface and suggested that over 20% of the molecules at the water surface have one −OH group projecting into the vapor.[53,54] In addition, Kuo and Mundy, using results from ab initio molecular dynamics simulations, have suggested that the air−water interface contains far more reactive sites than bulk water.[55]

So far, all of the optimized air−water interface descriptors that we have discussed have fallen inside the range of values found for organic solvents, and in this respect, the result is similar to that for two sets of effective solvent descriptors that were previously determined, namely, those for soil[56] and for a phospholipid bilayer.[57] The macroscopic surface tension solvent descriptor γ in this work does not, however, fall in such a range, and it poses an interesting challenge for several reasons. First, it is unclear how to define the macroscopic surface tension for surface water. One might be tempted to suggest that the value should be the same for interfacial water as for bulk water. However, the surface tension of a bulk liquid can be thought of as an amount of energy paid per unit area by molecules in the bulk solution to create a cavity for the absorbing solute.[58] For molecules adsorbing at the water interface, this energy might be expected to be much smaller, given that only the molecules at the surface need to rearrange significantly to accommodate the adsorbing molecule. We interpret the macroscopic surface tension solvent descriptor for surface water not as a measure of the energy *paid* due to an adsorbing solute, but instead as the energy *gained* because of favorable solute−solvent interactions at the air−water interface. In contrast to the bulk liquid phase, for which γ is often used to quantify cavitation effects accompanying the bulk solvation process,[11−22] the solvent descriptor γ for surface water should be thought of as a sticking

affinity. The optimized value for γ in this work is consistent with this interpretation—it is large and negative in all cases.

4.4. Previous Models. It is useful to use our databases to test the performance of some of the previously developed models that have been used to calculate $K_{i/a}$ values and to compare them to the three models presented here (SM5.0R-Surf, SESA, and STSA). The previously developed models described below have also been summarized in a review by Brusseau and Costanza.[3] In all of the equations below, $K_{i/a}$ and $K_{i/w}$ have units of m. Unless otherwise noted, all values for $K_{i/a}$, $K_{i/w}$, and the various solute parameters are for a temperature of 298 K. Finally, we note that, in this section, the parameters α and β correspond to the *solute* and are used to describe its hydrogen-bonding acidity and basicity, respectively, whereas above, these parameters were optimized for the SM5.0R-Surf model, where they were used to describe the hydrogen-bond acidity and basicity of *solvent* molecules at the air−water interface.

Recently, Roth et al. pointed out a correlation between $K_{i/a}$ and the solute's air−hexadecane partition coefficient ($K_{h/a}$), its α value, and its β value for 60 organic solutes[7]

$$\log K_{i/a}(288\text{ K}) = 0.635 \log K_{h/a} + 3.60\alpha + 5.11\beta - 8.47 \quad (26)$$

Earlier, Goss developed a predictive model for $K_{i/a}$ based on 28 organic solutes that uses the vapor pressure in the liquid (or subcooled liquid) state ($p_L^\bullet$) and the solute parameter β to predict $K_{i/a}$[4]

$$\log K_{i/a} = -0.615 \ln(p_L^\bullet) + 7.86\beta - 10.41 - [385 \ln(p_L^\bullet) - 6037\beta - 6611]\left(\frac{1}{T} - \frac{1}{323}\right) \quad (27)$$

where $p_L^\bullet$ is in units of Pa and T is in units of K.

The majority of previously developed models in the literature, as well as two of the three models developed here (SESA and STSA), take advantage of correlations between various physicochemical parameters and the interfacial water partition coefficient $K_{i/w}$. For example, Hoff et al. reported a correlation between the solute's aqueous solubility S_w (mol/L) and $K_{i/w}$ for 31 hydrophobic (low-aqueous-solubility) solutes[8]

$$\log K_{i/w} = -8.58 - 0.769 \log S_w \quad (28)$$

Hoff et al. noted[8] that the above correlation between $K_{i/w}$ and S_w decreased for solutes with an S_w exceeding 0.1 mol/L, and they suggested using an alternate model[8] (which is not tested here) to predict $K_{i/w}$ values for polar solutes. Recently, Thompson et al. demonstrated for a test set of 75 liquids and 15 solids that accurate predictions for the aqueous solubility can be made using the solute's standard-state air−bulk water free energy of solvation $\Delta G^\circ_{w/a}$ and its vapor pressure $p_L^\bullet$ by[59]

$$S_w = \left(\frac{p_L^\bullet}{p_L^\circ}\right) \exp\left[\frac{-\Delta G^\circ_{w/a}}{RT}\right] \quad (29)$$

where p_L° is the pressure (24.45 atm) of an ideal gas at 1 molar concentration and 298 K. Using eq 29, eq 28 can now be extended to include solutes for which experimental aqueous solubilities are not readily available

$$\log K_{i/w} = -8.58 - 0.769 \log\left(\frac{p_L^\bullet}{p_L^\circ} e^{-\Delta G^\circ_{w/a}/RT}\right) \quad (30)$$

TABLE 5: Comparison between Previous Models and the SM5.0R-Surf, SESA, and STSA Models

model	no. of solutes tested[a]	MUEL[b]
eq 26[c] (expt $K_{h/a}$)	76 (60/60)	0.20[d]
eq 26[c] (calc[e] $K_{h/a}$)	76 (60/60)	0.25
eq 27[f]	73 (28/28)	0.26
SM5.0R-Surf	85 (85/85)	0.47
SESA	78 (78/78)	0.34
STSA	85 (78/78)	0.51
Hydrophobic Solutes[g]		
eq 26[c] (expt $K_{h/a}$)	53 (41/60)	0.19[d]
eq 26[c] (calc[e] $K_{h/a}$)	53 (41/60)	0.28
eq 27[f]	51 (21/28)	0.24
eq 28[h] (expt S_w, $K_{a/w}$)	56 (31/31)	0.35
eq 30 (expt $\Delta G^\circ_{w/a}$, $K_{a/w}$, $p_L^\bullet$)	53 (31/31)	0.33
eq 30 (calc[e] $\Delta G^\circ_{w/a}$, $K_{a/w}$, expt $p_L^\bullet$)	53 (31/31)	0.31
eq 31[i] (expt $K_{o/w}$, $K_{a/w}$)	55 (13/18)	0.50
eq 31[i] (calc[e] $K_{o/w}$, $K_{a/w}$)	55 (13/18)	0.53
eq 32[j] (expt $K_{a/w}$)	56 (11/13)	0.78
eq 32[j] (calc[e] $K_{a/w}$)	56 (11/13)	0.87
SM5.0R-Surf	56 (56/85)	0.47
SESA	56 (56/78)	0.37
STSA	56 (56/78)	0.49

[a] *X* (*Y*/*Z*), where *X* refers to the number of solutes used here to test the model in this row, *Y* refers to the number of solutes tested here that were also in the original data set used to develop the model, and *Z* refers to the total number of solutes in the original data set used to develop the model. [b] Mean unsigned error in the logarithm of $K_{i/a}$. [c] Reference 7. [d] MUEL of $K_{i/a}$ calculated using experimental $K_{i/a}$ values for 288 K. [e] SM5.42R/HF/MIDI!. [f] Reference 4. [g] Solutes with an aqueous solubility (S_w) of less than 0.10 mol/L. [h] Reference 8. [i] Reference 9. [j] Reference 10.

