Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena

Andreas Klamt

Bayer AG, MD-IM-FA Computational Chemistry, Building Q18, D-51368 Leverkusen, Germany

Received: July 6, 1994; In Final Form: November 3, 1994[⊗]

Starting from the question of why dielectric continuum models give a fairly good description of molecules in water and some other solvents, a totally new approach for the calculation of solvation phenomena is presented. It is based on the perfect, i.e., conductor-like, screening of the solute molecule and a quantitative calculation of the deviations from ideality appearing in real solvents. Thus, a new point of view to solvation phenomena is presented, which provides an alternative access to many questions of scientific and technical importance. The whole theory is based on the results of molecular orbital continuum solvation models. A few representative solvents are considered, and the use of the theory is demonstrated by the calculation of vapor pressures, surface tensions, and octanol/water partition coefficients.

1. Introduction

The computational description of molecules in vacuum or in the gas phase has reach a remarkable state over the past four decades which now allows to calculate and predict most of the interesting features of isolated molecules or small clusters by more or less sophisticated molecular orbital (MO) calculations. But since nearly all technical chemistry as well as biochemistry does take place in the condensed phase, mainly in fluids, the proper theoretical and computational handling of molecules in solution and the proper calculation and prediction of the behavior and the thermodynamic properties of solute−solvent systems is one of the most important actual challenges for computational chemistry.

Due to the enormous technical importance of the knowledge of the thermodynamic behavior of fluid systems and due to the lack of more rigorous methods, powerful phenomenological approaches for the estimation of these data from the molecular structure have been developed over the past century, mostly by chemical engineers. These methods do heavily take advantage of the fact that the free energy of molecules in solution to a considerable degree is additive, i.e., it can be composed of incremental contributions from molecular fragments. As typical representatives of such methods, the Lyderson[1] and Benson[2] incremental methods, the UNIFAC method,[3,4] and the CLOGP method[5] shall be referenced here. Despite of the undoubted merits of these incremental methods, they have some inherent disadvantages: (i) They need large sets of experimental data for the derivation of a suitable fragmentation and the derivation of the corresponding increments, and they are restricted to molecules being composed of the parametrized fragments. (ii) In general, they are not able to describe nonadditive molecular effects arising from intramolecular electronic interactions of fragments, especially in conjugated or aromatic structures or from intramolecular hydrogen bonds. (iii) They only give a fairly good interpolation of the experimentally derived data, but they do not give real insight and thus a better understanding of the physics of the solute−solvent systems, which in many cases would be desirable for a more systematic access to technical problems. These inherent problems of the incremental methods make a more fundamental approach to solvation phenomena on molecular scale highly desirable, and the enormous evolution of computer power allows this goal to appear to become realistic now.

The proper description of a solute molecule in a solvent on a molecular scale is a very hard task, because the solute is interacting with several surrounding solvent molecules, and these again interact with others and so on. Periodic boundary conditions, which are crucial for the success of computational methods for crystalline systems, are missing in solution. Thus, a large ensemble of molecules, typically at least several hundreds of molecules, i.e., several thousands of atoms, has to be considered for a fairly good representation of the system. Even for these ensembles artificial periodic boundary conditions have to be assumed to avoid surface effects which otherwise would be of overwhelming influence in such, still quite small, ensembles. In addition, it is of no use to calculate the ground state of such an ensemble by energy minimization, because it is the thermodynamic average of conformationally excited states which determines the behavior of such disordered systems. Since molecular orbital calculations unavoidably scale with at minimum the third power of the number of atoms considered, now and in the foreseeable future there is no chance to handle real solvation phenomena without additional approximations.

The most widely used approximation is the use of force fields for the description of intra- and intermolecular interactions and to do the thermodynamic averaging by a dynamical (molecular dynamics, MD) or random (Monte Carlo, MC) sampling of the relevant states. Such calculations still are very time-consuming even on the fastest computers, but they may benefit considerably from the present trend to parallel computing. Nevertheless, the results of such calculations are restricted by the fundamental approximation, i.e., the description of the interactions by classical force fields. While intramolecular forces and also dispersive, i.e., van der Waals (vdW), interactions between molecules appear to be quite well parametrized in many force fields, the proper treatment of electrostatic interactions is an outstanding problem. At present they typically are handled as Coulomb interactions of frozen atomic charges. Despite the question of which charges should be used for the best description of reality, such a treatment does neglect polarization effects as well as directed forces arising from atomic dipole or quadrupole moments, e.g., from lone pairs. Since for most polar or polarizable molecules the majority of the solvation effects arise

[⊗] Abstract published in *Advance ACS Abstracts,* January 15, 1995.

0022-3654/95/2099-2224$09.00/0

from electrostatic screening of the molecular polarities, this is a crucial limitation for the application of MD or MC methods for the description of solvation phenomena. A review of this class of methods has been given by van Gunsteren.[6] Mixed molecular orbital (MO) force field methods have been developed[7] to overcome the problems with electrostatics, but even they are based on very simplified approximations regarding the electrostatics of the solvent. Quantum MC/MD methods are evolving[8] but are far from being capable of simulating real solvation problems.

Based on the early work of Born,[9] Kirkwood,[10] and Onsager,[11] continuum solvation models have been proposed, starting in the 1970s and being heavily developed during the past decade. These primarily focus on the solute molecule or a small cluster of the solute with a few solvent molecules and try to represent the influence of the rest of the solvent by an effective continuum surrounding them. Because dispersive interaction energies quite accurately can be expressed by a term proportional to the exposed surface area, and due to the outstanding importance of electrostatic interactions in solvation phenomena, the main task of these models is to give a proper description of the electrostatic screening within this continuum approximation. The use of the dielectric theory of macroscopic media is widely accepted in this approach. Different classes of these dielectric continuum solvation models, which often are called self-consistent reaction field (SCRF) models, have been developed, and a good review of them has recently been given by Cramer and Truhlar.[12] In all of these models a more or less vdW-like surface is constructed around the solute. This defines the boundary to the dielectric continuum. For a given electrostatic field of the solute, typically derived from a single molecule MO calculation, the dielectric medium screens this field by a polarization of the medium, which may as well be represented by well-defined surface charges on the boundary surface. Once these are calculated, the MO calculation can be repeated taking into account the field of these screening charges, and thus the cycle can be repeated until self-consistency is achieved. Since usually only the solute molecule has to be treated by quantum chemistry, there is no principal limit for the accuracy of the calculation of the electrostatic interaction energy within the dielectric continuum, including polarization effects, although consideration of correlation effects may bear some major complications.[13] Nevertheless, even semiempirical MO calculations appear to give very satisfying results in such calculations. Because the calculation of the dielectric screening charges on an arbitrarily shaped surface is by far not trivial and causes considerable numerical expense, simplified surface geometries, i.e., spheres or ellipsoids, are frequently used,[13,14] and Onsager's multipole expansion[11] for the energy of a source in such a dielectric cavity is applied. But although this approximation may be tolerable for some molecules, in general the use of more realistic surfaces is highly recommended, since the shape and size of the cavity are of crucial influence to the results. Despite the fact that the optimization of the parameters for the construction of such a real cavity is still outstanding, vdW-like surfaces are proven to give reliable results. Throughout this article the COSMO (conductor-like screening model) approach[15] will be used.[16] COSMO has been developed recently and is available in MOPAC7 and MOPAC93.[17,18] COSMO calculates the dielectric screening charges and energies on a vdW-like molecular surface in the approximation of a conductor, which is highly accurate and much more efficient compared to the solution of the dielectric boundary conditions. Thus, COSMO even allows for geometry optimization of the solute within the continuum. Nevertheless, the PISA-like implementations[19,20] could also have been used, probably without major differences.

Up to now dielectric continuum models are most widely applied to hydration phenomena, and they are proven to give surprisingly good results for hydration energies as well as for tautomeric and dissociation equilibria.[21] Although several successful applications of continuum solvation models to other solvents than water have been reported,[14,22] none of these solvents are considered nearly as intensive as water, and no general answer can be given yet as to whether the dielectric continuum approach also does work in general for other solvents.

The aim of this paper is to get a better understanding of the reason why dielectric continuum models do give such a good description of the solvent water. A surprising answer will be given to this question in section 2, which opens a totally new and instructive view to solvation phenomena. It provides access to a quite general quantitative theory for the calculation of a great variety of thermodynamic parameters of pure and mixed fluid systems, described in section 3. Section 4 presents a detailed consideration of a few representative solvents, and in section 5 the use of the theory is demonstrated by the three exemplary, but preliminary, applications, i.e., the calculation of vapor pressures, octanol/water partition coefficients, and surface tensions of organic compounds. In section 6 a survey of outstanding improvements is given, and in section 7 the possible fields of applications are discussed. Section 8 gives a final summary.

2. Why Do Continuum Solvation Models Work?

The theory of dielectric screening[23] has been developed for the description of the electrodynamic behavior of homogeneously polarizable macroscopic media. It describes the time-averaged, linear, first-order response of these media with respect to external electric fields, and thus it is a theory for weak electric fields. But the fields on the vdW surface of polar molecules and even more those of ions are by orders of magnitude stronger than the strongest technical electric fields. Furthermore, no solvent is a homogeneously polarizable medium on a molecular length scale. Thus, it would be very surprising if the macroscopic dielectric constant ϵ of a solvent could give a correct qualitative or even quantitative description of the ability of the solvent to screen the strong and highly inhomogeneous electric fields of solute molecules. As mentioned above, we do not know whether solvents in general behave like dielectrics on a molecular scale, and thus we should ask only why water appears to behave like a strong dielectric or a conductor with respect to the screening of polar molecules. But since water on a molecular scale resembles neither a homogeneously polarizable medium nor a conductor with freely movable charges, both pictures do not give a satisfying answer. It still appears strange that water is able to screen molecular polarities nearly as well as electric conductors, i.e., to screen them almost perfectly.

