General Semiempirical Quantum Mechanical Solvation Model for Nonpolar Solvation Free Energies. *n*-Hexadecane

David J. Giesen,† Joey W. Storer, Christopher J. Cramer,* and Donald G. Truhlar*

Contribution from the Department of Chemistry and Supercomputer Institute, University of Minnesota, Minneapolis, Minnesota 55455-0431

Received September 9, 1994. Revised Manuscript Received November 7, 1994⊗

Abstract: A new solvation model has been developed that accurately predicts solvation free energies in the nonpolar solvent *n*-hexadecane. The model is based on AM1-CM1A and PM3-CM1P partial charges, and it is based on a single set of parameters that is applicable to both the AM1 and PM3 Hamiltonians. To take account of both short-range and long-range solvation-shell interactions, each atom has two surface tensions associated with different effective solvent radii. For hydrogen, one of these surface tensions depends on the bond orders to carbon, nitrogen, oxygen, and sulfur, although only weakly. In addition to presenting the general parameterization, the article provides an analysis of the surface tension parameterization based on data for three rare gases. The model yields an rms error of 0.41 kcal/mol over a set of 306 data points (153 molecules, two Hamiltonians) that includes alkanes, alkenes, alkynes, aromatics, alcohols, ethers, aldehydes, ketones, esters, amines, nitriles, pyridines, thiols, sulfides, fluorides, chlorides, bromides, iodides, water, and ammonia.

1. Introduction

While most experimental chemistry takes place in the condensed phase, most computational research corresponds to gas-phase conditions. There has been great progress in recent years in developing techniques for including solvation effects in quantum mechanical electronic structure calculations,[1] and we have developed semiempirical models of useful accuracy for solvation free energies of organic and small inorganic solutes in water.[2−5] In the present work we extend this treatment to *n*-hexadecane solutions. *n*-Hexadecane has a dielectric constant, ϵ, of 2.06[6] at 298 K and is a particularly interesting solvent because it is sometimes used (as is octanol) as a model for lipid bilayers.[7,8] We also note that solvation free energies in *n*-hexadecane are experimentally indistinguishable (at the current level of experimental precision) from those for isooctane (2,2,4-trimethylpentane),[9] so *n*-hexadecane to some extent can serve as a surrogate for large alkane solvents in general and perhaps for other nonpolar solvent-like environments. As one possible area of application of nonpolar solvation, we note that the active site of an enzyme may be nonpolar, so it is important to understand differences in polar and nonpolar solvation when we model enzyme−substrate interactions. Nonpolar solvents, e.g. hexane, are also sometimes used for organic synthesis. Since *n*-hexadecane is used widely as a stationary phase in gas chromatography, a large body of experimental data is available, and this makes it a suitable choice for developing a model for nonpolar solvation.

Our solvation models for water, called SM*x*, where *x* denotes the specific parameterization of a particular solvation model,[2−5] account for the free energy of solvation as a sum of two components. The first, ΔG_{ENP}, accounts for the electric polarization of and by the solvent (treated as a continuum dielectric) and includes the associated costs of distorting the solute and solvent internal structures (electronic and nuclear) and the solvent molecules' orientation in order to increase the favorable solute−solvent polarization interactions. This term is based on a self-consistent reaction field (SCRF)[1,10] formalism for a homogeneous continuum solvent. The second component, G°_{CDS}, accounts for the free energy of forming a cavity in the solvent to make room for the solute and for the changes in dispersion interactions and solvent structure that accompany the solvation process. One critical extension was required for treating G°_{CDS} in *n*-hexadecane that was not required for water, and this is explained next.

In water we wrote the G°_{CDS} term as

$$G^{\circ}_{CDS} = \sum_{\gamma} \sigma_{\gamma} A_{\gamma}(R_S) \qquad (1)$$

where the index γ runs over atoms, σ_{γ} is a surface tension that depends on atom type, and $A_{\gamma}(R_S)$, which is a function of the solvent molecular radius, R_S, is the solvent-accessible surface area of atom γ. Although not indicated explicitly in the notation, it should be understood that A_{γ} depends not only on R_S but also on all the atomic radii, R_{γ}, of atoms in the molecule and on the molecular geometry. In particular, $A_{\gamma}(R_S)$ is calculated by rolling a sphere of radius R_S over the van der Waals surface of the molecule.[11,12] The center of the sphere maps out a surface

† Kodak Fellow.

* Authors to whom correspondence should be addressed.

⊗ Abstract published in *Advance ACS Abstracts,* January 1, 1995.

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

(2) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1991**, *113*, 8305, 9901(E).

(3) Cramer, C. J.; Truhlar, D. G. *Science* **1992**, *256*, 213.

(4) Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.* **1992**, *13*, 1089.

(5) Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 629.

(6) *CRC Handbook of Chemistry and Physics*, 71st ed.; Lide, D. R., Ed.; CRC Press: Boca Raton, FL, 1990−1991.

(7) Venable, R. M.; Zhang, Y.; Hardy, B. J.; Pastor, R. W. *Science* **1993**, *262*, 223.

(8) Solé-Violan, L.; Devallex, B.; Postel, M.; Riess, J. G. *New J. Chem.* **1993**, *17*, 581.

(9) Abraham, M. H.; Whiting, G. S. *J. Chem. Soc., Perkin Trans. 2* **1990**, 291.

(10) Tapia, O., Daudel, R., Pullman, A., Salem, L., Veillard, A., Eds.; *Quantum Theory of Chemical Reactions;* Reidel: Dordrecht, The Netherlands, 1980; Vol. 2, p 25.

(11) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.

(12) Richards, F. M. *Annu. Rev. Biophys. Bioeng.* **1977**, *6*, 151.

0002-7863/95/1517-1057$09.00/0

that passes through the center of a solvation shell of thickness $2R_S$ surrounding the molecule, and hence, the area of this surface is proportional (in the sense of a continuum treatment of the solvent) to the number of solvent molecules in the first solvation shell. G°_{CDS} computed in this manner is thus a measure of the extent to which the solvation free energy differs from the value of ΔG_{ENP} (which, as explained above, is due to homogeneous-solvent electrostatics) because of interactions of the solute with solvent molecules in the first solvation shell. The dispersion interactions are assumed proportional to the number of solvent molecules in the first solvation shell because dispersion is too short ranged for interactions of the solute with second-solvent-shell molecules to be significant. The solvent structural changes (both those due to solvent−solvent interactions and those due to solvent−solute interactions) are also assumed to be approximately proportional to this number because solvent structural changes beyond the first hydration shell are assumed to have a smaller effect than the typical error in the model due to other sources.