Valsaraj has correlated the octanol−water partition coefficient $K_{o/w}$ with theoretically estimated $K_{i/w}$ values for 18 hydrophobic solutes according to[9]

$$\log K_{i/w} = \frac{\log K_{o/w}}{1.41} - 8.68 \quad (31)$$

Shortly thereafter, Valsaraj also pointed out a correlation between the first-order molecular connectivity index χ^1 (which is calculated using the molecular topology of the solute) and theoretically estimated $K_{i/w}$ values for 13 of the above 18 hydrophobic solutes[10]

$$\log K_{i/w} = 0.2735\chi^1 - 7.53 \quad (32)$$

In the above equations, knowledge of one or more pieces of experimental data for the solute is required to predict its $K_{i/a}$ value, whether it be the data used for the parameters appearing in the equations themselves or the water−air partition coefficients required to convert $K_{i/w}$ to $K_{i/a}$ (as an exception, we note the recent work of Simcik,[60] in which he used calculated $K_{h/a}$ values in eq 26). One of the aims of the present paper is to investigate whether $K_{i/a}$ values can be predicted accurately if calculated instead of experimental data are used along with the equations above. To this end, we used the previously developed SM5.42R dielectric continuum solvent model[17,18] to calculate $K_{h/a}$, $K_{a/w}$, $\Delta G^\circ_{w/a}$, and $K_{o/w}$ values at the SM5.42R/HF/MIDI! level of theory[20] and then substituted these calculated values into some of the equations above. We also tested the STSA model developed here (which uses calculated instead of experimental $K_{a/w}$ values in eq 20) for all 85 solutes in the i/a data set using $K_{a/w}$ values calculated at the SM5.42R/HF/MIDI! level of theory. The results of this effort are summarized in Table 5. Details regarding the solutes chosen to test each of these models are given below.

Roth et al.'s eq 26, which requires the solute parameters $K_{h/a}$ and β, was tested against 76 of the 85 solutes in the i/a data set, which are the solutes for which experimental $K_{h/a}$ and β values are available. Goss's eq 27, which requires the solute parameters $p_L^\bullet$ and β, was tested against 73 of the 85 solutes in the i/a data set, which are the solutes for which experimental $p_L^\bullet$ and β values are available. The remaining three models were only tested against those solutes with an experimental aqueous solubility (S_w) less than 0.10 mol/L, as they were originally developed for use with hydrophobic solutes only. In addition, the remaining models tested here all require the solute's experimental $K_{a/w}$ value to convert $K_{i/w}$ to $K_{i/a}$. All data used to test these remaining three models were taken from a 56-solute subset of the 78-solute a/w data set (which we call the hydrophobic solute data set) that includes solutes with an experimental S_w value less than 0.10 mol/L. Hoff's eq 28, which requires the solute parameter S_w, was tested against all 56 solutes in the hydrophobic solute test set using experimental S_w values. We also tested eq 30 against 53 of the 56 solutes in the hydrophobic solute test set using two methods. For the first method, we used experimental $p_L^\bullet$, $\Delta G^\circ_{w/a}$, and $K_{a/w}$ values. For the second method, we used experimental $p_L^\bullet$ values and calculated (SM5.42R/HF/MIDI!) $\Delta G^\circ_{w/a}$ and $K_{a/w}$ values. Valsaraj's eq 31, which requires the solute parameter $K_{o/w}$, was tested against 55 of the 56 solutes in the hydrophobic solute test set, which are the solutes for which experimental $K_{o/w}$ values are available. Finally, Valsaraj's eq 32, which requires the solute parameter χ^1, was tested against all 56 solutes in the hydrophobic solute test set.

From Table 5, several conclusions can be made regarding previous models and the three models developed here. First, using calculated instead of experimental values in eqs 26−32 does not lead to a significant increase in the MUEL of $K_{i/a}$. The largest increase in the MUEL of $K_{i/a}$ that results from using calculated instead of experimental values is 0.17 when the STSA model is used instead of the SESA model for the 85 solutes in the i/a data set. The MUELs of $K_{i/a}$ for eqs 26 and 32 increase by only 0.05 and 0.03, respectively, when calculated instead of experimental values are used for the solute's $K_{h/a}$, $K_{o/w}$, and $K_{a/w}$ values. Using calculated instead of experimental S_w values in eq 30 actually *decreases* the MUEL of $K_{i/a}$.

Next, we focus on the performance of the three new models developed here (SM5.0R-Surf, SESA, and STSA) in comparison to the previous models that were tested here. The SESA model performs the best of the three models developed in this paper, giving a MUEL of $K_{i/a}$ of 0.34 for the 78 solutes in the a/w data set. Roth et al.'s eq 26 performs better than all of the previous models as well as the models developed here, yielding a MUEL of $K_{i/a}$ of 0.20 over 76 solutes. Even when calculated (SM5.42R/HF/MIDI!) instead of experimental $K_{h/a}$ values are used in eq 26, the model performs well, yielding a MUEL of $K_{i/a}$ of 0.25 over the same 76 solutes as above. Goss's eq 27 also performs quite well for the 73 solutes against which it was tested, yielding a MUEL of $K_{i/a}$ of 0.26. For the hydrophobic solutes (S_w less than 0.10 mol/L) tested here, Roth et al.'s eq 26 and Goss's eq 27 again perform the best, yielding MUELs of $K_{i/a}$ of 0.19 and 0.24, respectively. The SESA model performs the best of the three models developed here for the hydrophobic solutes, yielding a MUEL of $K_{i/a}$ of 0.37.

4.5. Analysis of $K_{i/a}$ Values for Pesticides, Chlorinated Arenes, and Polyaromatic Hydrocarbons (PAHs). Measured concentrations of organic solutes such as pesticides, chlorinated arenes, and PAHs in fog droplets are much higher than what would be expected from Henry's law.[61−65] Several explanations, including adsorption at the air−water interface,[1,2] have been used to explain this apparent increase in the solubility of these solutes in fog droplets. Here, $K_{i/a}$ values have been calculated for a test set of 16 pesticides, 6 chlorinated arenes, and 2 polyaromatic hydrocarbons (PAHs) using the three models developed in this paper (SM5.0R-Surf, SESA, and STSA), as well as two of the previously developed models that were shown to perform quite well above (Roth et al.'s eq 26 and Goss's eq 27). The molecular structures of the 16 pesticides are given in Figure 2. Throughout the rest of this paper, the 24-solute test set described above will be referred to as the environmental compound (EC) test set. Table 6 lists the $K_{i/a}$ values for all 24 solutes in the EC test set calculated using the SM5.0R-Surf, SESA, and STSA models; Roth et al.'s eq 26; and Goss's eq 27. Also given in Table 6 are previously calculated[2] and experimental $K_{i/a}$ values (described below) for some of the solutes in the EC test set. The physicochemical parameters required by the various models used in this section are listed in Table 7 and are described below. Also given in Table 7 are the $\Delta G^\circ_{3D\rightarrow 2D}$ values for the 24 solutes in the EC test set.