A surprising answer to our question can be derived from the following virtual experiment: Imagine a set of hypothetical cubic solvent molecules S and a cubic solute X which have been put into a conductor (see Figure 1a). Thus, they are perfectly screened by surface charges, and they gained their ideal screening energies Δ^S and Δ^X, respectively. Let us assume for simplicity that the screening charge densities are fairly constant on each face of the cubes. Due to the perfect screening, there is no interaction of the molecules. Thus, we can move and arrange them as we like without any change in energy. Let us assume that the solvent molecules S have a face with surface charge density $-\sigma_i$ for each surface charge density σ_i of the solute, and let us arrange the solvent molecules around the solute

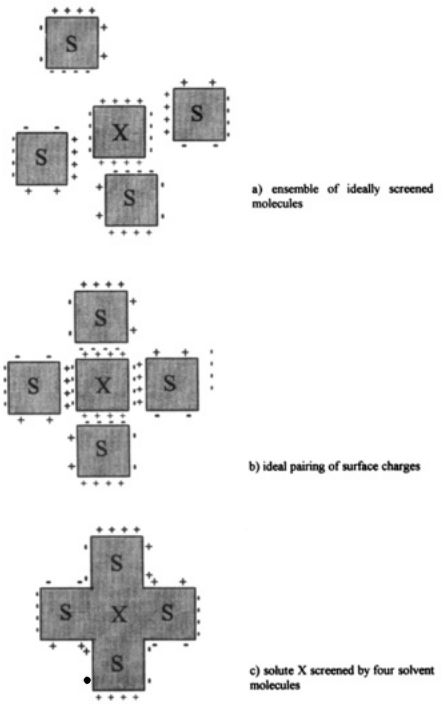

Figure 1. Virtual screening experiment.

X in a way that the net charge on adjacent surfaces is zero (see Figure 1b). Then it does not make any difference whether there is a conducting surface between the molecules or not, and we can remove these without any change in energy (see Figure 1c). Now the solute X is screened by the solvent molecules, but it still has the same energy as if it would be screened by a conductor. Thus, we may conclude the following: *If a solvent can offer the opposite ideal surface charge densities for all faces of a solute molecule, then it is able to screen the solute as good as a conductor.*

Water appears to have this capability for a great variety of molecules, and we assume that this is the reason for the conductor-like screening of most solutes in water. Other solvents can offer a less broad spectrum of surface charge density resulting in the need to undergo nonideal pairing with surface segments of many solute molecules as well as with those of their own species. This results in a less perfect screening in those solvents. Thus, we are led to a totally new view on fluids and fluid solutions. Instead of considering them as ensembles of molecules which are interacting via static and induced electric fields as well as via vdW interactions, we may look at them as ensembles of ideally screened molecules contacting on the vdW surface with pairwise interactions of adjacent surface charge densities and pairwise dispersive interactions through the contact areas. Because now all interactions are local and pairwise, the computational handling of such a system is much less complicated than in the traditional description. The only ingredients to the new description are the ideal, i.e., conductor-like, screening charge densities on the surfaces of the molecules. These can easily be derived from single molecule MO calculations with a continuum model like COSMO, which may be performed for each different compound of the system with small computational effort. Thus, this new approach to the description of solvation phenomena , which we shall call the "conductor-like screening model for real solvents" or COSMO-RS, promises to have great advantages compared to the traditional solvent models.

In addition, the above virtual experiment does point out the contact surface of molecules, i.e., a vdW-like surface, as the proper "screening" surface. Thus, it explains why it is just the vdW-surface which with only slight modifications has turned out to be the optimal cavity in nearly all continuum solvation models. Up to now this fact has been accepted as purely empirical, since the dielectric approach does not provide an answer to the question at which distance a solvent acts like a dielectric medium.

3. Construction of a Quantitative Theory

In this section we derive a quantitative theory for the treatment of solvation phenomena. Quantities related to a single molecule, mostly a solute, are marked by an upper capital index and those related to a solvent by a lower capital index.

3.1. Analysis of the Screening Charge Densities. The starting point of the COSMO-RS theory is the ideally screened molecules. As a result of a COSMO/MO calculation for a molecule X, we get the ideal screening energy Δ^X, i.e., the difference between its total energy in vacuum and in a conducting continuum, as well as the screening surface charge densities on a vdW-like surface around the molecule.

Since real molecules are not cubic or polyhedral as assumed in our virtual experiment, they have no natural faces. To transfer the basic idea of the virtual experiment to real molecules, we thus have to introduce the concept of typical contact segments of molecules. The size of these segments should be large enough to ensure that the probability of neighboring segments to have the same neighbor molecule is far less than unity. This is a necessary condition to make the concept of individual pairing of the segments meaningful. Furthermore, the mean surface charge densities on neighboring segments should be reasonably uncorrelated, while the surface charge density should be rather homogeneous within a segment, in order to allow to represent it by its mean value. It is by no means a priori obvious that such a proper size for typical contact segments can be defined. Nevertheless, a segment size of about 3 Å^2, corresponding to an effective segment radius of $R_{\text{eff}} = 1$ Å, has turned out to work very well under all of these different aspects, which in part are considered in more detail below, and we will adopt this value throughout this article. We take $n^X = A^X/A_{\text{eff}}$, with $A_{\text{eff}} = \pi R_{\text{eff}}^2$ being the effective segment area, for the number of segments of a molecule X. For water, for which the COSMO area A^X, i.e., the area of the screening surface in the COSMO algorithm, is about 25 Å^2, this setting leads to $n^X \simeq 8$, corresponding quite well with the number of nearest-neighbor molecules in typical water models.

Averaging these surface charge densities over such typical contact segments, we can derive an effective probability function $p^X(\sigma)$, which gives the probability of finding a mean screening charge density σ on a typical contact segment of the molecule X. For the sake of brevity we hereafter call this probability function the "σ-profile" of molecule X.

The σ-profiles are not only characteristic of single molecules but also of an ensemble of solvent molecules. Obviously, for a pure solvent $p_S(\sigma)$ is identical with the σ-profile $p^S(\sigma)$ of a single solvent molecule. For solvents composed of different constituents X_i with molar concentration x_i the σ-profile is given by weighted sums of the σ-profiles of the components:

$$p_S(\sigma) = \sum_i x_i n^{x_i} p^{x_i}(\sigma) / \sum_i x_i n^{x_i} \qquad (1)$$

As may be seen from Figure 2 and as discussed in section 4 the σ-profiles $p^X(\sigma)$ vary considerably between different typical organic solvent molecules.

3.2. The Interaction Energy of Surface Charges. As shown in section 2, the interaction energy of two segments with

opposite surface charge densities is zero, if we take the ideally screened molecules as reference. In a real fluid such ideal pairing will not be possible for all segments, and it is also not favorable under entropic conditions. Thus, a number of more or less nonideal pairings will be realized. For a quantitative theory it is crucial to quantify the energy of such nonideal pairs of segments, i.e., the interaction energy of two adjacent segments of radius R_{eff} with surface charges σ_1 and σ_2. Together these two closely adjacent segments represent a surface with total charge $\sigma_1 + \sigma_2$. As has been shown in ref 15, the self-energy of such a homogeneously charged surface segment in good approximation is

$$E_{misfit}(\sigma_1,\sigma_2) = {}^1/_2\alpha(\sigma_1 + \sigma_2)^2 \quad \text{with } \alpha = 1.2\pi^{5/2}R_{eff}{}^3/(4\pi\epsilon_0) \quad (2)$$

This is the interaction energy arising from the misfit of two adjacent surface charge densities. Obviously there will also be interactions of residual charges on neighboring segment pairs. If the charge densities on neighboring segments are not correlated, which we assume to be achieved by a proper choice of the segment radius, the sum of these interactions will vanish due to the fluctuating sign of the misfit charges. Therefore, we neglect these interactions further on. Now the only relevant energies in the system are the ideal screening energies Δ^X, the pairwise misfit energies, and contributions from the dispersive interaction of each molecule with the remainder, which we assume to be reasonably represented by a surface proportional term γA^X throughout this paper. It should be noted that this term does only include dispersion and not cavitation, since the latter results from electrostatic interactions, which are considered in the misfit term.

A necessary condition for the consistency of the model can be derived by consideration of a molecule X with ideal screening energy Δ^X in a nonpolar and nonpolarizable solvent. All solvent molecules will thus only have neutral surface segments. Hence, the total misfit energy of the solute is

$$E^x_{mf} = {}^1/_2\alpha\sum_\nu \sigma_\nu{}^2 = {}^1/_2\alpha N^x \int p^x(\sigma)\sigma^2\, d\sigma \quad (3)$$

Since X now is absolutely unscreened, this misfit energy has to be just the negative of the ideal screening energy Δ^X. Testing this condition for several molecules, i.e., the list of molecules given in Table 1, shows that it is very well met, if a value of $R_{eff} \approx 1$ Å is chosen.

3.3. Polarizability Correction. So far we have neglected that in the case of a misfit charge on a contact surface segment between two molecules both molecules will adjust their wave functions and thus reduce the misfit energy. In other words, both molecules will lower the misfit energy by their electronic polarizability. Fortunately, the electronic polarizability approximately is a constant for most organic solvents. It corresponds to a dielectric constant of $\epsilon_{pol} = n^2 = 1.8-2.2$ where n is the refractive index. Using the ϵ dependence $f(\epsilon) = (\epsilon - 1)/(\epsilon + {}^1/_2)$ discussed in ref 15 for the energy reduction in a dielectric medium, the misfit energies are reduced by a factor $f_{pol} \approx 0.6$ due to the polarization. Since for several important solvents ϵ_{pol} is slightly less than 2, we work with a polarizability factor of $f_{pol} = 0.64$ throughout this article. For shortness we introduce an effective self-energy constant $\alpha' = f_{pol}\alpha$ for the use in misfit energy calculations. It should be noted that ideal solvation energies are not affected by that correction.