Formation of a cavity *without* a solute involves changes in both dispersion and solvent structure, and the cavity formation component of the free energy is included in G°_{CDS} as well. Here we expect another significant difference between *n*-hexadecane and water. In water there is an entropically unfavorable contribution to the formation of a cavity due to loss of water−water hydrogen bonding possibilities for molecules on the surface of the cavity. In hydrocarbons, hydrogen bonding is not an issue, and instead, cavity formation is entropically favored due to destruction of the order that is characteristic of pure solvent.[13,14] Thus, for water, the most important *negative* contributions to G°_{CDS} are the energetic effects of dispersion and, when present, of solute−solvent hydrogen bonding. For hydrocarbons, on the other hand, it is the dispersion energy and the entropy of cavity formation that are expected to be important negative contributors.

In water, the choice for R_S is 1.4 Å,[11,12] an accepted value for the effective radius of a spherical model of a water molecule. In *n*-hexadecane, however, the choice for R_S is more ambiguous. For dispersion, the important part of the solvent probably consists only of those parts of the *n*-hexadecane molecules actually in contact with the van der Waals surface of the solute. The thickness of the "dispersion shell" is probably the same as if the solvent was methane. Thus, for short-range forces, we choose a solvent radius of 2.0 Å. This value comes from molecular beam data, in which two methane molecules are found to have a van der Waals minimum-energy structure at a distance of approximately 4.0 Å.[15,16] However, solvent structural changes surely extend much farther into solution. For long-range interactions, we choose a solvent radius of 4.9 Å. The model was relatively insensitive to radius choice once the radius was in excess of about 4 Å, so it was simply taken equal to the radius of a sphere that contains the same volume as that of a *n*-hexadecane molecule as calculated from the density of *n*-hexadecane at 298 K.[6,17] A geometric representation is provided in Figure 1. We refer to the short-range interactions as CD terms for cavitation and dispersion and the long-range

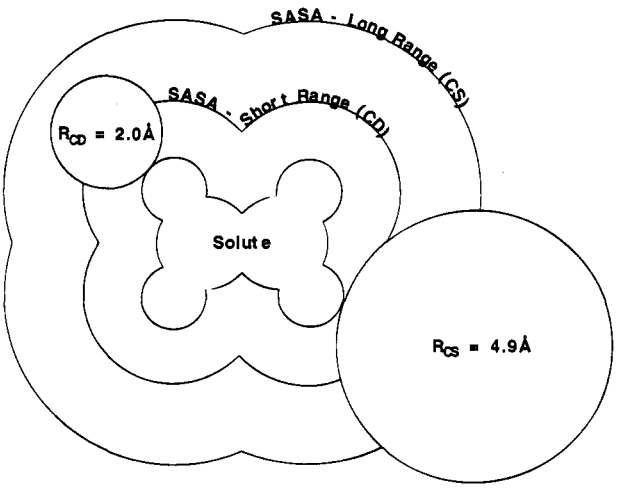

Figure 1. Planar depiction of the two differently sized spherical solvent probes tracing out the centers of their respective first solvation shells (whose areas are the solvent-accessible surface areas).

interactions as CS terms for cavitation and structure rearrangement. Thus, for large solvents like *n*-hexadecane, it seems more reasonable to write

$$G^{\circ}_{CDS} = G_{CD} + G_{CS} \tag{2}$$

where

$$G_{CD} = \sum_{\gamma} \sigma_{\gamma}^{CD} A_{\gamma}(R_S^{CD}) \tag{3}$$

and

$$G_{CS} = \sum_{\gamma} \sigma_{\gamma}^{CS} A_{\gamma}(R_S^{CS}) \tag{4}$$

Furthermore, we would expect σ_{γ}^{CD} to probably be negative and σ_{γ}^{CS} to probably be positive. As described above, R_S^{CD} is set to 2.0 Å and R_S^{CS} is set to 4.9 Å.

For a large solvent that can make specific interactions (e.g., hydrogen bonds) with some solute functional groups, σ_{γ}^{CS} could well depend on the type of atom γ. For *n*-hexadecane, however, the solvent structural changes probably depend more on the existence of the cavity than on what is actually in it, so σ_{γ}^{CS} might reasonably be expected to be almost independent of the type of atom γ. Equation 4 then becomes

$$G_{CS} = \sigma^{CS} \sum_{\gamma} A_{\gamma}(R_S^{CS}) \tag{5}$$

Dispersion interactions, though, depend on the polarizability of the solute atoms, so σ_{γ}^{CD} surely depends significantly on the type of atom γ. We will see that a reasonable set of parameters can be found that are in accord with the expectations just given.

A second aspect in which the present solvation model differs from our previous models is that it is based on a new method that we have recently developed for assigning partial charges, in particular the CM1 partial charge method.[18] The CM1 models, CM1A (based on semiempirical Austin Model 1[19] wave functions) and CM1P (based on semiempirical Parameterized Model 3[20] wave functions), are class IV charge models

(13) Eley, D. D. *Trans. Faraday Soc.* **1939**, *35*, 1421.
(14) Krestov, G. A. *Thermodynamics of Solvation. Solution and Dissolution; Ions and Solvents; Structure and Energetics*; Ellis Horwood: New York, NY, 1991; pp 167−170.
(15) Reid, B. P.; O'Loughlin, M. J.; Sparks, R. K. *J. Chem. Phys.* **1985**, *83*, 5656.
(16) Buck, U.; Kohlhase, A.; Secrest, D.; Phillips, T.; Scoles, G.; Grein, F. *Mol. Phys.* **1985**, *55*, 1233.
(17) Rossini, F. D.; Pitzer, K. S.; Arnet, R. L.; Braun, R. M.; Pimentel, G. C. *Selected Values of Physical and Thermodynamic Properties of Hydrocarbons and Related Compounds*; Carnegie Press: Pittsburgh, PA, 1953.
(18) Storer, J. W.; Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Comput.-Aided Mol. Des.*, in press.
(19) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. *J. Am. Chem. Soc.* **1985**, *107*, 3902.
(20) Stewart, J. J. P. *J. Comput. Chem.* **1989**, *10*, 221.