The SM5.0R-Surf model requires only the solute's three-dimensional geometry. The three-dimensional geometries of EPTC (*s*-ethyl dipropylthiocarbamate), benfluralin, diazinon, ethoprop, metamitron, trifluralin, 1,2,4-trichlorobenzene, 1,2,3,5-trichlorobenzene, 2,4,5-trichlorobiphenyl (PCB,2,4,5-), 2,2′,3,4,5-pentachlorobiphenyl (PCB,2,2′,3,4,5-), 2,3,7,8-tetrachlorodibenzodioxin (PCDD,2,3,7,8-), 1,2,3,4,7-pentachlorodibenzodioxin (PCDD,1,2,3,4,7-), phenanthrene, and benzo[*a*]pyrene, were optimized in the gas phase at the *m*PW1PW91/MIDI! level of theory, and a single molecular structure corresponding to the lowest-energy conformer was used to calculate the air−water interface coupling free energy $\Delta G_{coup(a\rightarrow i)}$. For the remaining 11 solutes in the EC test set (alachlor, atrazine, cyanazine, desmetryn, diazinon, metolachlor, pendimethalin, prometryn, propachlor, simazine, and terbutryn), more than one conformation was considered explicitly, and the following equation was used to calculate $\Delta G_{coup(a\rightarrow i)}$, $\Delta G_{coup(w\rightarrow i)}$, $\Delta G^\circ_{a/w}$, and $\Delta G^\circ_{h/a}$ values

$$\exp\left[-\frac{\Delta G}{RT}\right] = \sum_C P_C \exp\left[-\frac{\Delta G(C)}{RT}\right] \qquad (33)$$

where P_C is the equilibrium mole fraction of conformation C in the gas phase. The gas-phase absolute energies required for computing P_C in the above equation were calculated at the B97-2/6-31+G(d,p)//B97-2/6-31G(d) level of theory, where the B-97-2 denotes a density functional explained elsewhere.[66]

The SESA model, in addition to the solute's three-dimensional geometry, also requires the solute's experimental $K_{a/w}$ value. Experimental $K_{a/w}$ values for the 24 solutes in the EC test set were taken from several different sources.[67−78] Many of the experimental $K_{a/w}$ values for the 16 pesticides that were considered here were originally determined at temperatures other than 298 K, and these values were adjusted to 298 K using an average enthalpy of volatilization of 47 kJ/mol.[68] All of the $K_{a/w}$ values used here for the chlorinated arenes and PAHs were originally determined at 298 K, so no temperature correction was necessary. The mean value was used in cases where more than one $K_{a/w}$ value was available for a single solute. Three of the 87 experimental $K_{a/w}$ values that were considered here were disqualified because they significantly deviated from the mean $K_{a/w}$ value of the other determinations for that molecule. The remaining experimental data points were within two standard

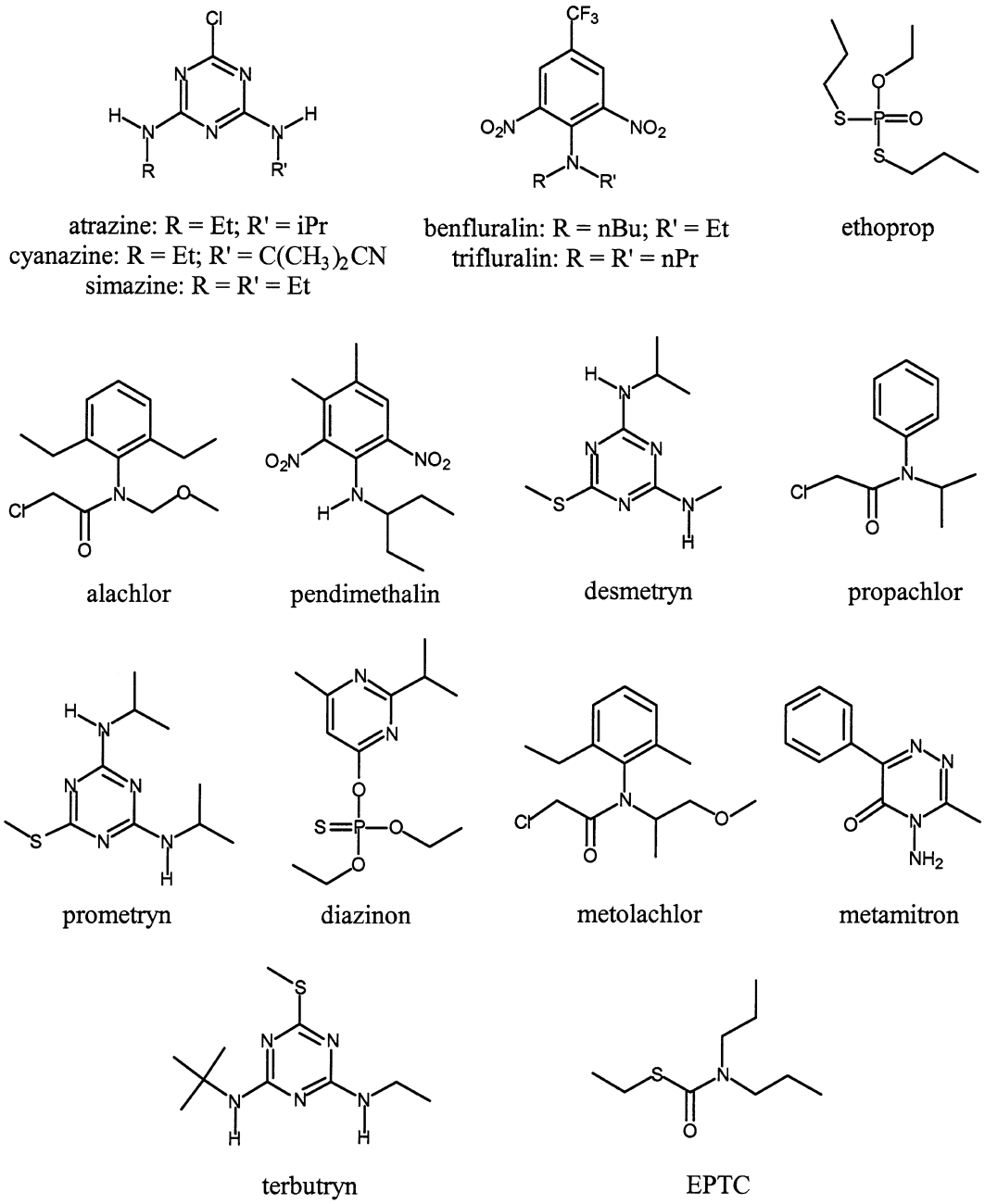

Figure 2. Molecular structures of the pesticides in the EC test set.

deviations of the mean. Only two experimental $K_{a/w}$ values were available for benfluralin, and they differed by over 3 orders of magnitude from one another. We considered both $K_{a/w}$ values, and as a result, the $K_{a/w}$ value for benfluralin is listed as a range of values. All of the experimental $K_{a/w}$ values that were used in the calculations below are given in Table 7.

For the STSA model, $K_{a/w}$ values calculated at the SM5.42R/HF/MIDI! level of theory were used. This level of theory yielded a MUEL of $K_{a/w}$ of 0.95 for 23 of the 24 solutes in the EC test set (benfluralin was not used to calculate the MUEL of $K_{a/w}$ because its experimental $K_{a/w}$ value is listed here as a range of values). We also calculated $K_{a/w}$ values using several other solvent models developed within this group,[11,20,79] and a complete list of all $K_{a/w}$ values calculated as part of this work is provided in the Supporting Information. For ethoprop, the SM5.42R/HF/MIDI! level of theory yields a $K_{a/w}$ value that is 5.69 log units greater than its experimental value. The SM5.43R/*m*PW1PW91/MIDI! level of theory[79] yields a calculated $K_{a/w}$ value that is in better agreement with the experimental value (error of 1.99 log units), so this value was used instead. All of the $K_{a/w}$ values that were used for the STSA calculations in this section are listed in Table 7.