Whether $f_{pol} = 0.64$ is the optimal choice and whether molecular and directional variations of the polarizability are of importance for the results have to be decided later. At present we assume them to be negligible in order to keep the theory as simple as possible.

TABLE 1: COSMO−RS Descriptor Values for the Test Set Compounds

	Δ^X [a]	A^X [b]	E^{GS}_X [a]	μ'^X_X [a]	$\Delta\mu^X_{OW}$ [a]	μ^{vac}_X [c]
methane	−0.2	30.1	0.0	0.2	2.4	0.004
ethane	−0.3	38.8	0.0	0.3	3.1	0.004
n-pentane	−0.3	61.8	0.0	0.4	4.9	0.004
n-octane	−0.4	84.1	0.1	0.6	6.7	0.004
neopentane	−0.3	57.8	0.0	0.4	4.6	0.004
cyclobutane	−0.7	49.6	0.1	0.8	3.6	0.008
cyclopentane	−0.6	55.3	0.1	0.8	4.1	0.007
cyclohexane	−0.3	59.7	0.0	0.3	4.8	0.003
ethene	−1.1	37.3	0.1	0.9	2.4	0.017
ethine	−2.8	36.0	0.2	1.7	1.4	0.042
propene	−1.4	45.8	0.2	1.2	2.9	0.017
propine	−2.9	44.4	0.2	1.9	2.0	0.037
propadione	−2.2	44.5	0.1	1.7	2.1	0.033
propanone	−9.2	49.4	2.8	5.3	−0.0	0.045
butanone	−8.5	56.6	2.5	5.3	1.0	0.037
2-pentanone	−8.3	64.2	2.3	5.4	1.6	0.033
3-pentanone	−7.7	63.8	2.4	5.4	1.9	0.029
3-methylbutanone	−7.9	62.6	2.6	5.3	1.7	0.031
2-heptanone	−8.4	79.1	2.2	5.7	2.8	0.029
3-heptanone	−7.6	78.9	2.2	5.5	3.1	0.025
2-octanone	−8.4	86.7	2.2	5.8	3.4	0.027
dimethyl ether	−5.4	44.0	2.7	4.0	1.2	0.027
methyl propyl ether	−5.2	58.8	2.4	4.1	2.6	0.020
dipropyl ether	−4.9	74.1	2.1	4.1	4.1	0.015
diisopropyl ether	−4.4	69.7	2.3	3.9	4.1	0.013
1,4-dioxane	−9.5	56.4	4.0	6.6	0.0	0.039
chloromethane	−3.1	42.8	0.0	1.7	1.8	0.040
dichloromethane	−2.1	74.0	0.1	1.4	5.0	0.013
trichloromethane	−3.8	64.7	0.7	2.8	3.6	0.024
tetrachloromethane	−4.3	54.3	0.5	2.7	2.2	0.039
chloroethane	−3.2	50.3	0.2	1.9	2.4	0.033
hexachloroethane	−3.2	94.7	0.1	2.2	6.1	0.016
tetrachloroethen	−2.6	79.6	0.1	1.8	5.1	0.016
fluoromethane	−3.5	33.7	0.6	1.8	0.9	0.053
bromomethane	−3.3	48.2	0.0	1.8	2.2	0.036
dibromoethane	−5.0	64.0	0.3	3.0	2.6	0.039
trifluorochloromethane	−1.9	52.2	0.1	1.9	2.8	0.022
difluorodichloromethane	−1.4	60.5	0.0	1.4	3.8	0.015
fluorotrichloromethane	−1.6	67.6	0.0	1.2	4.6	0.012
1,2-dichlorotetrafluoroethane	−1.4	71.9	0.1	1.4	4.7	0.013
methanol	−6.9	35.0	1.0	3.3	0.2	0.050
ethanol	−6.6	43.5	0.9	3.5	1.0	0.036
1-butanol	−6.5	58.7	0.8	3.7	2.3	0.028
2-butanol	−5.9	57.1	1.0	3.6	2.5	0.023
1-pentanol	−6.5	66.1	0.7	3.8	2.9	0.025
1-heptanol	−6.6	81.2	0.9	4.1	4.1	0.021
ethandiol	−12.0	47.9	1.0	5.1	−1.1	0.081
phenol	−7.9	62.8	0.1	4.2	1.3	0.059
p-cresol	−8.0	69.5	0.2	4.5	1.8	0.050
benzyl alcohol	−8.6	69.8	0.7	4.9	1.4	0.053
benzene	−3.5	59.0	0.3	2.5	2.5	0.037
toluene	−3.6	66.0	0.5	2.8	3.0	0.032
4-xylene	−3.7	72.7	0.8	3.1	3.5	0.028
styrene	−4.3	71.8	0.3	3.2	2.9	0.039
2-picoline	−7.5	65.1	1.8	5.0	1.8	0.038
chinoline	−8.8	77.4	1.3	5.5	1.7	0.050
aniline	−9.1	64.0	0.4	4.7	0.8	0.064
dichlorobenzene	−4.1	80.4	0.2	2.8	4.0	0.030
hexachlorobenzene	−3.7	113.1	0.1	2.8	7.1	0.018
1-chloronaphthalene	−5.3	86.7	0.0	3.7	3.6	0.040
fluorobenzene	−4.0	61.8	0.0	2.5	2.6	0.041
bromobenzene	−4.6	74.4	0.0	2.9	3.3	0.037
water	−9.2	24.7	0.1	2.5	−1.7	0.145
methylamine	−6.5	37.1	1.3	3.5	0.7	0.037

[a] In kcal/mol. [b] In Å^2. [c] In kcal/(mol Å^2).

3.4. Decoupling of the Segments. In a real fluid the arrangement of surface segments is constrained by the geometry of the involved molecules. Because the handling of these

constraints, i.e., the calculation of sterically allowed realizations of the considered ensemble of molecules, is quite complicated, we shall neglect these constraints in this article in order to achieve an extreme simplification of the problem. We are aware that this decoupling of the segments is a very severe approximation of a priori unknown accuracy,[24] probably bearing a tendency toward oversolvation, since it allows for segment combinations, which might be forbidden otherwise. Nevertheless, we decided to evaluate how far we can get within this approximation, and the encouraging results show that this approximation appears to be much less critical than may be expected.

Thus, we now consider a liquid S as an ensemble of N decoupled segments, each carrying a surface charge σ. The relative frequencies of the charge densities σ is given by the σ-profile $p_S(\sigma)$, which can be that of a pure liquid or a mixture (see eq 1). Neglecting surface effects, we postulate that any segment has to be associated with another one so that $N/2$ pairs of segments are formed. Obviously, for each of these pairs there has to be a physical site in the considered amount of the liquid. Thus, there are $N/2$ pair sites in the liquid, and for a fixed pairing of segments there exist $(N/2)!$ arrangements on these sites. Taking into account that each pair has two possible orientations on a given site, the multiplicity of each pairing is $2^{N/2}(N/2)!$.

The total solvation energy of a given pairing of segments is the sum of all ideal solvation energies $E_{is}^X = \Delta^X + \gamma A^X$ of the molecules in the ensemble, plus the sum of the misfit energies $E_{misfit}(\sigma_i,\sigma_j)$ of all pairs of segments. Since the ideal solvation energy is independent of the the concrete realization of the ensemble and, moreover, is easy to calculate, we only consider the sum of misfit energies in more detail now.

3.5. Ground State of the Decoupled Ensemble. The ground state (GS) of an ensemble of decoupled surface segments, i.e., the optimal pairing of the segments with respect to the misfit energy, can be easily derived from the σ-profile $p_S(\sigma)$ by the following procedure. Order all segments by their charge density σ, and number them from $-N/2$ to $N/2$, beginning at the minimum of σ and leaving out the zero. Then the optimal pairing is given by all pairs $(-i,i)$, and hence the ground state misfit energy is

$$E_S^{GS} = {}^1/_2\alpha' \sum_{i=1}^{N/2} (\sigma_i + \sigma_{-i})^2 \qquad (4)$$

Obviously for any symmetric distribution σ_i is equal to σ_{-i}, and thus there is no misfit energy in the ground state. But in most cases the distribution will not be symmetric, and we find a nonvanishing ground state misfit energy. Although for relevant temperatures in real fluids this extremely ordered ground state is of negligible thermodynamical importance, the ground state misfit energy is a good indicator for the ability of a fluid to screen its own molecules. There are considerable and interesting differences in the ground state misfit energies of common solvents (see examples in section 4).