(explained below) that yield atomic charges similar to high-level *ab initio* charges but with a computational effort the same as obtaining the very inexpensive AM1 or PM3 wave functions.[18] Because class IV charges based on either of these Hamiltonians produce similar, physically reasonable electrostatic interactions, we find that we are able to obtain a single set of solvation parameters that may be used to calculate accurate free energies of solvation with either the AM1 or PM3 Hamiltonian. Moreover, these solvation parameters should be applicable with any set of high-quality partial charges and geometries, whether obtained from experiment or from high-level electronic structure calculations. We have also used this class IV charge method for improved calculations in aqueous solutions,[21] and all solvation models developed with the new models are denoted as level-4 parameterizations, e.g. SM4 or SM4*x*, where *x* specifies the specific level-4 parameter set. The present model, in particular, will be denoted SM4. It will be included in a future release (5.0) of the AMSOL program.[22]

Section 2 presents details of the theory, with section 2.1 reviewing the aspects of previous SM*x* models that we shall incorporate into the SM4 model for *n*-hexadecane solvent and section 2.2 presenting the new elements required for eqs 2−5 and for using the CM1A and CM1P charge models. Section 3 presents the parameterization and results. Section 4 discusses the resulting solvation model, and section 5 presents concluding remarks.

2. Theory

2.1. General Aspects. A successful series of methods, called the SM*x* models, has already been introduced that calculates free energies of solvation in water.[2−5] This formalism has been presented in full in earlier papers,[2−5] and the relevant elements are briefly reviewed here.

We begin by writing the free energy of solvation as

$$\Delta G_S^\circ = G_{(sol)}^\circ - G_{(g)}^\circ \tag{6}$$

where $G_{(sol)}^\circ$ is the standard-state free energy of the solute in solution and $G_{(g)}^\circ$ is its standard-state free energy in the gas phase. We use a standard state of 1 mol/L in both phases. Furthermore, we restrict our attention in this paper to a temperature of 298 K.

We can partition ΔG_S° as

$$\Delta G_S^\circ = E_{EN(sol)} - E_{EN(g)} + \Delta G_{vib} + \Delta G_{elec} + G_P + G_{CDS}^\circ \tag{7}$$

where E_{EN} is the sum of the solute electronic kinetic and electronic nuclear Coulombic energies either in the gas phase or in the presence of the solvent, ΔG_{vib} and ΔG_{elec} are the change in internal vibrational and electronic free energy upon solvation, G_P is the free energy of the electric polarization, including both solute−solvent interaction energy and solvent reorganization energy, and G_{CDS}° is the remaining solvation free energy, which is assumed to correlate with solvent-accessible surface area. The three dominant contributions to G_{CDS}° were identified in the Introduction; a more complete list includes not only cavitation, dispersion, and energetic and entropic effects from structural changes in the solvent but also the sum of the changes in free energy due to ΔPV terms, the creation of libron modes (i.e., the three-to-six new modes of the solution that replace the gas-phase translation and rotation of the solute), the deviation of all first-solvation-shell electrostatic effects from their values calculated by the SM*x* treatment based on the bulk dielectric constant (e.g., dielectric saturation), specific interactions such as hydrogen bonding, and any deviation of the entropy of mixing from its ideal value that correlates with solute area. We will return to the discussion of the entropy of mixing later in this subsection.

In the present paper, we ignore ΔG_{vib} and ΔG_{elec}, which are expected to be very small. This allows us to write

$$\Delta G_S^\circ = \Delta G_{ENP} + G_{CDS}^\circ \tag{8}$$

where ΔG_{ENP} is now the sum of G_P and $E_{EN(sol)}$ minus $E_{EN(g)}$. G_P can be calculated by the generalized[1,10,23−30] Born[31,32] equation

$$G_P = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{k,k'} q_k q_{k'} \gamma_{kk'} \tag{9}$$

where ϵ is the dielectric constant of the solvent, q_k is the partial atomic charge on atom k, and $\gamma_{kk'}$ is a Coulomb integral. We use a modified[2−5] form of the Coulomb integral suggested by Still *et al.*,[30] namely

$$\gamma_{kk'} = (r_{kk'}^2 + \alpha_k\alpha_{k'}\exp(-r_{kk'}^2/4\alpha_k\alpha_{k'}) + d_{kk'}^{(1)}\exp\{-d_{kk'}^{(2)}/\{1 - [(r_{kk'} - r_{kk'}^{(1)})/r_{kk'}^{(2)}]^2\}\})^{-1/2} \tag{10}$$

where $r_{kk'}$ is the distance between atoms k and k' and α_k is the effective Coulomb radius of atom k. For a single atom, α_k is set equal to an intrinsic Coulomb radius, ϱ_k, given by

$$\varrho_k = \varrho_k^{(0)} + \varrho_k^{(1)}\left[-\frac{1}{\pi}\arctan\frac{q_k + q_k^{(0)}}{q_k^{(1)}} + \frac{1}{2}\right] \tag{11}$$

where $\varrho_k^{(0)}$, $\varrho_k^{(1)}$, $q_k^{(0)}$, and $q_k^{(1)}$ are empirical parameters such that ϱ_k is a smooth, sigmoidal function that switches from $\varrho_k^{(0)}$ at large positive q_k to $\varrho_k^{(0)} + \varrho_k^{(1)}$ at large negative q_k. The parameter $q_k^{(0)}$ is the value of the inflection point of this curve, and $q_k^{(1)}$ defines the steepness of the switch from $\varrho_k^{(0)}$ to $\varrho_k^{(0)} + \varrho_k^{(1)}$. In all models thus far, $q_k^{(1)}$ has been fixed at 0.1. In multicenter cases, α_k is determined numerically according to the dielectric screening model of Still *et al.*[30] with one of two possible radial quadrature schemes presented elsewhere.[5,21,30,33] Note that α_k depends on the set of all $\varrho_{k'}$ and that this method specifically allows for the treatment of arbitrarily shaped solutes. The dielectric screening model may be considered to be an approximate solution of the Poisson equation. Although it is