For eq 26, the solute parameters $K_{h/a}$, α, and β are required. Experimental values for $K_{h/a}$, α, and β, are available[80] for 1,2,4-trichlorobenzene, 1,2,3,5-tetrachlorobenzne, and phenanthrene, and these values were used in eq 26. Experimental α and β values are available[80] for benzo[*a*]pyrene, and these were also used in eq 26. The $K_{h/a}$ values for benzo[*a*]pyrene and for the remaining solutes in the EC test set were all calculated at the SM5.42R/HF/MIDI! level of theory. The group contribution method of Hickey and Passino-Reader[34] was used to estimate α and β values for the solutes in the EC test set for which experimental α and β values are not available. The above method could not be used to estimate α and β for ethoprop, given that parameters for phosphorodithioate groups are not available in this method.[34] As a result, $K_{i/a}$ values for this solute

TABLE 6: Calculated and Experimental $K_{i/a}$ Values for the Solutes in the Environmental Compound (EC) Test Set

solute	SM5.0R-Surf[a]	SESA[b]	STSA[c]	eq 26[d]	eq 27[e]	Goss[f]	expt
			Pesticides				
EPTC[g]	0.7	−0.2	−0.6	1.0	0.9		
alachlor	3.3	5.3	3.0	2.3	3.0	−0.1	>0.5[h]
atrazine	4.3	3.8	4.6	6.5	4.3	0.7	
benfluralin	3.2	0.6/2.8	1.0	1.8	1.8		
cyanazine	6.4	7.4	6.6	9.0	6.5		
desmetryn	5.5	4.7	5.7	8.2	5.0		
diazinon	7.3	5.0	4.8	0.4	1.9	1.8	−0.4/1.1[h]
ethoprop	3.7	3.7	−2.0				
metamitron	6.0	6.2	5.9	4.7	4.7		
metolachlor	3.5	6.2	3.6	2.5	3.2		>0.5[h]
pendimethalin	4.3	2.5	2.8	1.8	1.7	0.3	0.6/0.9[h]
prometryn	6.5	5.0	6.0	8.4	4.8		
propachlor	0.8	2.2	1.8	0.8	1.0		
simazine	4.0	3.2	4.6	6.5	4.7		
terbutryn	6.0	4.0	5.5	8.3	4.9		
trifluralin	3.8	2.3	1.6	1.8	2.1		
			Chorinated Arenes				
1,2,4-trichlorobenzene	−4.8	−5.0	−4.6	−5.7	−5.1	−5.1	−5.1[i]
1,2,3,5-tetrachlorobenzene	−4.2	−4.9	−4.4	−5.5	−4.7	−4.7	
PCB, 2,4,5-	−1.6	−1.4	−1.4	−2.9	−2.1	−2.2	
PCB, 2,2′,3,4,5′-	−0.1	−0.4	−0.1	−2.7	−1.4	−1.6	
PCDD, 2,3,7,8-	2.7	0.0	1.1	−1.2	0.5	1.0	
PCDD, 1,2,3,4,7-	2.9	0.9	1.1	−1.2	1.4	1.9	
			PAHs				
phenanthrene	−2.6	−1.7	−1.4	−4.1	−2.0	−2.3	−1.0[i]
benzo[*a*]pyrene	0.5	2.1	2.1	0.5	1.4	0.8	

[a] $\Delta G^{\circ}_{3D\rightarrow 2D}$ values taken from Table 7. [b] $\Delta G^{\circ}_{3D\rightarrow 2D}$ and experimental $K_{a/w}$ values taken from Table 7. [c] $\Delta G^{\circ}_{3D\rightarrow 2D}$ and calculated $K_{a/w}$ values taken from Table 7. [d] $K_{h/a}$, β, and α values taken from columns 7, 8, and 10, respectively, of Table 7. [e] β and $p^{\bullet}_{L}$ values taken from columns 8 and 12, respectively, of Table 7. [f] Caclulated values, taken from ref 2. [g] *S*-Ethyl dipropylthiocarbamate. [h] Determined using eq 34 with enrichment factors taken from refs 61 and 77 (8-μm droplet diameter used). [i] Taken from i/a data set.

TABLE 7: Physicochemical Parameters for Solutes in the EC Test Set

solute	MW (g/mol)	$\Delta G^{\circ}_{3D\rightarrow 2D}$ (kcal/mol)	A_{vdW}[b] (Å^2)	log $K_{a/w}$[a] expt	log $K_{a/w}$[a] calc[c]	log $K_{h/a}$[c,d]	β this work[d]	β Goss[e]	α[d]	mp (K)	$p^{\bullet}_{L}$[f] (Pa) (298 K)
					Pesticides						
EPTC	189.3	15.18	255.4	−2.51	−2.15	4.90	1.25		0		4.23
alachlor	269.8	15.28	320.3	−5.94	−3.66	6.77	1.26	0.48[g]	0	314	5.10×10^{-3}
atrazine	215.7	15.22	247.7	−6.78	−7.64	9.78	1.48	0.61[h]	0.34	449	1.62×10^{-3}
benfluralin	335.3	15.35	333.2	−0.92[i]	−1.34	6.25	1.12		0.15	339	4.40×10^{-2}
				−3.11[j]							
cyanazine	240.7	15.25	263.8	−9.95	−9.15	9.53	1.85		0.54	440	1.62×10^{-4}
desmetryn	213.3	15.22	252.4	−7.57	−8.56	10.09	1.60		0.57	358	7.38×10^{-4}
diazinon	304.3	15.32	350.2	−4.74	−4.48	5.89	1.01	1.01	0		1.31×10^{-2}
ethoprop	242.3	15.25	295.4	−5.18	−3.19[k]						5.32×10^{-2}
metamitron	202.2	15.20	214.5	−10.27	−9.96	8.20	1.30		0.36	440	3.32×10^{-1}
metolachlor	283.8	15.30	336.4	−6.33	−3.78	7.12	1.26		0		3.01×10^{-3}
pendimethalin	281.3	15.30	316.8	−3.28	−3.63	7.50	0.96	0.64[l]	0.17	329	8.11×10^{-3}
prometryn	241.4	15.25	295.3	−6.43	−7.44	10.50	1.60		0.57	392	1.67×10^{-3}
propachlor	211.7	15.21	244.7	−5.31	−4.95	6.61	1.00		0	350	1.09×10^{-1}
simazine	201.7	15.20	226.4	−6.94	−8.32	9.75	1.48		0.34	499	4.23×10^{-4}
terbutryn	241.4	15.25	291.1	−5.66	−7.15	10.26	1.60		0.57	378	1.28×10^{-3}
trifluralin	335.5	15.35	335.4	−2.52	−1.79	6.24	1.12		0.15	322	1.85×10^{-2}
					Chlorinated Arenes						
1,2,4-trichlorobenzene	181.5	15.17	158.0	−0.95	−1.30	(4.43)	(0)	0	(0)		61
1,2,3,5-tetrachlorobenzene	215.9	15.22	172.9	−1.34	−1.09	(4.73)	(0)	0	(0)		19.2
PCB,2,4,5-	257.6	15.27	236.3	−2.05	−2.04	7.52	0.16	0.16	0		4.40×10^{-2}
PCB,2,2′,3,4,5′-	326.4	15.34	268.1	−2.00	−2.38	8.49	0.08	0.08	0		2.30×10^{-3}
PCDD,2,3,7,8-	322.0	15.34	256.3	−2.80	−3.92	7.52	0.48	0.48	0		1.18×10^{-4}
PCDD,1,2,3,4,7-	353.4	15.37	268.8	−3.58	−3.51	7.76	0.45	0.45	0		4.23×10^{-6}
					PAHs						
phenanthrene	178.2	15.05	200.5	−2.83	−3.04	(4.74)	(0.26)	0.20	(0)		0.113
benzo[*a*]pyrene	252.3	15.26	259.0	−4.73	−4.78	10.63	(0.44)	0.30	(0)		2.13×10^{-5}