3.6. Partition Function, Free Energy, and Chemical Potential. Of real interest is the behavior of liquids at elevated temperatures, i.e., from ca. 200 to 500 K. Here misfit energies of ${}^1/_2\alpha'\sigma^2 \approx {}^1/_2kT = 0.25-0.5$ kcal/mol corresponding to σ values of about 0.015 e/Å^2 become thermodynamically tolerable so that compared to the ground state considerable disorder, but for most solvents by far not a total disorder, will occur. Hence, the thermodynamic behavior of an ensemble S of N segments has to be derived from its partition function Z_S, which is related to the free energy F_S by $F_S = -kT \ln Z_S$. As mentioned above, we only consider the misfit energies further on, i.e., take the ideally solvated state, where each molecule has gained its ideal solvation energy $E_{is}^X = \Delta^X + \gamma A^X$, as reference. Since in the intermediate temperature region the analytical treatment of the partition function turns out to be quite complicated, and because the statistical evaluation by Monte Carlo simulations is possible, but not very instructive, we have derived a self-consistency equation for the chemical potential $\mu_S(\sigma) = dF_S/dN(\sigma)$ of segments with charge density σ in the ensemble S:

$$\mu_S(\sigma) = kT \ln[N^2 \int d\sigma'\, p_S(\sigma') \times \exp\{({}^1/_2\alpha'(\sigma+\sigma')^2 \mu_S(\sigma'))/kT\}] \qquad (5)$$

For the derivation of this equation see Appendix A. Introducing the chemical potential $\mu_S'(\sigma)$ of a normalized ensemble of 1 mol of segments, called the σ-potential further on, eq 5 becomes

$$\mu_S'(\sigma) = -kT \ln [\int d\sigma'\, p_S(\sigma') \times \exp\{-({}^1/_2\alpha'(\sigma+\sigma')^2 - \mu_S'(\sigma'))/kT\}] \qquad (6)$$

together with $\mu_S(\sigma) = -kT \ln n + \mu_S'(\sigma)$, where $n = N/N_{mole}$ is the number of moles of segments in the ensemble. Only for Gaussian σ-profiles can eq 6 can be solved analytically. Since this is quite instructive, we present this example in Appendix B. For these cases $\mu_S'(\sigma)$ comes out to be a parabola $a + b\sigma^2$. In general, the self-consistency equation has to be solved by numerical iteration. Starting from $\mu_S'(\sigma) \equiv 0$ the iteration turns out to converge rapidly, if after each iteration step the average of the old and the new chemical potentials is used for the next step in order to avoid numerical oscillations. This iteration only takes seconds on a modern computer. For the analytical solvable case of the Gaussian σ-profiles as well as by comparison with the results of Monte Carlo simulations on different representative systems,[25] we have checked that the resulting σ-potential always is the correct solution.

The σ-potential $\mu_S'(\sigma)$ is the key to all interesting thermodynamical properties of the liquid S. The chemical potential of any molecule X in this liquid now can be calculated as

$$\mu_S^X = \Delta^X - \gamma A^X + n^X \int d\sigma\, p^X(\sigma)\, \mu_S'(\sigma) - n^X kT \ln(N) \qquad (7)$$

The two leading terms are due to the ideal solvation of the molecule, the third term is due to the misfit interactions of X in the solvent, and the last term is due to the possibilities to place the n^X segments of the solute X within the N segments of the solvent ensemble. This term has been artificially enlarged by the decoupling of the segments. In reality, only the entire molecule and not all n^X segments separately can be placed in the solvent. Therefore, the factor n^X should be replaced by 1 to correct this error. Then we have our final result

$$\mu_S^X = \Delta^X - \gamma A^X + n^X \int d\sigma\, p^X(\sigma)\, \mu_S'(\sigma) - kT \ln(N) \qquad (8)$$

It should be noted that eq 8 gives the chemical potential of any molecule X in pure as well as in mixed fluids. Thus, it opens the way to calculate all the thermodynamical equilibrium constants like Henry constants, activity coefficients, partition coefficients, and also phase diagrams of fluid systems. Furthermore, we can calculate the chemical potential of an uncharged and unpolarizable segment, i.e., a segment of an interface to the vacuum (or a gas) as

$$\mu_S^{vac} = -kT \ln[\int d\sigma\, p_S(\sigma) \exp\{-({}^1/_2\alpha\sigma^2 - \mu_S'(\sigma))/kT\}] \qquad (9)$$

Note that according to the missing polarizability of the vacuum the constant in the misfit energy is α and not α'. Thus $\tau_S^{vac} =$

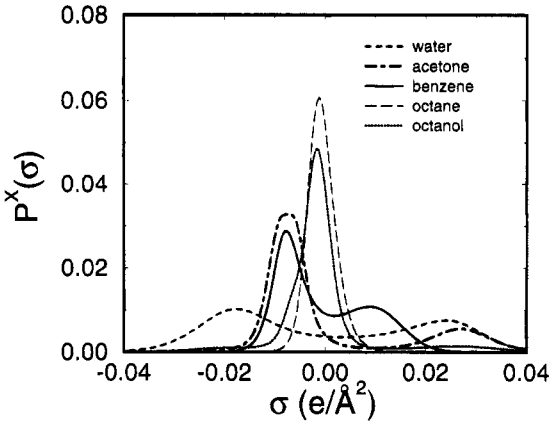

Figure 2. σ-Profiles of representative solvents.

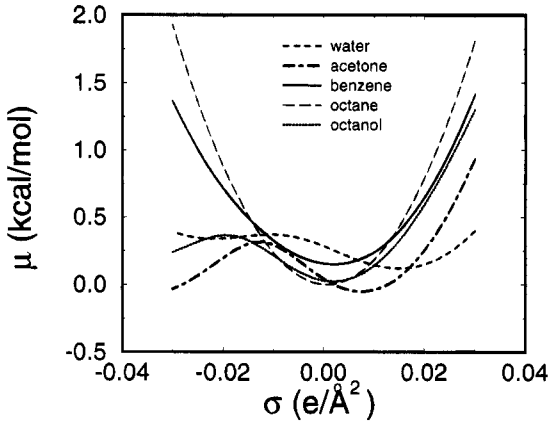

Figure 3. σ-Profiles of representative solvents.

μ_S^{vac}/A_{eff} is a measure for the electrostatic contribution to the surface tension of the liquid S under consideration, and we will prove its use below. In analogy the electrostatic surface tension of an interface between two liquids S and S′ can be calculated as

$$\tau_S^{vac} = -\frac{kT}{A_{eff}} \ln[\int d\sigma\, p_S(\sigma) \int d\sigma'\, p_{S'}(\sigma') \times \exp\{-(^1/_2\alpha(\sigma+\sigma')^2 - \mu_{S'}'(\sigma) - \mu_S'(\sigma'))/kT\}] \quad (10)$$

4. Some Representative Solvents

In this section we apply the theory developed in section 3 to some representative solvents, in order to demonstrate its potential to give a new qualitative insight into solvation phenomena. The underlying surface charge density profiles $p_S(\sigma)$ of these solvents have been calculated with the COSMO–MOPAC implementation, keeping the radii as well as all COSMO defaults for the cavity construction as they are set in MOPAC93.[16,18] The AM1 Hamiltonian[26] has been used, taking into account the conformation with the lowest energy in the conducting medium. These σ-profiles are shown in Figure 2, while Figure 3 presents the σ-potentials $\mu_S'(\sigma)$ due to the misfit energies.

4.1. Water. The ideal screening energy of a water molecule is −9.3 kcal/mol, and its COSMO surface A^X is 24.7 Å^2. Water has a relatively broad σ-profile ranging from −0.03 to 0.03 e/Å^2 with two maxima at −0.02 and 0.025 e/Å^2, which correspond to the positive hydrogen atoms and the lone pairs of the negative oxygen atom, respectively. The profile is quite symmetric so that in the ground state each segment can find a nearly ideal partner. As a consequence, its ground state misfit energy is only 1% of Δ^X. Due to the broad and symmetric σ-profile, the σ-potential is flat in a broad σ-range, but it has a relatively high level compared to the other solvents at $\sigma = 0$, i.e., for nonpolar segments. The first fact resembles the capability of water to screen very well almost all polarities, i.e., all relevant surface charge densities of organic molecules. Only for charge densities outside the interval ±0.035 e/Å^2 is the limit of the screening capability of water reached, and $\mu'(\sigma)$ starts to increase. Such σ-values do only occur for small ions like NH_4^+ or Cl^-, and thus we find the well-known fact that water is not able to screen small ions as perfectly as larger ones.

The high level of the chemical potential resembles the high cavitation energy, i.e., the high surface tension for interfaces to solutes, which is due to the strong affinity of the water molecules among each other. This is responsible for the hydrophobicity of unpolar molecules like alkanes. Since a constant part of σ-potential $\mu'(\sigma)$ corresponds to a surface-proportional contribution to the chemical potential of solutes (see Appendix C), and since accidently the level of the chemical potential of water is nearly of the same value but opposite sign as the dispersion energy constant γ, the surface dependence of the chemical potential of molecules in water almost vanishes. This reflects the fact that the Henry constants for the solvent water do not vary significantly for homologous series of solutes, which only differ in the length of alkane chains.

4.2. Acetone. Acetone is as polar as water, i.e., it has the same ideal screening energy, but it has a larger surface of 48 Å^2. While the oxygen peak in the σ-profile is nearly at the same position as for water, there are much more, but less charged, segments on the negative side. These correspond to the positive countercharges of the electronegative carbonyl oxygen, which are distributed over the six hydrogens of the two methyl groups instead of being concentrated on two hydrogens as in water. As a consequence, acetone has a highly asymmetric σ-profile. Since adequate partners are missing, the strong positive segments of the oxygens have to arrange with the slightly negative hydrogen segments, resulting in a considerable misfit energy of these pairs. Even more uncomfortable feel the remaining negative segments, since for them, there are no positive partners left and they have to pair with each other. Hence acetone does not like itself very much, and it has a high ground state misfit energy of about one-third of the ideal solvation energy. This is the reason why acetone has a considerably lower boiling point than water, although it has the same polarity and even twice the surface area as water.

The σ-potential $\mu'(\sigma)$ of acetone shows pronounced structure. As is expected from the above consideration, there is a lack of highly negatively charged segments and of weekly positively charged segments in acetone, leading to a low chemical potential in these σ-regions. Conversely, acetone does not like additional highly positive or slightly negative segments, corresponding to high chemical potentials in these σ-regions. For neutral segments the σ-potential also is quite low in accordance with the affinity of acetone to nonpolar solutes. As a consequence of the pronounced structure in the σ-potential, acetone should be very attracted to small positive ions, e.g., protons, while it does like almost no negative ion. Indeed, this behavior of acetone is well known from experiments.