(21) Storer, J. W.; Giesen, D. J.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G.; Liotard, D. A. In *Structure and Reactivity in Aqueous Solution*; ACS Symposium Series 568; Cramer, C. J., Truhlar, D. G., Eds.; American Chemical Society: Washington, DC, 1994; p 24.
(22) Cramer, C. J.; Lynch, G. C.; Hawkins, G. D.; Giesen, D. J.; Truhlar, D. G.; Liotard, D. A. *QCPE Bull.* **1994**, *14*, 55.
(23) Hoijtink, G. J.; De Boer, E.; Van der Meij, P. H.; Weiland, W. P. *Recl. Trav. Chim. Pays-Bas* **1956**, *75*, 487.
(24) Peradejordi, F. *Cah. Phys.* **1963**, *17*, 393.
(25) Jano, I. *C. R. Hebd. Seances Acad. Sci.* **1965**, *261*, 103.
(26) Fischer-Hjalmars, I.; Hendriksson-Enflo, A.; Hermann, C. *Chem. Phys.* **1977**, *24*, 167.
(27) Constanciel, R.; Contreras, R. *Theor. Chim. Acta* **1984**, *65*, 1.
(28) Kozaki, T.; Morihasi, M.; Kikuchi, O. *J. Am. Chem. Soc.* **1989**, *111*, 1547.
(29) Tucker, S. C.; Truhlar, D. G. *Chem. Phys. Lett.* **1989**, *157*, 164.
(30) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. *J. Am. Chem. Soc.* **1990**, *112*, 6127.
(31) Born, M. *Z. Phys.* **1920**, *1*, 45.
(32) Rashin, A. A.; Honig, B. *J. Phys. Chem.* **1985**, *89*, 5588.
(33) Liotard, D. A.; Hawkins, G. D.; Lynch, G. C.; Cramer, C. J.; Truhlar, D. G. *J. Comput. Chem.*, in press.

probably a reasonably physical solution in most cases,[34] it does have some obvious shortcomings. For example, it sometimes predicts a positive polarization energy, which is nonphysical. This occurs because favorable Coulomb interactions are not screened fully consistently with unfavorable Coulomb interactions. Fortunately, such positive values do not appear to be large, and we accept them in the same way as other inevitable modeling errors.

When developing a model, G°_{CDS} is determined by rearranging eq 8 to yield

$$G^{\circ}_{CDS} = \Delta G^{\circ}_{S} - \Delta G_{ENP} \tag{12}$$

where ΔG°_{S} is now a reference free energy of solvation calculated from experimental data. We then substitute eqs 2, 3, and 5 on the left-hand side and determine the σ_{γ}^{CD} and σ^{CS} values by a multilinear regression to a model-specific equation involving empirical surface tensions and solvent-accessible surface areas as defined by Lee and Richards.[11,12]

In this work, as in our previous solvation models, we neglect any contribution from a nonideal entropy of mixing. Although the ideal entropy of mixing approximation that is implicit in our treatment can be derived only in very special cases, e.g. for solute and solvent being isotopes[35] or at least having equal free volumes in a liquid mixture,[36] or for linear arrangements of long and short molecules,[35,37] no better treatment is available for the general case. There is evidence that the ideal-solution approximation for the entropy can provide a good approximation even for molecules of quite different size.[38,39] In our work, any nonideality of the entropy of mixing is included in the first-solvation-shell terms.[40] Additional experimental evidence relevant to the present investigation is provided by solutions of *n*-hexane and *n*-heptane in *n*-hexadecane; data for the former show only slight deviations from ideal solution behavior;[35,41,42] data for the latter show a more significant nonideality, although it is still small.[35,43] Finally, recent investigations have suggested that nonideal entropy of mixing contributions for a given solute *do* scale proportionally to the size of the first solvation shell;[40,44,45] G°_{CDS} thus has the proper functional form to absorb such effects into the parameterization.

2.2. SM4 Formalism for *n*-Hexadecane. The new SM4 model for *n*-hexadecane involves two major improvements over previous models; in particular, a new method is used to obtain partial charges that allows us to achieve a general parameter set that is transferable among Hamiltonians, and the form of G°_{CDS} has been redefined.

Incorporation of Charge Model 1. In the original SM*x* models,[2-5] partial charges were obtained from the semiempirical wave function using a zero-overlap Mulliken population analysis[46] of the AM1[19] or PM3[20] semiempirical molecular orbital wave functions. The Mulliken analysis method has been highly criticized in the literature, especially for extended basis sets. For the minimal basis set employed in AM1 and PM3, it yields charges that follow physically reasonable trends but that are not quantitatively accurate. In an effort to obtain more physical partial charges, we have developed a class IV charge method called Charge Model 1 (CM1).[18] Class I charges are those that are obtained directly from experiment by simple models. Class II charges are obtained directly from a quantum mechanical wave function, while class III charges are obtained from a physical observable that is predicted from the wave function, e.g. charges derived from electrostatic-potential-fitting. Finally, class IV charge methods transform class II or class III charges to more accurately predict the desired experimental property. Thus, the parameters in the CM1A and CM1P models were determined[18] to map the Mulliken charges obtained from the wave function computed using the AM1 and PM3 Fock operators[19,20] to charges that more accurately reproduce experimental gas-phase dipole moments.

A critical element of the CM1A and CM1P models is that the charges are linear functions of the one-particle density matrix elements, which are themselves bilinear functions of the molecular orbital coefficients. In particular, if molecular orbitals, ψ_{α}, are linear combinations of atomic orbitals, ϕ_{λ}, then

$$\psi_{\alpha} = \sum_{\lambda} c_{\alpha\lambda}\phi_{\lambda} \tag{13}$$

Then, for a closed-shell molecule treated by restricted Hartree–Fock theory with real atomic orbitals, the one-particle density matrix elements are

$$P_{\lambda\sigma} = 2\sum_{\alpha}^{occ} c_{\alpha\lambda}c_{\alpha\sigma} \tag{14}$$

The notation indicates a sum over occupied molecular orbitals. By obtaining charges as simple functions of the density matrix elements, we retain a reasonably high level of computational efficiency and numerical stability.