[a] Logarithm of the water−air partition coefficient (dimensionless Henry's law constant). [b] van der Waals surface area computed using OMNISOL, version 1.1. [c] SM5.42R/HF/MIDI!. [d] Experimental values (ref 80) in parentheses. [e] Reference 2. [f] Vapor pressure in the liquid (or subcooled liquid) state. [g] Oxygen atom of the amide group used to estimate β. [h] Triazine ring and chlorine atom used to estimate β. [i] Reference 78. [j] Reference 76. [k] SM5.43R/*m*PW1PW91/6-31G(d) value used. [l] Phenyl group and both nitro groups used to estimate β.

TABLE 8: Mean Unsigned Differences in the Logarithm (MUDL) of Calculated $K_{i/a}$ Values for the Solutes in the EC Test Set

	SM5.0R-Surf	SESA	STSA	eq 26	eq 27
SESA	1.22[a]				
STSA	1.08[b]	0.95[a]			
eq 26	2.00[a]	2.19[c]	1.79[a]		
eq 27	1.10[a]	0.99[c]	0.75[a]	1.35[a]	
Goss	4.13[d]	3.48[d]	3.13[d]	2.78[d]	2.05[d]

[a] MUDL of $K_{i/a}$ over 23 solutes. [b] MUDL of $K_{i/a}$ over 24 solutes. [c] MUDL of $K_{i/a}$ over 22 solutes. [d] MUDL of $K_{i/a}$ over 4 solutes.

were not calculated using eqs 26 and 27. The α and β values used here are given in the 8th and 10th columns, respectively, of Table 7.

For eq 27, the solute parameters β and $p_L^\bullet$ are required. The β values used in this equation are the same as those used above. Experimental $p_L^\bullet$ values were taken from several different sources.[2,72,76,78,81] Some of the experimental $p_L^\bullet$ values for the 16 pesticides were determined at temperatures other than 298 K, and these were adjusted to 298 K using an average enthalpy of phase change of 50 kJ/mol.[2] All of the $p_L^\bullet$ values for the chlorinated arenes and PAHs were taken from ref 2 and are for 298 K. The melting points given in Table 7 were taken from ref 81 and were used to convert vapor pressures over a solid into the vapor pressures over the corresponding subcooled liquid according to the method of Mackay et al.[82] The mean $p_L^\bullet$ value was used in cases where more than one value was available for a single solute. All of the β and $p_L^\bullet$ values that were used here in eq 27 are given in the 8th and 12th columns, respectively, of Table 7.

Goss has also used eq 27 to predict $K_{i/a}$ values for several of the solutes in the EC test set,[2] and Table 7 also lists the β values that he used in his work (ninth column). All of these β values were estimated using the method of Hickey and Passino-Reader. For most of the solutes, the β values used here in eqs 26 and 27 and those used by Goss in eq 27 are in exact agreement with one another. Exceptions to this are as follows: We have used experimental β values for phenanthrene and benzo[*a*]pyrene, whereas Goss used calculated β values, resulting in small differences between these two sets of values. The β values used here for alachlor, atrazine, and pendimethalin are much larger than those used by Goss. This is because Goss calculated β values for these three solutes considering only the molecule components in the adsorption plane of the molecule, which he defined as the functional group(s) with the strongest hydrogen-bonding ability.[2] Thus, for alachlor, Goss calculated β using only the amide oxygen atom.[2] Similarly, for atrazine, the β value was calculated using only the triazine ring and chlorine atoms,[2] and for pendimethalin, the β value was calculated using only the phenyl ring and the two nitro groups.[2]

From the data in Table 6, we can draw several important conclusions regarding the SM5.0R-Surf, SESA, and STSA models. First, for most of the solutes in the EC test set, the $K_{i/a}$ values calculated using these three models are in fairly good agreement with each other. Table 8 reports mean unsigned differences in the logarithm (MUDLs) of $K_{i/a}$ values calculated using the models developed here, calculated using the two previously developed models (eqs 26 and 27), and previously calculated by Goss.[2] The MUDLs of $K_{i/a}$ values calculated using the SM5.0R-Surf and SESA models and those calculated using the SM5.0R-Surf and STSA models are 1.22 and 1.08 log units, respectively. The largest differences between $K_{i/a}$ values calculated using the SM5.0R-Surf and SESA models are 2.7, 2.7, and 2.3 log units for the solutes metolachlor, PCDD 2,3,7,8-, and diazinon, respectively. Second, contrary to the data presented in Table 3, the SM5.0R-Surf model does not seem to be making as drastic an overestimation error as it did for the larger solutes in the a/w data set (recall that, for the six solutes in the a/w data set with A_{vdW} values > 200 Å^2, the SM5.0R-Surf model gave a MUEL of $K_{i/a}$ of 1.17 log units, whereas the SESA model gave a MUEL of $K_{i/a}$ of 0.26 log units for these same six solutes). For 7 of the 24 solutes in the EC test set, the SESA model gives larger (i.e., more positive) $K_{i/a}$ values than does the SM5.0R-Surf model. It is important to point out that 7 of the 8 solutes in the i/a data set with A_{vdW} values larger than 200 Å^2 are hydrocarbons, which suggests that the overestimation error shown in Table 3 might be the result of the SM5.0R-Surf model's inability to predict $K_{i/a}$ values for hydrocarbons and not the result of some type of systematic overestimation error for larger solutes in general. Unfortunately, reliable experimental $K_{i/a}$ values for solutes other than those included in the i/a data set, which could be used to test the above hypothesis, are not currently available.

Because differences between $K_{i/a}$ values calculated using the SESA and STSA models depend solely on the theoretical method chosen to calculate *bulk water*−air partition coefficients (i.e., both of these models yield identical $K_{i/w}$ values for a given solute), comparisons between these two models cannot offer much additional insight into the relative performance of previous models for the air−water interface and the models developed here. Nevertheless, it is encouraging that, for the majority of the solutes in the EC test set (with a notable exception being the solute ethoprop), the SESA and STSA models give quite similar results. This is encouraging because, for many solutes, reliable experimental $K_{a/w}$ values have not been measured or are not readily available in the literature.

Next, we compare the $K_{i/a}$ values calculated using the two previously developed models (eqs 26 and 27) to each other and to the values calculated using the models developed here. The MUDL of $K_{i/a}$ values calculated using eqs 26 and 27 is 1.35 log units, which is slightly higher than the MUDL of $K_{i/a}$ values calculated using the SM5.0R-Surf and SESA models developed here (1.17 log units). The largest differences between $K_{i/a}$ values calculated using eqs 26 and 27 are 3.6, 3.4, and 3.2 log units for the solutes prometryn, terbutryn, and desmetryn, respectively. All three of these solutes contains a thiol functional group, and they are the only solutes in the EC test set that do. For these three solutes, the $K_{i/a}$ values calculated using eq 28 (which does not require the solute parameter $K_{h/a}$) are in quite good agreement with the values calculated using the three models developed here. This suggests that the SM5.42R universal model might be making some type of systematic error in calculating $K_{h/a}$ for larger solutes containing sulfide groups.