4.3. Benzene. Although benzene has no dipole moment, due to its quadrupole moment, its ideal solvation energy is −3.5 kcal/mol. Its σ-profile looks quite asymmetric due to a high, but narrow peak at −0.01 e/Å^2, which corresponds to the positively charged hydrogens, and a lower but broader peak from the π-electrons of the ring at 0.01 e/Å^2. Nevertheless, since the integral of both peaks is nearly the same there is only 5% misfit in the ground state of benzene. The σ-potential of

benzene does lock quite parabolic with a slightly less curvature near $\sigma = 0$. Therefore, benzene can only screen low charge densities up to ± 0.01 e/Å^2 rather well, and thus it is a good solvent for nonpolar or slightly polar compounds, while it is not capable to screen any stronger polarities efficiently.

4.4. *n*-Octane. *n*-Octane is nearly nonpolar, and its ideal screening energy is -0.4 kcal/mol. As expected, the σ-profile is a narrow and nearly Gaussian peak around $\sigma = 0$. Consequently, the σ-potential is a parabola with curvature $^1/_2\alpha'$ (see Appendix B) corresponding to a medium, which exclusively screens by its polarizability. The σ-profiles of different conformations of *n*-octane do not differ significantly.

4.5. 1-Octanol. 1-Octanol has an ideal solvation energy of -5.6 kcal/mol, which is solely due to the polar hydroxy group. The σ-profile shows a strong central peak from the alkane chain with a slight shoulder at the negative side resulting from the slightly positive segments near the oxygen and broad wings from the oxygen and the polar hydrogen. This single polar hydrogen is not capable of fully compensating the oxygen segments, neither with respect to the value of the charge density nor with respect to the number of polar segments. Hence, its misfit energy is 10% of the ideal screening energy, and like in acetone the σ-potential shows minima and maxima, but far less pronounced.

This exemplary consideration of some typical solvents within the COSMO-RS approach reveals significant differences between them and gives instructive explanations for several of the differences in their solvation behavior. It demonstrates the capability of the COSMO-RS model to throw new light on solvation phenomena.

5. Three Quantitative Applications

In this section we present three quantitative applications of the COSMO-RS model in order to demonstrate its capability to give useful quantitative results. Since yet there are a lot of optimizations left to be done, it must be emphasized that these applications are in a preliminary state, and considerable improvements can be expected in future.

As exemplary applications we took the calculation of vapor pressures, octanol/water partition coefficients, and surface tensions. All three parameters are considered at room temperature. We have chosen a set of 64 compounds for this purpose, which covers a broad range of organic chemistry. The experimental data for vapor pressure and surface tension are taken from the in-house property databank of Bayer.[27] In parts they are interpolated to room temperature. The partition coefficients are from the THOR data bank.[28]

5.1. Vapor Pressures. Apart from a small correction for the density, the logarithmic vapor pressure of a liquid X is proportional to the difference of the chemical potential of a molecule of X in the gas phase and the liquid phase X. For pressures which are much lower than the critical pressure, we can safely use the ideal gas approximation for the chemical potential in the gas phase, which is kT, independent of the compound. Thus, for a fixed temperature the gas phase chemical potential can be absorbed in a regression constant, and only the chemical potential μ_X^X of the compound X in its own liquid phase should be responsible for the variations in the vapor pressure. Within the COSMO-RS approach this chemical potential is given by eq 8 so that we ideally expect an expression of the form

$$\log(p_v) = \{\Delta^X - \gamma A^X + \mu'^X_X\}/(kT \ln(10)) + \text{const} \qquad (11)$$

for the logarithmic vapor pressure. Using the three terms of the chemical potential as independent descriptors, a multilinear regression on the basis of 59 of our test compounds, which have melting points less or at maximum slightly higher than room temperature, yields

$$\log\left(\frac{p_v}{\text{bar}}\right) = \left\{1.06\Delta^X - 0.10\frac{kcal}{\text{mol}\,\text{Å}^2}A^X + 1.24\mu'^X_X - 1.26\frac{\text{kcal}}{\text{mol}}n^X_{\text{ring}}\right\}/\left(1.36\frac{\text{kcal}}{\text{mol}}\right) + 4.72 \qquad (12)$$

with a correlation coefficient of $r^2 = 0.93$ and a standard deviation of $s = 0.41$, corresponding to a factor 2.5 for the vapor pressure itself. Here n^X_{ring} is the number of rings in molecule X. Apart from this ring correction, eq 12 is in quite good accordance with the ideal expectation given in eq 11, since the coefficients of Δ^X and μ'^X_X are close to unity. The coefficient of the surface area A^X has to be taken as an estimate of the dispersion energy constant γ, and it is in good agreement with other estimates for γ. Obviously, the appearance of the ring correction resembles a weakness of the present state of the model. Together with the results for LOGPOW (see below) it appears to be probable that this ring correction is mainly caused by a systematic error for aromatic rings, going along with another reason for the deviations at aliphatic rings. The error on aromatic rings most probably is due to deficiencies in the semiempirical wave function. The standard error of a factor 2.5 surely is far from perfect, but in consideration of the broad range of compounds and the large variation of more than 6 orders of magnitude for the vapor pressure it is very satisfying. On restricted classes of compounds much more precise predictions should be achievable.

Experimental and calculated values of the vapor pressure are given in Table 2 and displayed in Figure 4. It should be noted that even for methane and 1-chloronaphthalene, which have been left out in the regression analysis due to the extremely high vapor pressure and due to the quite far extrapolation below the melting point, respectively, the predictions of eq 12 are in best agreement with the experimental data.

5.2. Octanol/Water Partition Coefficients. The *n*-octanol/water partition coefficient P_{OW} or more often its decadic logarithm LOGPOW is frequently taken as a measure of lipophilicity of organic compounds. There do exist several incremental systems for the calculation of LOGPOW.[5,28] The best of these have a prediction standard error of about 0.3–0.4, corresponding to a factor of 2–2.5, if they are applied to a broad variety of organic compounds.[29]

Apart from a constant for density correction of the two liquids the logarithmic partition coefficient of a molecule X between two liquids S and S′ is given by the difference of the chemical potential of X in these two liquids. Thus, within the COSMO-RS approach the difference $\Delta\mu^X_{OW} = \mu^X_{\text{water}} - \mu^X_{\text{octanol}}$ should be a good descriptor for LOGPOW, and ideally we would expect a coefficient of $1/(RT \ln(10)) = 0.74$ mol/kcal at room temperature. A linear regression on 63 of the test compounds yields

$$\text{LOGPOW} = 0.75\frac{\text{mol}}{\text{kcal}}\Delta\mu^X_{OW} + 0.89n^X_{ar} - 0.55 \qquad (13)$$

with $r^2 = 0.95$ and $s = 0.31$, corresponding to a factor of 2. While the coefficient of $\Delta\mu^X_{OW}$ is almost ideal, we again have to take into account a correction for ring structures, but in this case for aromatic rings only. Apart from the constant and this correction, which hopefully will be removed by future improvements, eq 13 is a direct calculation of LOGPOW from the COSMO-RS model. Keeping this in mind it is remarkable that

TABLE 2: Experimental and Calculated Physicochemical Properties[a]

	log(p_v)		LOGPOW			τ	
	exp	eq 12	exp	eq 13	eq 14	exp	eq 15
methane		2.5	1.1	1.2	1.1		−0.01
ethane	1.6	1.8	1.8	1.8	1.6		0.00
n-pentane	−0.2	0.1	3.4	3.1	2.9	0.02	0.01
n-octane	−1.9	−1.5	4.9	4.5	4.3	0.03	0.03
neopentane	0.2	0.5	3.1	2.9	2.7	0.02	0.01
cyclobutane	0.1	0.2	2.2	2.2	2.1	0.02	0.03
cyclopentane	−0.5	−0.1	3.0	2.6	2.5	0.03	0.03
cyclohexane	−0.9	−0.6	3.4	3.1	2.8	0.04	0.03
ethene	1.8	1.8	1.1	1.2	1.2		0.01
ethine	1.7	1.3	0.4	0.5	0.6		0.03
propene	1.0	1.3	1.8	1.6	1.7	0.01	0.01
propine	0.7	0.8	0.9	0.9	1.1	0.02	0.03
propadiene	0.9	1.2	1.5	1.0	1.3	0.01	0.03
propanone	−0.6	−1.4	−0.2	−0.6	−0.5	0.03	0.04
butanone	−0.9	−1.3	0.3	0.2	0.2	0.04	0.04
2-pentanone	−1.3	−1.7	0.9	0.7	0.7	0.03	0.04
3-pentanone	−1.3	−1.2	0.8	0.9	0.9	0.04	0.04
3-methylbutanone	−1.2	−1.4	0.6	0.7	0.7	0.04	0.04
2-heptanone	−2.3	−2.7	2.0	1.5	1.6		0.05
3-heptanone	−2.1	−2.2	1.8	1.8	1.8	0.04	0.04
2-octanone	−2.9	−3.2	2.4	2.0	2.0		0.05
dimethyl ether	0.7	0.9	0.1	0.4	0.4		0.02
methyl propyl ether	−0.3	0.0	1.2	1.4	1.3		0.03
dipropyl ether	−1.1	−1.0	2.0	2.5	2.3	0.03	0.03
diisopropyl ether	−0.7	−0.4	1.5	2.5	2.2	0.03	0.03
1,4-dioxane	−2.4	−1.9	−0.4	−0.5	−0.1	0.05	0.06
chloromethane	0.8	0.7	0.9	0.8	1.0	0.02	0.03
dichloromethane	−0.9	−1.3	2.8	3.2	3.2	0.04	0.03
trichloromethane	−0.7	−0.6	2.0	2.2	2.1	0.04	0.03
tetrachloromethane	−0.3	−0.3	1.3	1.1	1.3	0.04	0.04
chloroethane	0.1	0.2	1.4	1.2	1.4	0.03	0.03
hexachloroethane		−3.0	4.1	4.0	4.1	0.04	0.05
tetrachloroethene	−1.7	−1.7	3.4	3.3	3.4	0.05	0.04
fluoromethane	1.5	1.1	0.5	0.1	0.3		0.03
bromomethane	0.3	0.1	1.2	1.1	1.2	0.03	0.03
dibromomethane	−1.3	−1.3	1.5	1.4	1.7	0.06	0.04
trifluorochloromethane	1.5	1.0	1.7	1.5	1.9		0.02
difluorodichloromethane	0.8	0.3	2.2	2.3	2.5	0.01	0.02
fluorotrichloromethane	0.0	−0.5	2.5	2.9	2.9	0.03	0.03
1,2-dichlorotetra fluoroethane	0.3	−0.6	2.8	3.0	3.2	0.02	0.03
methanol	−0.8	−0.4	−0.8	−0.4	−0.7	0.04	0.03
ethanol	−1.3	−0.5	−0.3	0.2	0.0	0.03	0.03
1-butanol	−2.2	−1.4	0.9	1.1	0.9	0.04	0.03
2-butanol	−1.8	−1.0	0.6	1.3	1.0	0.03	0.03
1-pentanol	−2.6	−1.9	1.6	1.6	1.4	0.04	0.03
1-heptanol	−3.6	−2.9	2.7	2.5	2.3	0.04	0.04
ethandiol	−4.0	−3.7	−1.4	−1.4	−1.4	0.07	0.07
phenol	−3.4	−3.3	1.5	1.3	1.2	0.06	0.06
p-cresol	−3.7	−3.6		1.7	1.6	0.05	0.06
benzyl alcohol	−4.2	−3.8	1.1	1.4	1.4	0.06	0.06
benzene	−1.0	−1.1	2.1	2.2	2.3	0.04	0.04
toluene	−1.6	−1.5	2.7	2.6	2.7	0.04	0.04
4-xylene	−2.0	−1.8	3.2	3.0	3.0	0.04	0.04
styrene	−2.1	−2.1	3.0	2.5	2.8	0.05	0.05
2-picoline	−2.0	−2.5	1.1	1.7	1.4	0.05	0.04
chinoline	−4.0	−4.0	2.0	1.6	1.8	0.06	0.06
aniline	−3.4	−3.9	0.9	0.9	0.9	0.06	0.06
dichlorobenzene		−2.9	3.4	3.4	3.4	0.05	0.05
hexachlorobenzene		−5.1	5.3	5.6	5.5		0.06
1-chloronaphthalene		−4.5	3.9	4.0	3.9	0.06	0.06
fluorobenzene	−1.0	−1.7	2.3	2.3	2.3	0.04	0.04
bromobenzene	−2.4	−2.7	3.0	2.8	2.9	0.05	0.05
water	−1.7	−2.1	−1.4	−1.8	−2.0	0.10	0.10
methylamine	0.5	0.0	−0.6	0.0	−0.4	0.03	0.02