The CM1 models begin with the Mulliken charge $q_i^{(M)}$ on atom *i*, which, since the atomic orbitals are assumed orthogonal in NDDO methods, reduces to

$$q_i^{(M)} = Z_i - \sum_{\lambda\in i} P_{\lambda\lambda} \tag{15}$$

and they apply a semiempirical mapping in the form of

$$q_i = q_i^{(M)} + B_i[c_i q_i^{(M)} + d_i] - \sum_{i'\neq i} B_{ii'}[c_{i'} q_{i'}^{(M)} + d_{i'}] \tag{16}$$

where $B_{ii'}$ is the covalent bond index[47] (or, for short, the bond order) between atoms *i* and *i'*, B_i is the sum of all bond orders to atom *i*, and c_i and d_i are empirical scale factors and offsets respectively defined as

$$c_i = \hat{c}_i + \sum_{i'\neq i} f^{(c)}(B_{ii'})c_{ii'} \tag{17}$$

$$d_i = \hat{d}_i + \sum_{i'\neq i} f^{(d)}(B_{ii'})d_{ii'} \tag{18}$$

In eqs 17 and 18, $\hat{c}_i$ and $\hat{d}_i$ are parameters that depend only

(34) See, e.g.: Cramer, C. J.; Hawkins, G. D.; Truhlar, D. G. *J. Chem. Soc., Faraday Trans.* **1994**, *90*, 1802.
(35) Hildebrand, J. H.; Scott, R. L. *The Solubility of Nonelectrolytes*, 3rd ed.; Reinhold: New York, 1950; pp 17, 46, 106–118, 210–211, 452–453.
(36) Shinoda, K. *Principles of Solution and Solubility*; Marcel Dekker: New York, 1978; pp 111–116.
(37) Hildebrand, J. H. *J. Am. Chem. Soc.* **1937**, *59*, 794.
(38) Shinoda, K.; Hildebrand, J. H. *J. Phys. Chem.* **1958**, *62*, 481.
(39) Shinoda, K.; Hildebrand, J. H. *J. Phys. Chem.* **1961**, *65*, 1885.
(40) Giesen, D. J.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1994**, *98*, 4141.
(41) Brønsted, J. N.; Koefed, J. *K. Dan. Vidensk. Selsk.* **1946**, *22*, No. 17.
(42) Hildebrand, J. H.; Sweny, J. W. *J. Phys. Chem.* **1939**, *43*, 109, 297.
(43) van der Waals, J. H.; Hermans, J. J. *Recl. Trav. Chim. Pays-Bas* **1949**, *68*, 181.
(44) Rashin, A. A.; Bukatin, M. A. *J. Phys. Chem.* **1994**, *98*, 386.
(45) Lazaridis, T.; Paulaitis, M. E. *J. Phys. Chem.* **1994**, *98*, 635.
(46) Mulliken, R. S. *J. Chem. Phys.* **1955**, *23*, 1833.
(47) Armstrong, D. R.; Perkins, P. G.; Stewart, J. J. P. *J. Chem. Soc., Dalton Trans.* **1973**, 838.

on the atomic number of atom i, and $c_{ii'}$ and $d_{ii'}$ are parameters that depend on the atomic numbers of atoms i and i'. $f^{(c)}(B_{ii'})$ and $f^{(d)}(B_{ii'})$ are functions[18] of the covalent bond indices and are specified in the supplementary material. $B_{ii'}$ is defined as

$$B_{ii'} = \sum_{\lambda \in i} \sum_{\sigma \in i'} \left(\frac{1}{2}P_{\lambda\sigma} + \frac{1}{2}P_{\sigma\lambda}\right)^2 \tag{19}$$

In the present work, the orbitals are real and $P_{\lambda\sigma} = P_{\sigma\lambda}$. Nevertheless, we write the bond order in the manifestly symmetric form of eq 19 to insure that analytic derivatives of functions of the covalent bond indices with respect to density matrix elements are symmetric (i.e., the same for $B_{ii'}$ and $B_{i'i}$); this proves useful when taking the derivative of the CM1 charge with respect to the density matrix, as required below.

The second term in eq 16 is the mapping of the partial charge on atom i. The third term is a renormalization-like term that accounts for the change in charge of all other atoms bonded to i. This term is needed to preserve the overall charge of the system.[18]

The full specifications of the CM1A and CM1P charge models are provided elsewhere.[18] We simply note here that the parameters in eqs 17 and 18 were obtained by least squares fitting to experimental dipole moments and high-level ab initio partial charges for 195 test molecules. The parameters in the charge models are different for CM1A and CM1P, but we will obtain a single set of solvation model parameters that may be used with either set of resulting charges.

In order to determine the molecular orbitals self-consistently in the presence of a reaction field, that field must enter into the SCF equations. Modification of the Fock operator (relative to the gas phase) will result in a new secular equation for the molecular orbital coefficients. Different coefficients imply a different density matrix. Of course, that different density matrix then results in a change in the Fock matrix, and the whole procedure must be repeated until convergence is reached, just as in any SCF strategy. The point, however, is that the density matrix for the solvated solute will be different from that for the gas-phase solute. The question, then, is how to form the Fock operator.

In SCRF theory,[1,10] we have the following general equation

$$\mathbf{F} = \frac{\partial G}{\partial \mathbf{P}} \tag{20}$$

where $\mathbf{F}$ is the Fock matrix, G is the energy functional (which is the sum of the internal electronic and internuclear Coulomic energy of the solute and the free energy of the solvent), and $\mathbf{P}$ is the density matrix. For a gas-phase restricted Hartree–Fock calculation in which core electrons are not treated explicitly

$$G = \frac{1}{2}\sum_{\mu\nu} P_{\mu\nu}(h_{\mu\nu} + F^{(0)}_{\mu\nu}) + \sum_{i<j} \frac{Z_i Z_j}{r_{ij}} \tag{21}$$

where μ and ν run over the set of valence atomic orbital basis functions, Z_i is the nuclear charge of atom i minus its number of core electrons, and $P_{\mu\nu}$ is the density matrix formed from the coefficients of the molecular orbitals in the usual fashion, as in eqs 13 and 14. Matrix elements for the one-electron Hamiltonian $h_{\mu\nu}$ are defined by

$$h_{\mu\nu} = \left\langle \mu \left| -\frac{1}{2}\nabla^2 \right| \nu \right\rangle - \sum_i \left\langle \mu \left| \frac{Z_i}{r_i} \right| \nu \right\rangle \tag{22}$$

where $\langle f|g|h\rangle$ indicates an integral with the Dirac bra-ket notation and r_i is the distance to nucleus i. Matrix elements for the gas-phase Fock operator are defined by[48]

$$F^{(0)}_{\mu\nu} = h_{\mu\nu} + \sum_{\lambda\sigma} P_{\lambda\sigma}\left(\left\langle \mu(1)\lambda(2)\left|\frac{1}{r_{12}}\right|\nu(1)\sigma(2)\right\rangle - \frac{1}{2}\left\langle \mu(1)\lambda(2)\left|\frac{1}{r_{12}}\right|\sigma(1)\nu(2)\right\rangle\right) \tag{23}$$

where the integrals are the usual Coulomb and exchange integrals, respectively. It is easily verified that eqs 20–23 are consistent, i.e. the Fock matrix is indeed the partial derivative of the energy functional with respect to the density matrix.