The best overall agreement between any of the models developed here and the previously developed models is between the STSA model and eq 27 (MUDL of 0.75 log units). The SESA model, which requires experimental $K_{a/w}$ values, gives a MUDL of $K_{i/a}$ of 0.99 log units when compared to the values calculated using eq 27. The SM5.0R-Surf model and eq 27 also yield similar $K_{i/a}$ values for the solutes in the EC test set. The MUDL of $K_{i/a}$ for the values calculated using these two models is 1.11 log units. It is encouraging that, for most of the solutes in the EC test set, the $K_{i/a}$ values calculated using the three models developed here (SM5.0R-Surf, SESA, and STSA) are in fair agreement with the values calculated using eqs 26 and 27, especially given that it was shown above that eqs 26 and 27 performed quite well when tested against solutes for which experimental $K_{i/a}$ values are available. A notable exception is

diazinon, for which the SM5.0R-Surf, SESA, and STSA models all yield $K_{i/a}$ values that are significantly higher than those calculated using eqs 26 and 27. Because eqs 26 and 27 both use the solute parameter β but the SM5.0R-Surf, SESA, and STSA models do not (the SM5.0R-Surf model uses the parameter β to describe the *solvent*, and this value was optimized above), we used another group contribution method[33] to calculate the parameter β for diazinon, and then used this value in eqs 26 and 27 to calculate $K_{i/a}$. This second method[33] yields a β value of 2.59, which is much higher than the β value estimated using the method of Hickey and Passino-Reader (1.01). Using this value of 2.59 in eqs 26 and 27 to calculate $K_{i/a}$ values for diazinon yields values of 8.2 and 8.5 log units, respectively, which are in somewhat closer agreement with the values calculated using the three models developed here. Although not used here, we also calculated β values using this second method[33] for 18 other solutes in the EC data set. For these solutes, the agreement between the β values calculated by the two methods is quite good (MUD = 0.18). Because reliable experimental $K_{i/a}$ values are not available for diazinon, it is unclear which of these methods (if any) yields the correct β value for diazinon, so we will continue the rest of this work using the original value for β that was calculated with the method of Hickey and Passino-Reader (1.01). The above situation, however, does illustrate one of the serious problems faced by predictive models that rely on the use of physicochemical parameters to describe the adsorbing solute (including the SESA and STSA models developed here), namely, that, for solutes where limited experimental data exist, other predictive models must be used to derive the physicochemical parameters required by these models.

Finally, we compare the $K_{i/a}$ values calculated using the five models described above (SM5.0R-Surf, SESA, STSA, eq 26, and eq 27) to previously calculated $K_{i/a}$ values and some of the experimental data available for these solutes. Experimental $K_{i/a}$ values have been determined using inverse gas chromatography for the solutes 1,2,4-trichlorobenzene[7] and phenanthrene.[26] These two $K_{i/a}$ values are also included in the i/a data set. All five models yield $K_{i/a}$ values for 1,2,4-trichlorobenzene that are in good agreement with the experimental value. In contrast, all five models underestimate the experimental $K_{i/a}$ value for phenanthrene, suggesting that this experimental value might be in error.

Experimental $K_{i/a}$ values have not been determined using inverse gas chromatography for any of the remaining solutes in the EC test set, although experimentally measured enrichment factors in filtered fogwater (defined as the ratio of the bulk water−air partition coefficient $K_{a/w}$ to the measured distribution ratio between air and filtered fogwater) have been reported for alachlor,[77] atrazine,[77] diazinon,[61] metolachlor,[77] and pendimethalin.[77] The experimental $K_{i/a}$ values listed in the last column of Table 6 for these solutes were determined using the following equation[1,2]

$$\mathrm{EF} = K_{a/w}(SK_{i/a} + K_{a/w}{}^{-1}) \qquad (34)$$

where EF is the reported enrichment factor in filtered fogwater and S is the surface area-to-volume ratio for a fog droplet. In the above equation, we used a droplet diameter of 8 μm to calculate S. It is worth noting here that an experimental EF value of 0.05 has been reported for atrazine,[77] although using this value in the above equation leads to a negative value for $K_{i/a}$, which does not make physical sense.

From Table 6, we see that the $K_{i/a}$ values calculated using the three models developed here (SM5.0R-Surf, SESA, and STSA) agree poorly with the experimental $K_{i/a}$ values derived using experimental EF values. With the exception of diazinon, the $K_{i/a}$ values obtained using eqs 26 and 27 are also in poor agreement with the experimental values. One possible explanation for this poor agreement is that, for larger solutes with multiple functional groups, a significant portion of the molecule will be unable to interact with solvent molecules at the air−water interface, leading to an overestimation error in their $K_{i/a}$ values. This potential problem has been pointed out by Goss,[2] and indeed, the $K_{i/a}$ values reported by Goss[2] for alachlor and pendimethalin are both in good agreement with the experimental values. Furthermore, this good agreement is almost entirely attributable to the values for the parameter β that Goss used in eq 27, which were calculated using only the partial structure of the solute.[2]

However, this analysis requires one to assume that there is no real "width" to the air−water interface, which does not seem physically realistic. Another explanation for the above discrepancies is that eq 34 might not be valid for some of the solutes investigated here (i.e., other factors might actually decrease the enrichment of some organic compounds in fog droplets). As noted above, when eq 34 is used to calculate $K_{i/a}$ for atrazine using its experimental EF value of 0.05, a negative value for $K_{i/a}$ is obtained, which is clearly nonphysical. Another problem with using eq 34 to derive $K_{i/a}$ values is that these values are sensitive to the value chosen for S, and it is not immediately obvious what value should be used for this parameter. Following the work of Goss,[2] we have used a droplet diameter of 8 μm (which leads to an S value of 75000 m^{-1}), although Valsaraj[1] has used droplet diameters between 1 and 4 μm in eq 34. (Of course, using a smaller value for S in eq 34 would lead to even smaller $K_{i/a}$ values, thus leading to an even larger disagreement with the majority of the values given in Table 6.) Finally, the experimental EF values that were considered here were originally calculated using water−air partition coefficients and fogwater air distribution factors determined at quite different temperatures,[61,77] which will most likely have a large influence on the value for EF. Unfortunately, additional experimental data for these solutes are not available in the literature.

For the reasons pointed out above, we feel that the $K_{i/a}$ values derived using eq 34 are not reliable and that the overall performance of the three models presented here (SM5.0R-Surf, SESA, and STSA) for the solutes in the EC test set is satisfactory. For most of the solutes in the EC test set, the $K_{i/a}$ values calculated using the three models developed here are in relatively good agreement with each other and with the $K_{i/a}$ values calculated using two of the previous developed models from the literature (eqs 26 and 27). In addition, two of the models presented here (SM5.0R-Surf and STSA), should be useful for modeling the adsorption of environmentally important solutes at the air−water interface because they require no experimental data for the solute to predict its $K_{i/a}$ value.