[a] Vapor pressures p_V are given as log(p_V/bar), octanol/water partition coefficients p_{OV} as log(p_{OW}), and surface tensions τ in kcal/(mol Å^2). All values are for room temperature (T = 298 K).

the standard error of eq 13 compares with those of the incremental methods even in this early state of COSMO-RS.[30]

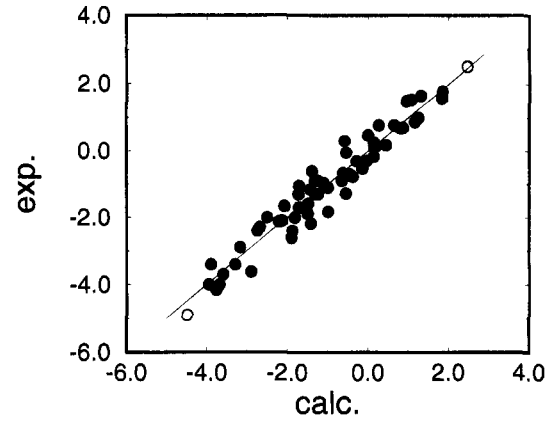

Figure 4. Experimental vs calculated logarithmic room temperature vapor pressures (in bar; see eq 12).

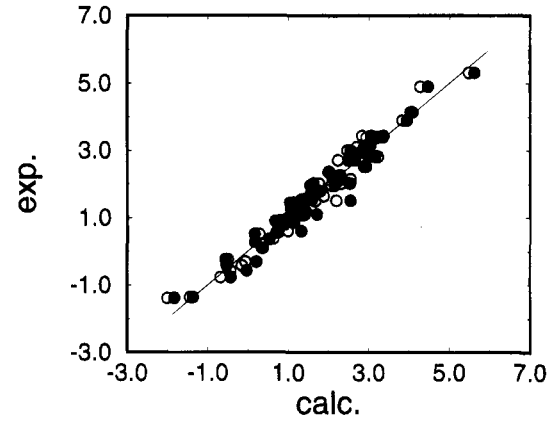

Figure 5. Experimental vs calculated logarithmic octanol/water partition coefficients (filled symbols, eq 13; open symbols, eq 14).

An even more precise equation for LOGPOW can be derived using the ideal screening energy Δ^X and the COSMO surface A^X instead of the chemical potential difference:

$$\text{LOGPOW} = 0.30\ \text{mol/kcal}\ \Delta^X + (0.061\ \text{Å}^{-2})A^X + 0.43n_{ar}^X + 0.72 \quad (14)$$

Here we have $r^2 = 0.96$ and $s = 0.28$. As shown in Appendix C, this expression corresponds to a σ-profile of the form $\mu'(\sigma) = a\sigma^2 + b\sigma + c$, fitted to the "true" σ-profile in the relevant σ-region. Thus, both equations are in good accordance.

The results of eqs 13 and 14 are given together with the experimental values of LOGPOW in Table 2 as well as in Figure 5. For both equations the strongest outlier is diisopropyl ether. We could not clarify the reason for this, yet.

Both equations provide a local measure of LOGPOW on the surface of the molecule, since, apart from the ring corrections, they can be written as surface integrals for the solute X. Thus, they may be used to visualize the lipophilicity of molecules as local property on the molecular surface. In addition, they allow for the study of the conformation dependence of the LOGPOW, since the evaluation of these equations can be done for different conformations of a molecule. These two advantages compared to the incremental methods should be of considerable use in many molecular modeling studies.

5.3. Surface Tensions. The surface tension τ or, more physically expressed, the specific surface energy of a fluid/gas interface shall be considered as a third exemplary application, because it is a little different from the equilibrium constants studied before. In eq 9 we have derive a result for the misfit contribution to the surface tension. In addition, we expect that

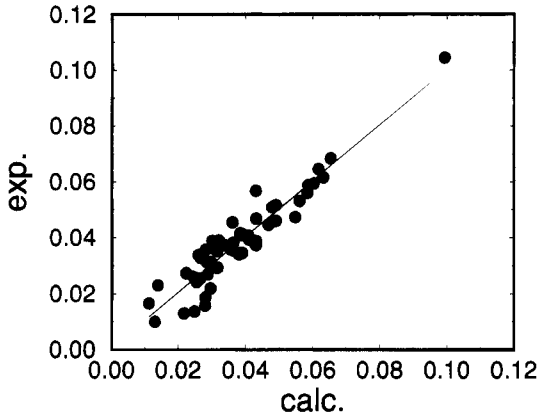

Figure 6. Experimental vs calculated surface tensions (in kcal/(mol Å^2); see eq 15).

the dispersive interaction should give a constant contribution $\gamma \approx 0.1$ kcal/(mol Å^2). Since the relationship between the microscopic surface, especially the COSMO surface, and the macroscopic surface is not fully clear, we expect a proportionality factor between this COSMO-RS surface tension contributions and the macroscopic surface tension, which should be not too far from unity. Nevertheless, in a multilinear regression on 53 of the test compounds, in addition to the dominating τ_S^0 term we find a significant contribution proportional to the molecular surface area A^X. We have no physical explanation for this yet. Accepting this empirical finding as well as a correction for nonaromatic rings, we have[31]

$$\tau = 0.79\tau_X^{vac} + 0.0007\frac{\text{kcal}}{\text{mol Å}^4}A^X + 0.016\frac{\text{kcal}}{\text{mol Å}^2}n_{nonar}^X - 0.032\frac{\text{kcal}}{\text{mol Å}^2} \quad (15)$$

with $r = 0.89$ and $s = 0.0056$ kcal/(mol Å^2). The experimental and calculated surface tensions are given in Table 2 and Figure 6. The latter underlines the extraordinary position of water with respect to surface tension.

Several further applications of this type could be given. In order to keep this article concise, we will present them in a forthcoming paper, together with a more detailed discussion of the results and probably after some improvement of the model.

6. Improvements

After we have demonstrated the encouraging results of the COSMO-RS approach even in its present state, we give an outlook over possible improvements in this section. In the above derivation of the COSMO-RS model we have made several approximations. Some of them are crucial for the model itself; others could be overcome at the cost of additional effort. At the moment we consider a few topics as most relevant for a refinement of the model. These are listed below, beginning with the underlying COSMO calculations:

(i) The surface parameters, i.e., the vdW radii used in the COSMO calculations, have to be optimized in order to get better ideal screening energies as well as improved charge densities. Probably a considerable part of the present error is due to this source.

(ii) The errors from the underlying semiempirical Hamiltonian have to be overcome. For example, the AM1 Hamiltonian as well as other semiempirical Hamiltonians give wrong partial dipole moments for several electronically sophisticated functional groups like nitro or cyano. This causes considerable errors in the screening energies of molecules containing these groups. We have therefore left out such problematic groups in our test set. Since this error at least to first order is constant for a given group, one could introduce correction dipoles located at the group atoms prior to the COSMO calculation. We have proved that this idea works quite well for a few functional groups, but obviously this post-parametrization is not the most elegant way out. Instead, a semiempirical parametrization would be desirable, which from the start focuses on a reliable reproduction of experimental dipole and quadrupole moments taking into account larger errors in other properties like bond lengths or spectra. Such a parametrization would be a good basis for solvation models like COSMO-RS. Nevertheless, it should be emphasized that AM1 yields good results for a wide range of organic chemistry.