Within the generalized Born approximation, the energy functional for a solute dissolved in a liquid solvent may be written as

$$G_{\mathrm{ENP}} = \frac{1}{2}\sum_{\mu\nu} P'_{\mu\nu}(h_{\mu\nu} + F^{(0)'}_{\mu\nu}) + \sum_{i<j} \frac{Z_i Z_j}{r_{ij}} - \frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{k,k'} q_k q_{k'} \gamma_{kk'} \tag{24}$$

where P' is the relaxed (with respect to solvation) density matrix, $F^{(0)'}$ is the Fock matrix defined by eq 23 but formed using the relaxed density matrix, and the final term accounts for electric polarization using the generalized Born formalism. Note that the first term continues to represent simply the electronic energy of the solute, including electron–nuclear attraction, although it will generally be higher in energy than was the corresponding gas-phase term since the density matrix has been changed from the gas-phase optimum. The gas-phase nuclear repulsion term is unchanged; the dielectric screening of nuclear repulsion is, however, one of the effects included in the third term of eq 24.

In order to arrive at the relaxed density matrix, we must solve for the orbitals with a proper Fock matrix as defined by eq 20. Doing so requires us to take the partial derivative of eq 24 with respect to the density matrix. This yields

$$F^{(1)'}_{\mu\nu} = \frac{\partial G_{\mathrm{ENP}}}{\partial P'_{\mu\nu}} = F^{(0)'}_{\mu\nu} - \left(1 - \frac{1}{\epsilon}\right)\sum_{k,k'} \frac{q_k}{\gamma_{kk'}}\left(\frac{\partial q_{k'}}{\partial P'_{\mu\nu}}\right) - \frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\sum_{k,k'} q_k q_{k'} \frac{\partial \gamma_{kk'}}{\partial P'_{\mu\nu}} \tag{25}$$

We emphasize that $F^{(1)'}_{\mu\nu}$ is required for solution of the SCF equations but $F^{(0)'}_{\mu\nu}$ is used in the calculation of G_{ENP} according to eq 24.

Appendix A (in the supplementary material) contains further information on forming the Fock matrix with the new charge models.

Our solvation model is directly dependent on the underlying Hamiltonian in two ways. First, the partial charges resulting from an analysis of the wave function are used to calculate the polarization energy and the new Fock matrix according to eqs 9 and 25. Second, the geometry of the molecule determines the solvent-accessible surface areas that are used in eqs 3 and 5. Thus, previously,[3–5] new parameters were required for use

(48) Roothaan, C. C. J. *Rev. Mod. Phys.* **1953**, *23*, 61.

Table 1. SM4 Results for Selected Compounds using AM1 and PM3[a]