5. Conclusions

In this paper, we have partitioned the free energy of adsorption into separate terms that account for the free energy associated with coupling the solute to molecules at the bulk water surface and for the free energy associated with a change in the spatial dimensionality of the solute. This allows for meaningful comparisons to be made between the free energy associated with specific solute−solvent interactions at the water surface and the free energy associated with specific solute−solvent interactions in bulk water without the introduction of ad hoc terms based on either the distance between molecules

in the vapor and adsorbed phases or the thickness of the adsorbing surface.

Using this statistical mechanical method to identify the coupling part of the free energy of adsorption, we have optimized effective solvent descriptors that allow the SM5.0R universal solvation model to be used to predict the air−water interface adsorption coefficient $K_{i/a}$ for any solute containing H, C, N, O, F, S, Cl, Br, I, and/or P atoms. The solvent descriptors were obtained by minimizing the root-mean-square error between predicted and experimental ln $K_{i/a}$ values for 85 solutes, and the model involving these descriptors is denoted as SM5.0R-Surf. We have also shown that adsorption from the bulk water phase to the air−water interface is strongly dependent on the solute's calculated van der Waals surface area (A_{vdW}), and we have developed two additional models, denoted SESA and STSA, that can be used to predict the interfacial-water partition coefficient $K_{i/w}$ based solely on the total exposed van der Waals surface area of the solute.

The SESA model, which requires knowledge of the solute's experimental bulk water−air partition coefficient $K_{a/w}$ to convert $K_{i/w}$ to $K_{i/a}$, gives the lowest errors between predicted and experimental adsorption coefficients of the three models developed here. The remaining two models developed in this paper (SM5.0R-Surf and STSA) are less accurate than the SESA model and several of the previously developed models tested here, but the SM5.0R-Surf and STSA models have a set of advantages not shared by many of the previously developed models in the literature or by the SESA model (the SESA model does share the first and third advantages, but not the second). First, the SM5.0R-Surf and STSA models have been tested for a wide variety of solutes and are designed to be applicable to any organic solute, whereas several of the previous models are applicable to restricted classes of solutes. Second, no experimental data are needed for a new solute once the molecular structure is known. This is very important because experimental data are not readily available for many solutes of environmental interest. Third, the SM5.0R-Surf and STSA models do not use atom typing (i.e., the atoms in a solute do not have to be assigned to functional groups or atom types, as in force fields used in molecular mechanics calculations). This is convenient, and beyond convenience, it is especially important for solutes containing functional groups for which there are not enough data to define a parameter, which is a serious problem for models with types. Additionally, both models allow for very rapid computations. This is especially useful for performing calculations on large systems, such as many environmentally and/or biologically active molecules, or for performing calculations on large libraries of compounds. Finally, we note that the solvent descriptors optimized for the SM5.0R-Surf model can be used to make meaningful inferences about the nature of the air−water interface.

Acknowledgment. Special thanks go to Jason Thompson and Matt Simcik for providing help and insightful comments throughout the course of this work. This work was supported in part by the National Science Foundation.

Supporting Information Available: The Supporting Information includes a full list of all solutes that were used in the SM5.0R phosphorus data set (13 solutes) along with their experimental $\Delta G^{\circ}_{org/a}$ values, the experimental $K_{i/a}$ and $\Delta H_{i/a}$ values used to parametrize eq 2, and the $K_{a/w}$ values calculated with the SM5.0R, SM5.42R, and SM5.43R models for the 24 solutes in the environmental compound (EC) data set. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Valsaraj, K. T.; Thoma, G. J.; Reible, D. D.; Thibodeaux, L. J. *Atmos. Environ.* **1993**, *27*, 203.
(2) Goss, K.-U. *Atmos. Environ.* **1994**, *28*, 3513.
(3) Costanza, M. S.; Brusseau, M. L. *Environ. Sci. Technol.* **2000**, *34*, 1.
(4) Goss, K.-U. *Environ. Sci. Technol.* **1994**, *28*, 640.
(5) Abraham, M. H. *Chem. Soc. Rev.* **1993**, 73.
(6) Abraham, M. H. In *Quantitative Treatments of Solute/Solvent Interactions*; Politzer, P., Murray, J. S., Eds.; Theoretical and Computational Chemistry Series 1; Elsevier: Amsterdam, 1994; pp 83−134.
(7) Roth, C. M.; Goss, K.-U.; Schwarzenbach, R. P. *J. Colloid Interface Sci.* **2002**, *252*, 21.
(8) Hoff, J. T.; Gillham, R. W.; Mackay, D.; Shiu, W. Y. *Environ. Sci. Technol.* **1993**, *27*, 2174.
(9) Valsaraj, K. T. *Chemosphere* **1988**, *17*, 875.
(10) Valsaraj, K. T. *Chemosphere* **1988**, *17*, 2049.
(11) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1997**, *101*, 7147.
(12) Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Org. Chem.* **1998**, *63*, 4305.
(13) Chambers, C. C.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 16385.
(14) Giesen, D. J.; Gu, M. Z.; Cramer, C. J.; Truhlar, D. G. *J. Org. Chem.* **1996**, *61*, 8720.
(15) Giesen, D. J.; Hawkins, G. D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1997**, *98*, 85.
(16) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. B* **1998**, *102*, 3257.
(17) Zhu, T.; Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **1998**, *109*, 9117; **1999**, *111*, 5624(E); **2000**, *113*, 3930(E).
(18) Li, J.; Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1998**, *288*, 293.
(19) Hawkins, G. D.; Zhu, T.; Li, J.; Chambers, C. C.; Giesen, D. J.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. In *Combined Quantum Mechanical and Molecular Mechanical Methods*; Gao, J., Thompson, M. A., Eds.; ACS Symposium Series 712; American Chemical Society: Washington, DC, 1998; pp 201−219.
(20) Li, J.; Zhu, T.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *Theor. Chem. Acc.* **1999**, *103*, 9.
(21) Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2000**, *104*, 2178.
(22) Dolney, D. M.; Hawkins, G. D.; Winget, P.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **2000**, *21*, 340.
(23) Chatterjee, A. K.; King, J. W.; Karger, B. L. *J. Colloid Interface Sci.* **1972**, *41*, 71.
(24) Hartkopf, A.; Karger, B. L. *Chem. Res.* **1973**, *6*, 209.
(25) Dorris, G. M.; Gray, D. G. *J. Phys. Chem.* **1981**, *85*, 3628.
(26) Raja, S.; Yaccone, F. S.; Ravikrishna, R.; Valsaraj, K. T. *J. Chem. Eng. Data* **2002**, *47*, 1213.
(27) Goss, K.-U.; Schwarzenbach, R. P. *Environ. Sci. Technol.* **1999**, *33*, 3390.
(28) Abraham, M. H.; Andovian-Haftvan, J.; Whiting, G. S.; Leo, A. *J. Chem. Soc., Perkin Trans. 2* **1994**, 1777.
(29) Yaffe, D.; Cohen, Y.; Espinosa, G.; Arenas, A.; Giralt, F. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 85.
(30) Cabani, S.; Gianni, P.; Mollica, V.; Lepori, L. *J. Solution Chem.* **1981**, *10*, 563.
(31) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.
(32) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.
(33) Platts, J. A.; Butina, D.; Abraham, M. H.; Hersey, A. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 835.
(34) Hickey, J. P.; Passino-Reader, D. R. *Environ. Sci. Technol.* **1991**, *25*, 1753.
(35) Murray, J. S.; Politzer, P. *J. Chem. Res., Synop.* **1992** (3), 110.
(36) Ben-Naim, A. *Solvation Thermodynamics*; Plenum Press: New York, 1987; p 4.
(37) McQuarrie, D. A. *Statistical Mechanics*; University Science Books: Sausalito, CA, 2000.
(38) de Boer, J. H. *The Dynamical Character of Adsorption*; Clarendon Press: Oxford, U.K., 1968; p 112.
(39) Kemball, C.; Rideal, E. K. *Proc. R. Soc. A* **1946**, *187*, 53.
(40) Hawkins, G. D.; Lynch, B. J.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G. *OMNISOL*, version 1.1; University of Minnesota: Minneapolis, MN, 2003.
(41) http://comp.chem.umn.edu/omnisol.
(42) AMSOL, version 7.0, by Hawkins, G. D.; Giesen, D. J.; Lynch, G. C.; Chambers, C. C.; Rossi, I.; Storer, J. W.; Li, J.; Thompson, J. D.; Winger, P.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G.