(iii) Alternatively, the use of ab initio methods in combination with the COSMO model should be tested. Since in the COSMO-RS model only single molecule MO calculations are used, ab initio calculations would be affordable for most applications. Different ab initio/COSMO implementations are in preparation. Thus, in the near future we will be able to decide whether these are a good basis for COSMO-RS.

(iv) The contributions of zero-point and thermal vibrations and rotations to the free energy of solvation should be studied and taken into account, if they turn out to be relevant. For doing this the COSMO model should be extended with respect to the calculation of vibrational spectra in solution. Obviously, we have to be aware that the proper handling of librations may require considerable effort.

Now we consider outstanding improvements of the COSMO-RS approach itself:

(i) The dependence of the results on the detailed values of the two model parameters R_{eff} and f_{pol} should be studied, and an optimized pair of these parameters should be fixed.

(ii) The severeness of the neglect of the neighborhood constraints of segments should be tested by statistical simulations on simplified systems, consisting of hypothetical cubic or octahedral molecules, having different patterns of surface charge densities on their faces. The results of these simulations can be compared with the results derived for the decoupled segments. Thus, we get a feeling for the influence of frustration, i.e., the hindering of good segment pairing by neighborhood constraints, as well as for the influence of the neglect of misfit charge interactions, which at least to first order could be taken into account in such simulations.[24]

(iii) Up to now only a single conformation of a molecule is considered in COSMO-RS. For flexible molecules this may introduce some error, because different conformations can have different σ-profiles. The way to take multiple conformations of a molecule into account in COSMO-RS is quite obvious. First for each relevant conformation a COSMO calculation has to be done. Then their relative populations have to be calculated according to their total energies in a conducting continuum. Starting with this population the fluid can now be treated like a mixture of different compounds, and a σ-potential $\mu_S'(\sigma)$ can be calculated. From this we get the chemical potentials of the different conformations in the fluid. This may now be used for a new population analysis of the conformations, and the cycle is to be repeated until a self-consistent population is achieved. It can be expected that this procedure will converge in a few cycles, so that it also should take only seconds. Nevertheless, for large flexible molecules the conformational problem may remain as severe as it is in other approaches.

(iv) Hydrogen bonds are at present not considered explicitly. They appear to be quite well represented by the electrostatic

interaction taken into in COSMO-RS. Nevertheless, by a more detailed consideration or by the use of other Hamiltonians, it may become desirable to introduce an extra interaction term for hydrogen bonds. Since hydrogen bond accepting and donating groups are represented as highly charged segments in COSMO-RS, this could be done by an additional interaction energy of the form $-c\ (\sigma - \sigma')^{2n}$, where c is an appropriate positive constant and $n > 1$. This term only contributes considerably for segment pairs with high oppositely signed charge densities, i.e., typical hydrogen bonding pairs. Such an additional energy in the exponent in eq 10 would not make any difference, neither to its validity nor to its applicability. Thus, the COSMO-RS model is open for such interaction energy correction, which depend on the charge densities of the segment pairs, exclusively.

(v) Finally, the concept of COSMO-RS may be generalized to more than one segment property dimension without major problems. In the moment the surface charge density σ is the only segment property considered, but it should be no problem to take into account other useful properties. Especially the local polarizability, i.e., the polarization answer of the molecule to a misfit surface charge density applied at an individual segment, should be an interesting property, since it could help to overcome the average polarizability approximation. Furthermore, this property might provide a key to replace the global dispersion term γA^X by a more sophisticated local description, based on the local polarizabilities of the adjacent segments. Obviously, several other local properties for the description of dispersion or hydrogen bonds might be considered alternatively or in addition.

7. Possible Fields of Application

Because COSMO-RS is a quite general theory for pure or mixed fluid systems, with one or the other improvement or modification it should be applicable to a great variety of problems in chemistry. Without trying to be complete a few possible areas of application are addressed in this chapter:

As has been shown in section 5, COSMO-RS provides access to the thermodynamic parameters of solvents as well as of solute molecules. It provides the knowledge of the chemical potential of any compound in a fluid and thus enables the calculation of phase diagrams. Since phase diagrams often are very sensitive to small errors in the chemical potential, we think that in the present state of COSMO-RS a reliable calculation of phase diagrams is not possible for many systems, but taking into account the considerable potential for improvements they should become achievable in the near future. In this context it is important that COSMO-RS handles multicomponent mixtures without additional complications, because it does not make any mean-field approximation. Thus, it can be expected that multicomponent phase diagrams can be described as reliable as binary or fluid-gas phase diagrams in the future.

If the σ-profiles and σ-potentials of the most common solvents are stored in a database, a highly efficient screening of pure or mixed solvents, well suited for a new solvation or separation problem, can be performed, since the calculation of the chemical potentials of the solutes of interest in all of these solvents can be done automatically within minutes or hours.

The COSMO-RS approach should also be capable to yield useful parameters for the estimation of solubility and miscibility properties of polymer melts. A first preliminary study on these questions has given encouraging results.

By the introduction of a one-dimensional position dependence in the σ-profile $p_S(\sigma)$ as well as in the σ-potential $\mu_S'(\sigma)$, i.e., introducing $p_S(\sigma,x)$ and $\mu_S'(\sigma,x)$, also the treatment of inhomogeneous systems like membranes should be achievable. In such a field the optimal position of a solute molecule, which now is considered with full geometry, could easily be calculated from the chemical potential of the surface segments at their entire positions.

Apart from the above-mentioned more thermodynamically oriented applications, COSMO-RS may also be of considerable use for molecular modeling and drug design. As discussed in context with the LOGPOW, it provides a new access to questions of hydrophilicity and lipophilicity. Furthermore, the representation of enzymatic receptors by a set of surface segments, which correspond to hydrogen bond donors and acceptors, medium polar and lipophile regions, respectively, and which characterize the enzymatic receptor, should be an interesting new approach for the fitting of molecules to receptors.

8. Summary

On the basis of screening energies, surface areas, and screening surface charge densities, calculated for the compounds of interest with a continuum solvation model like COSMO, the intermolecular interactions in a liquid system can be described as pairwise interactions of surface segments. This description is much more efficient than the traditional description via electrostatic and vdW interactions, because in contrast to these the explicit position of all molecules in space has not to be taken into account. Thus, a highly efficient calculation of the chemical potential of any solute in a pure or mixed solvents is enabled, which provides access to most of the relevant thermodynamic parameters of liquid systems.

Even in its present very preliminary state, the COSMO-RS model provides interesting qualitative insight into solvation phenomena as well as encouraging quantitative results for different thermodynamic parameters. Nevertheless, there is a considerable potential for improvements of the model, which should be achievable with reasonable effort. Thus, it can be expected that COSMO-RS will become a valuable supplement for the traditional incremental methods in physical chemistry and chemical engineering, as well as a powerful new tool for molecular modeling.

Acknowledgment. I express my gratitude to my collegues for several valuable hints and discussions. I especially thank Jannis Batoulis and Peter Hoever for their patience in our difficult discussions on the partition function of the "sandwich ensemble". I am indebted to Vera Nordhausen for correcting my English.

Appendix A: Derivation of the Self-Consistency Equation

Select one of the segments, give to it the index 0, and number the remainder from 1 to $N - 1$. Let Z be the partition function of this ensemble. Assume that the chemical potential $\mu(\sigma) = -kT\ \mathrm{d}Z/\mathrm{d}N(\sigma)$ is known for any value of σ. Then in the thermodynamic limit of large N the partition function of an ensemble deminished by two segments i and j is given by the old partition function weighted by a Boltzmann factor due to the free energy necessary for the removal of the two segments, i.e., $Z \exp\{(\mu(\sigma_i) + \mu(\sigma_j))/kT\}$. As noted in section 3.2, the degeneracy of an ensemble of segments is $2^{N/2}(N/2)!$. Thus, the degeneracy of the orignal ensemble is by a factor N larger than that of the ensemble diminished by two segments.

Now the probability $p(0,i)$ to find a pair $(0,i)$ in the original ensemble is given by the Boltzmann weight of the misfit energy of the pair $(0,i)$, the partition sum of the ensemble without segments 0 and i, and the increase in deneracy of the starting

ensemble compared to the smaller one, i.e.

$$p(0,i) = N \exp\{-{}^1/_2\alpha'(\sigma_0 + \sigma_i)^2/kT\}\, Z \exp\{(\mu(\sigma_0) + \mu(\sigma_i))/kT\}$$

Since in the original ensemble segment 0 always has to be paired with one of the $N-1$ remaining segments, the partition function of the starting ensembles must be the sum of all $N-1$ probabilities to find segment 0 in a pair $(0,i)$. Hence we have

$$Z = \sum_{i=1}^{N-1} p(0,i) = \sum_{i=1}^{N-1} N \exp\{-{}^1/_2\alpha'(\sigma_0 + \sigma_i)^2/kT\} Z \times \exp\{(\mu(\sigma_0) + \mu(\sigma_i))/kT\}$$

If we now first divide by $Z \exp\{-\mu(\sigma_0)/kT\}$, then switch to the logarithm, and finally multiply by $-kT$, we obtain the equation

$$\mu(\sigma_0) = -kT \ln[N \sum_{i=1}^{N-1} \exp\{-({}^1/_2\alpha'(\sigma_0 + \sigma_i)^2 - \mu(\sigma_i))/kT\}]$$