	AM1-SM4					PM3-SM4					
	ΔG_{ENP}	G_{CD}	G_{CS}	G^{o}_{CDS}	ΔG^{o}_{S}(SM4)	ΔG_{ENP}	G_{CD}	G_{CS}	G^{o}_{CDS}	ΔG^{o}_{S}(SM4)	ΔG^{o}_{S}(expt)[b]
methane	−0.02	−10.21	10.36	0.15	0.13	0.00	−10.24	10.30	0.07	0.07	0.45[c]
ethane	−0.03	−12.20	11.68	−0.52	−0.55	−0.01	−12.26	11.64	−0.62	−0.63	−0.67[d]
cyclopropane	−0.17	−13.43	12.18	−1.25	−1.42	−0.08	−13.53	12.17	−1.36	−1.44	−1.78[e]
n-hexane	0.06	−19.21	15.70	−3.52	−3.46	0.02	−19.34	15.68	−3.67	−3.65	−3.64[d]
2-methylpentane	0.06	−18.49	15.16	−3.33	−3.27	0.02	−18.53	15.11	−3.42	−3.40	−3.45[e]
cyclohexane	0.02	−17.01	14.37	−2.63	−2.61	0.01	−17.12	14.36	−2.76	−2.75	−4.04[d]
n-hexadecane	0.40	−36.62	25.72	−10.90	−10.50	0.15	−36.94	25.72	−11.21	−11.06	−10.52[d]
1-pentene	−0.14	−17.38	14.44	−2.94	−3.08	−0.09	−17.48	14.42	−3.06	−3.15	−2.79[d]
1-butyne	−0.94	−14.53	13.20	−1.33	−2.27	−0.74	−14.59	13.18	−1.41	−2.15	−2.07[d]
anthracene	−2.23	−25.74	17.77	−7.97	−10.20	−1.40	−25.78	17.74	−8.04	−9.44	−10.32[f]
ethanol	−1.43	−12.91	12.23	−0.68	−2.11	−1.32	−12.96	12.20	−0.75	−2.07	−2.03[d]
1-decanol	−1.11	−26.84	20.25	−6.59	−7.70	−1.16	−27.03	20.24	−6.79	−7.95	−7.68[d]
tert-butyl alcohol	−0.93	−15.90	13.76	−2.14	−3.07	−0.92	−15.97	13.76	−2.21	−3.13	−2.74[d]
butanal	−1.53	−15.56	13.82	−1.74	−3.27	−1.61	−15.65	13.82	−1.83	−3.44	−3.10[d]
2-butanone	−1.62	−15.57	13.82	−1.76	−3.38	−1.63	−15.67	13.81	−1.87	−3.50	−3.12[d]
methyl propanoate	−0.76	−16.69	14.52	−2.17	−2.93	−0.28	−16.83	14.51	−2.32	−2.60	−2.68[c]
pentyl acetate	−0.80	−22.18	17.67	−4.50	−5.30	−0.48	−22.38	17.65	−4.72	−5.20	−5.24[d]
ethyl octadecanoate	−0.15	−44.89	30.64	−14.24	−14.39	0.03	−45.33	30.68	−14.65	−14.62	−13.69[g]
diethyl ether	−0.67	−17.15	14.53	−2.62	−3.29	−0.50	−17.25	14.51	−2.74	−3.24	−2.75[d]
tetrahydrofuran	−1.02	−14.96	13.40	−1.56	−2.58	−0.77	−15.07	13.38	−1.69	−2.46	−3.60[d]
trimethylamine	−0.63	−15.07	13.28	−1.79	−2.42	−0.23	−15.21	13.28	−1.93	−2.16	−2.21[d]
n-methylaniline	−1.72	−19.67	15.26	−4.41	−6.13	−1.25	−19.74	15.24	−4.50	−5.75	−6.19[e]
acetonitrile	−2.45	−11.62	11.71	0.09	−2.36	−2.72	−11.67	11.68	0.01	−2.71	−2.37[d]
4-methylpyridine	−1.83	−17.56	14.39	−3.18	−5.01	−1.39	−17.64	14.36	−3.29	−4.68	−4.89[c]
fluorobenzene	−1.07	−16.67	13.94	−2.73	−3.80	−0.79	−16.70	13.92	−2.78	−3.57	−3.80[d]
dichloromethane	−0.83	−14.15	12.23	−1.92	−2.75	−0.87	−14.22	12.22	−1.99	−2.86	−2.75[d]
(*E*)-1,2-chloroethene	−0.37	−15.81	13.00	−2.81	−3.18	−0.35	−15.80	12.97	−2.83	−3.18	−3.11[d]
chlorobenzene	−1.05	−18.70	14.40	−4.30	−5.35	−0.67	−18.71	14.37	−4.34	−5.01	−4.99[d]
1,2-dibromoethane	−0.72	−17.81	13.76	−4.05	−4.77	−0.71	−17.71	13.67	−4.04	−4.75	−4.62[d]
iodoethane	−0.64	−15.94	12.99	−2.95	−3.59	−0.55	−15.80	12.88	−2.92	−3.47	−3.51[d]
ethanethiol	−0.70	−14.88	12.73	−2.15	−2.85	−0.55	−15.04	12.75	−2.29	−2.84	−2.96[d]
dimethyl sulfide	−0.69	−14.74	12.73	−2.01	−2.70	−0.76	−14.91	12.77	−2.15	−2.91	−3.05[d]
water	−1.67	−8.23	9.50	1.27	−0.40	−1.52	−8.26	9.50	1.24	−0.28	−0.35[d]
ammonia	−2.52	−9.36	9.86	0.50	−2.02	−1.31	−9.55	9.86	0.31	−1.00	−0.93[d]
tetramethylsilane	−0.26	−18.56	14.89	−3.67	−3.93	−0.02	−18.91	14.97	−3.93	−3.95	−2.92[c]
tetraethylsilane	−0.12	−22.77	17.20	−5.57	−5.69	−0.52	−23.17	17.31	−5.87	−6.38	−5.86[c]
tetramethylstannane[h]	−0.39	−19.99	15.54	−4.46	−4.85	na	na	na	na	na	−3.98[c]
tetraethylstannane[h]	−0.17	−24.37	18.01	−6.36	−6.53	na	na	na	na	na	−6.93[c]
helium[i]	0.00	−7.47	9.85	2.38	2.38	0.00	−7.47	9.85	2.38	2.38	2.38[c]
neon[i]	0.00	−7.66	9.82	2.16	2.16	0.00	−7.66	9.82	2.16	2.16	2.16[c]
argon[i]	0.00	−9.02	9.97	0.95	0.95	0.00	−9.02	9.97	0.95	0.95	0.95[c]

[a] Free energies are in kcal/mol. [b] Letter indicates the reference from which the experimental data was taken. [c] Reference 9. [d] Reference 54. [e] Reference 52. [f] Reference 53. [g] Reference 55. [h] Parameters for tin are not available with PM3. [i] Calculated without use of either AM1 or PM3.

with each new semiempirical Hamiltonian. This was necessary for the most part because differences in the underlying wave functions caused the electrostatic interactions calculated from the Mulliken partial charges to differ markedly between the AM1 and PM3 models. However, both the CM1A and CM1P models[18] give partial charges that are consistent with high-level *ab initio* electrostatic-potential-fitting[49,50] charges. The inclusion of these models causes electrostatic interactions calculated using either AM1 or PM3 to be quite similar (see Table 1). Because of this, and because of the fact that the CDS term is fairly insensitive to small changes in the molecular geometry (again see Table 1), we can for the first time create a set of parameters that are Hamiltonian-independent. As a consequence, the parameters presented in this paper were developed for use with either the AM1 or the PM3 Hamiltonians. Indeed, because the parameters were developed using partial charges that agree with high-level *ab initio* calculations and also using reasonable geometries, they should be valid for use with any Hamiltonian that is accurate enough to yield sufficiently reasonable partial charges and geometries. In this paper, we develop and apply the SM4 model parameters using the AM1 and PM3 Hamiltonians.

New Definition of G^{o}_{CDS}. In SM2,[3] SM2.1,[33] and SM3,[4] the surface tension of a non-hydrogenic atom depends on the bond order to hydrogen for that atom and hydrogen is given a zero surface area. This allows us to treat each hydrogen as part of a group and thus to account for the fact that the properties of hydrogen change depending on the atom to which it is bonded. In the present parameterization, hydrogenic solvent-accessible surface area is calculated as for any other atom but the hydrogen CD surface tension varies according to the type of atom that it is bonded to; nonhydrogenic heavy atoms have a constant surface tension. In addition, a bond-order-dependent surface tension was added for carbon bonded to carbon in order to fit the alkyne experimental data. This gives us an equation for G^{o}_{CDS} in the SM4 formalism for *n*-hexadecane as

$$G^{o}_{CDS} = \sigma^{CS}\sum_{k} A_k^{CS}(\beta_k^{CS},\{\beta_{k'}^{CS}\}) + \sum_{k} \hat{\sigma}_k^{CD} A_k^{CD}(\beta_k^{CD},\{\beta_{k'}^{CD}\}) \quad (26)$$

where σ_k^{CD} takes the form

(49) Chirlian, L. E.; Francl, M. M. *J. Comput. Chem.* **1987**, *8*, 894.
(50) Breneman, C. M.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361.