University of Minnesota, Minneapolis, MN 2003, based in part on AMPAC, version 2.1, by Liotard, D. A.; Healy, E. F.; Ruiz, J. M.; Dewar, M. J. S.

(43) http://comp.chem.umn.edu/amsol.

(44) Xidos, J. D.; Li, J.; Thompson, J. D.; Hawkins, G. D.; Winget, P. D.; Zhu, T.; Rinaldi, D.; Liotard, D. A.; Cramer, C. J.; Truhlar, D. G.; Frisch, M. J. *MN-GSM*, version 3.2; Univeristy of Minnesota, Minneapolis, MN, 2003.

(45) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision A.1; Gaussian, Inc.: Pittsburgh, PA, 2003.

(46) Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1995**, *16*, 422.

(47) (a) Perdew, J. P. In *Electronic Structure of Solids;* Ziesche, P., Eschrig, H., Eds.; Akademie Verlag: Berlin, 1991; p 11. (b) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais C. *Phys. Rev. B* **1992**, *46*, 6671. (c) Perdew, J. P.; Chevary, J. A.; Vosko, S. H.; Jackson, K. A.; Pederson, M. R.; Singh, D. J.; Fiolhais, C. *Phys. Rev. B* **1993**, *48*, 4978. (d) Perdew, J. P.; Burke, K.; Wang, Y. *Phys. Rev. B* **1996**, *54*, 16533.

(48) Easton, R. E.; Giesen, D. J.; Welch, A.; Cramer, C. J.; Truhlar, D. G., *Theor. Chim. Acta* **1996**, *93*, 281.

(49) Cramer, C. J. *Essentials of Computational Chemistry: Theories and Models*; Wiley & Sons: Chichester, U.K., 2002.

(50) Leo, A. J. *Masterfile* from MedChem Software; BioByte Corp.: Claremont, CA, 1994.

(51) *CRC Handbook of Chemistry and Physics*, 75th ed.; Lide, D. R., Ed.; CRC Press: Boca Raton, FL, 1995.

(52) Gragson, D. E.; Richmond, G. L. *J. Phys. Chem. B* **1998**, *102*, 3847.

(53) Du, Q.; Superfine, R.; Freysz, E.; Shen, Y. R. *Phys. Rev. Lett.* **1993**, *70*, 2313.

(54) Du, Q.; Freysz, E.; Shen, Y. R. *Nature* **1994**, *264*, 826.

(55) Kuo, I.-F. W.; Mundy, C. J. *Science* **2004**, *303*, 658.

(56) Winget, P.; Cramer, C. J.; Truhlar, D. G. *Environ. Sci. Technol.* **2000**, *34*, 4733.

(57) 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.; IMA Volume in Mathematics and Its Applications 108; Springer: New York, 1999; pp 51−72.

(58) Levine, I. N. *Physical Chemistry*, 4th ed.; McGraw-Hill Inc.: New York, 1995; p 352.

(59) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Chem. Phys.* **2003**, *119*, 1661.

(60) Simcik, M. F. *Atmos. Environ.* **2004**, *38*, 491.

(61) Glotfelty, D. E.; Majewski, M. S.; Seiber, J. N. *Environ. Sci. Technol.* **1990**, *24*, 353.

(62) Schomburg, C. J.; Glotfelty, D. E.; Seiber, J. N. *Environ. Sci. Technol.* **1990**, *25*, 155.

(63) Capel, P. S.; Gunde, R.; Zürger, F.; Giger, W. *Environ. Sci. Technol.* **1990**, *24*, 722.

(64) Capel, P. D.; Leunberger, C.; Giger, W. *Atmos. Environ.* **1991**, *25*, 1335.

(65) Sagebiel, J. C.; Seiber, J. N. *Environ. Tox. Chem.* **1993**, *12*, 813.

(66) Wilson, P. J.; Bradley, T. J.; Tozer, D. J. *J. Chem. Phys.* **2001**, *115*, 9233.

(67) Gautier, C.; Le Calvé, S.; Mirabel, P. *Atmos. Environ.* **2003**, *37*, 2347.

(68) Staudinger, J.; Roberts, P. V. *Chemosphere* **2001**, *44*, 561.

(69) Rice, C. P.; Chernyak, S. M.; McConnell L. L. *J. Agric. Food Chem.* **1997**, *45*, 2291.

(70) Mackay, D.; Shiu, W.-Y.; Ma, K.-C. *Illustrated Handbook of Physical-Chemical Properties and Environmental Fate for Organic Chemicals*; Lewis Publishers: New York, 1997; Vol. 5.

(71) Gish, T. J.; Sadeghi, A.; Weinhold, B. J. *Chemosphere* **1995**, *31*, 2971.

(72) Montgomery, J. H. *Agrochemicals Desk Reference: Environmental Data;* Lewis Publishers: New York, 1993.

(73) Chesters, G.; Simsiman, G. V.; Levy, J.; Alhajjar, B. J.; Fathulla, R. N.; Harkin, J. M. *Rev. Environ. Contam. Toxicol.* **1989**, *110*, 1.

(74) Fendinger, N. J.; Glotfelty, D. E.; Freeman, P. *Environ. Sci. Technol.* **1989**, *23*, 1528.

(75) Fendinger, N. J.; Glotfelty, D. E. *Environ. Sci. Technol.* **1988**, *22*, 1289.

(76) Suntio, L. R.; Shiu, W.-Y.; Mackay, D.; Sieber, J. N.; Glotfelty, D. *Rev. Environ. Contam. Toxicol.* **1987**, *103*, 1.

(77) Glotfelty, D. E.; Sieber, J. N.; Liljedahl, L. A. *Nature* **1987**, *325*, 602.

(78) Syracuse Research Corporation. *Physical/Chemical Property Database (PHYSPROP)*; SRC Environmental Science Center: Syracuse, NY, 1994. (online demo version available at http://esc.syrres.com/interkow/physdemo.htm).

(79) Thompson, J. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **2004**, *108*, 6532.

(80) Abraham, M. H.; Chadha, H. S.; Whiting, G. S.; Mitchell, R. C. *J. Pharm. Sci.* **1994**, *83*, 1085.

(81) Worthing, C. R., Walker, S. B., Eds. *The Pesticide Manual—A World Compendium*, 8th ed.; The British Crop Protection Council (The Lavenham Press Limited): Lavenham, Suffolk, U.K., 1987.

(82) Mackay, D.; Bobra, A.; Chan, D. W.; Shiu, W.-Y. *Environ. Sci. Technol.* **1982**, *16*, 645.