If we now let the sum run over all N segments instead of the $N-1$ remaining segments, which can safely be done for large values of N, and then replace the sum by $N \int p(\sigma')\, d\sigma'$ and rename σ_0 by σ, then we have our final result

$$\mu(\sigma) = -kT \ln[N^2 \int d\sigma'\, p(\sigma') \exp\{-({}^1/_2\alpha'(\sigma + \sigma')^2 - \mu(\sigma'))/kT\}]$$

Appendix B: Analytical Treatment of Gaussian σ-Profiles

For Gaussian σ-profiles centered at $\sigma = 0$, i.e., for σ-profiles of the form

$$p(\sigma) = (2\pi s^2)^{-1} \exp\{-{}^1/_2(\sigma/s)^2\}$$

where s is the width of the profile, the self-consistency equation (8) can be solved by analytical means. Since for a given σ-profile the solutions of eq 8 are unique, we only need to prove that a trial solution for $\mu'(\sigma)$ really satisfies this equation. We use the trial functions $\mu'(\sigma) = \mu'_0 + {}^1/_2a\sigma^2$ here. Inserting this and the Gaussian σ-profile into eq 6, we have

$$\mu_0' + {}^1/_2a\sigma^2 = -kT \ln \int d\sigma'\, (2\pi s^2)^{-1/2} \exp\{-{}^1/_2(\sigma'/s)^2\} \times \exp\{[-{}^1/_2\alpha'(\sigma + \sigma')^2 + \mu_0' + {}^1/_2a\sigma'^2]/kT\}$$

This integral is still of Gaussian type, and it can thus be solved easily, yielding

$$\mu_0' + {}^1/_2a\sigma^2 = -kT \ln\left[\left(1 + \frac{\alpha' - a}{kT} s^2\right)^{-1/2} \times \exp\left\{\left[\mu_0' - {}^1/_2\alpha'\sigma^2\left(1 + \frac{\alpha' s^2}{kT + (\alpha' - a)s^2}\right)\right]/kT\right\}\right]$$

$$= {}^1/_2kT \ln\left(1 + \frac{\alpha' - \alpha}{kT} s^2\right) - \mu_0' + {}^1/_2\alpha'\sigma^2\left(1 + \frac{\alpha' s^2}{kT + (\alpha' - a)s^2}\right)$$

If this equation holds for any value of σ, the constants as well as the σ^2 coefficients on both sides must be equal. Thus we have the two equations:

$$2\mu_0' = {}^1/_2kT \ln\left(1 + \frac{\alpha' - a}{kT} s^2\right)$$

and

$$a = \alpha' \frac{kT - as^2}{kT + (\alpha' - a)s^2} \leftrightarrow a^2s^2 - (2\alpha's^2 + kT)a + \alpha'kT = 0$$

The solution of the quadratic equation for a is

$$a = \alpha' + \frac{kT}{2s^2} \pm \alpha'\left[1 + \left(\frac{kT}{2\alpha's^2}\right)^2\right]^{1/2}$$

Inserting this result into the equation from the constant terms yields

$$2\mu_0' = {}^1/_2kT \ln\left\{{}^1/_2 - \pm{}^1/_2\left[1 + \left(\frac{2\alpha's^2}{kT}\right)^2\right]^{1/2}\right\}$$

To avoid the logarithm of a negative number, we have to choose the negative sign in the solution for a. Thus we get

$$\mu_0' = {}^1/_4kT \ln\left\{\frac{1}{2} + \frac{1}{2}\left[1 + \left(\frac{1\alpha's^2}{kT}\right)^2\right]^{1/2}\right\}$$

For very low temperatures ($kT \ll 2\alpha's^2$) these results become $a \approx 0$ and $\mu_0' \approx 0$, i.e., $\mu'(\sigma) \equiv 0$. This reflects the fact that for symmetric σ-profiles the ground state misfit energy and its gradient with respect to the segment numbers are zero. For high temperature with $kT \gg 2\alpha's^2$ we are in the limit of total disorder, where the pairing of segments does not contribute to the screening of an additional segment. Thus this is screened by polarization only. In agreement with this we find $a = \alpha'$ and $\mu_0' = 0$ in this limit.

Appendix C: Parabolic σ-Potentials

If the σ-potential $\mu_S'(\sigma)$ of a solvent S or more general the difference $\Delta\mu_{SS'}$ of the σ-potentials of two liquids S and S' can be approximated by a general parabola, i.e., if it is $\Delta\mu_{SS'} \approx a\sigma^2 + b\sigma + c$, then for a solute X the difference of its chemical potentials in these liquids is given by

$$\Delta^X_{SS'} = N^X \int d\sigma\, p^X(\sigma)\, \Delta'_{SS'}(\sigma) = N^X \int d\sigma\, p^X(\sigma)(a\sigma^2 + b\sigma + c)$$

Realizing that within the COSMO-RS approximations, we have

$$N^X \int d\sigma\, p^X(\sigma)\sigma^2 \approx 2\Delta^X/\alpha$$

and

$$N^X \int d\sigma\, p^X(\sigma)\sigma = Q^X$$

where Q^X is the total charge of solute X and

$$N^X \int d\sigma\, p^X(\sigma) = A^X$$

and finally we get for neutral solutes the result

$$\Delta^X_{SS'} = 2(a/\alpha)\Delta^X + cA^X$$

Since in a limited σ-range near $\sigma = 0$ the parabolic approximation will be reasonably accurate for most solvents S or solvent pairs S and S', many logarithmic Henry constants and partition coefficients should be describable by a bilinear regression with respect to the ideal screening energy Δ^X and the molecular

surface A^X, at least for not too polar solutes. This is just what we have observed for the octanol/water partition coefficients.

References and Notes

(1) Lydersen, A. L. Estimation of Critical Properties of Organic Compounds, University of Wisconsin, College Engineering, Eng. Exp. Stn. Rep. 3, Madison, WI, 1955.

(2) Benson, S. W. *Chem. Rev.* **1969**, *69*, 279–324.

(3) Fredenslund, A.; Gmehling, J.; Rasmussen, P. *Vapor Liquid Equilibria Using UNIFAC*; Elsevier: Amsterdam, 1977.

(4) Abrams, D.; Prausnitz, J. M. *AIChE J.*, **1975**, *21*, 116.

(5) Hansch, C.; Leo, A. J. *Substituent Parameters for Correlation Analysis in Chemistry and Biology*, Wiley: New York, 1979; CLOGP program, Daylight CIS, Irvine, CA.

(6) van Gunsteren, W. F. *Angew. Chem.* **1990**, *102*, 1020.

(7) Luzhkov, V.; Warshel, A. *J. Comput. Chem.* **1992**, *13*, 199.

(8) Laasonen, K.; Sprik, M.; Parrinello, M.; Car, R. *J. Chem. Phys.* **1993**, *99*, 9080.

(9) Born, M. *Z. Phys.* **1920**, *1*, 45.

(10) Kirkwood, J. G. *J. Chem. Phys.* **1934**, *2*, 351.

(11) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58*, 1486.

(12) Cramer, C. J.; Truhlar, D. G. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; VCH Publishers: New York, Vol. 6, in press.

(13) Wong, M. W.; Frisch, M. J.; Wiberg, K. B. *J. Am. Chem. Soc.*, **1991**, *113*, 4776.

(14) Karelson, M. M.; Katritzky, A. R.; Zerner, M. C. *Int. J. Quantum Chem. Symp.* **1984**, *113*, 4776.

(15) Klamt, A.; Schüürmann, G. *J. Chem. Soc., Perkin Trans. 2* **1993**, 799.

(16) Throughout this paper the MOPAC93 version of COSMO is used with all parameters kept from the defaults (see MOPAC93 manual[18]). Note that in contrast to the original COSMO paper[15] now atomic quadrupole moments are taken into account. The vdW radii used in the surface construction are (in Å) 1.08 for H, 1.52 for C, 1.48 for N, 1.36 for O, 1.30 for F, 1.65 for Cl, and 1.80 for Br.

(17) Stewart, J. J. P. *J. Comput.-Aided Mol. Design* **1990**, *4*, 1.

(18) Stewart, J. J. P. MOPAC program package, QCPE-No. 455, 1993. Stewart, J. J. P. MOPAC93.00 Manual 1993, Fujitsu Ltd., Tokyo, Japan.

(19) Miertus, S.; Scrocco, E.; Tomasi, J. *Chem.Phys.*, **1981**, *55*, 117.

(20) Negre, N.; Orozco, M.; Luque, F. J. *Chem. Phys. Lett.* **1992**, *196*, 26.

(21) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1993**, *115*, 8810.

(22) Szafran, M.; Karelson, M. M.; Katritzky, A. R.; Koput, J.; Zerner, M. C. *J. Comput. Chem.* **1993**, *14*, 371.

(23) See e.g.: Jackson, J. D. *Classical Electrodynamics*; Wiley: New York, 1975.

(24) Note added in proof: Recent MC simulations on cubic molecules by J. Batoulis and R. Merkt indicate that the decoupling approximation is highly accurate, as long as the system is disordered. Only below the phase transition to an ordered system, which corresponds to crystallization in real fluids, considerable differences between the MC results and the results derived within the decoupling approximation have been found. The results of these simulation will be discussed in more detail elsewhere.

(25) These MC simulations have been performed by Jannis Batoulis, Bayer, AG.

(26) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.

(27) The Bayer in-house property database is a collection of data from different sources. Some of these data have been averaged or interpolated in order to yield room temperature data.

(28) THOR database, Daylight CIS, Irvine CA.

(29) Meylan, W.; Howard, P. *Users Guide for LOGKOW*; Syracuse Research Corporation: Syracuse, NY, 1994.

(30) It should be noted that on this data set the incremental methods give a smaller standard deviation, but most of these quite small compounds have been used for the fitting of the large number of increments. Thus on this data set they are not really predictive.

(31) Surface tensions are given in kcal/(mol Å^2) here. The conversion factor to the more common units dyn/cm ≡ mN/m is 696.5.

JP9416883