$$\hat{\sigma}_k^{CD} = \begin{cases} \sigma_k^{CD} & k \neq H, C \\ \sum_{k'} \sigma_{k'H}^{CD} B_{kk'} & k = H \\ \sigma_k^{CD} + \sum_{k'=C} \sigma_{CC}^{CD} B_{kk'}{}^3 & k = C \end{cases} \quad (27)$$

and $\beta_k{}^X$ is defined as

$$\beta_k^X = R_k + R_S^X \quad (28)$$

where X can be either CD or CS. R_k is the radius of atom k and is the same for CD and CS terms. Equation 27 allows the surface tension of a hydrogen to gradually change during a proton or hydride transfer in which the hydrogen environment is also changing. This overcomes one of the shortcomings in SM1a,[2] a model that, while accurate, was limited in its utility due to its inability to allow environmentally specific surface tensions to change continuously during a reaction. Also, by explicitly including the bond order calculated from the density matrix, the atomic environments need not be specified.

3. Parameterization for *n*-Hexadecane

Method. Throughout the parameterization step, all geometries are optimized in the presence of the solvent. In calculating the effective Coulomb radii, α_k, from the intrinsic Coulomb radii, ρ_k, the radial integrals of the dielectric screening algorithm were all well converged, using the force trapezoid algorithm of ref 33.

First we divide the parameters into two groups: those given standard values or taken from previous models and those optimized in the present work. The former set includes all $q_k^{(1)}$ and some values of R_k, $\rho_k^{(0)}$, $\rho_k^{(1)}$, and $q_k^{(0)}$. As has been done in all previous models,[2-5] $q_k^{(1)}$ was set to 0.1. All noncarbon R_k values are from Bondi;[51] the value for carbon was set to 1.60 Å. With the exception of C, N, and O, $\rho_k^{(0)}$ remains the same value as in SM2.1.[33] For H, $\rho_k^{(1)}$ and $q_k^{(0)}$ remain the same as in SM2.1;[33] for O, these two parameters have been reoptimized. All other $\rho_k^{(1)}$ have been set to zero, and all other $q_k^{(0)}$ have been removed and are no longer parameters. In SM2, SM2.1, and SM3, $d_{kk'}^{(1)}$ in eq 10 is taken equal to zero unless k and k' correspond to N and H or to O and O. In the present parameterization we found that the former was unnecessary. We make $d_{kk'}^{(1)}$ nonzero only for O−O interactions. The parameters used have the same value as in all previous models.[2-5] (This means that only bonded and geminal O−O interactions are affected. The fact that such interactions are not treated well in any of our models or in any solvent tends to indicate that the problem is with the underlying wave functions.) For completeness, R_k, $\rho_k^{(0)}$, $\rho_k^{(1)}$, $q_k^{(0)}$, $d_{kk'}^{(1)}$, $d_{kk'}^{(2)}$, $r_{kk'}^{(1)}$, and $r_{kk'}^{(2)}$ are given in Appendix B of the supplementary material. The only remaining parameters are the surface tensions, and these were obtained by linear regression using eqs 12, 26, and 27.

Reference data for ΔG_S° are required for the linear regression using eq 12. In parameterizing the present model, we have taken data from five sources.[9,52-55] A test set of 153 compounds was selected to develop the surface tensions. Each compound was computed using both the AM1 and PM3 Hamiltonians, creating a total of 306 data points in the parameterization.

(51) Bondi, A. *J. Phys. Chem.* **1964**, *68*, 441.
(52) Zhang, Y.; Dallas, A. J.; Carr, P. W. *J. Chromatogr.* **1993**, *638*, 43.
(53) Abraham, M. H. *J. Chromatogr.* **1993**, *644*, 95.
(54) Abraham, M. H. *Chem. Soc. Rev.* **1993**, *22*, 73.
(55) Abraham, M. H. Personal communication.

Table 2. Surface Tensions for *n*-Hexadecane SM4

atom	σ_k^{CD}, cal mol^{-1} Å^{-2}	$\sigma_{k'k}^{CD}$,[a] cal mol^{-1} Å^{-2}	σ_k^{CS}, cal mol^{-1} Å^{-2}
H	*b*	*b*	17.25
He	−40.31	*b*	17.25
C	−76.80	−50.21	17.25
N	−36.29	−56.29	17.25
O[c]	−39.18	−56.13	17.25
F	−37.42	*b*	17.25
Ne	−41.55	*b*	17.25
S	−59.77	−54.64	17.25
Cl	−55.35	*b*	17.25
Ar	−47.66	*b*	17.25
Br	−59.79	*b*	17.25
I	−62.58	*b*	17.25
C bonded to C	*b*	1.19	*b*

[a] Unless otherwise indicted, is $\sigma_{k'H}^{CD}$. [b] Not a parameter. [c] Not recommended for nitro compounds. See text.

We did not include any carboxylic acids in our data set because the experimental data may be unreliable for such compounds. This is discussed further below.

Results. The surface tensions are obtained by considering eqs 2, 3, 5, and 8. In particular, they are obtained by minimizing the rms error in ΔG_S° calculated from

$$\Delta G_S^\circ = \Delta G_{ENP} + \sum_\gamma [\hat{\sigma}_\gamma^{CD} A_\gamma(R_\gamma^{CD}) + \sigma^{CS} A_\gamma(R_\gamma^{CS})] \quad (29)$$

with respect to σ^{CS} and to the σ_k^{CD} and $\sigma_{kk'}^{CD}$ parameters in $\hat{\sigma}_\gamma^{CD}$ (see eq 27). Since the unknown parameters all occur linearly at this stage, this step is accomplished noniteratively for a given set of geometries. The geometries were then reoptimized in solution with the new parameters, and this cycle was repeated until the geometries and the parameters were simultaneously converged. Note that ΔG_{ENP} depends on whether we use the AM1 or PM3 method. The rms error for a preliminary fit to the 153 data points based on the AM1 Hamiltonian is 0.39 kcal/mol, with a mean unsigned error of 0.28 kcal/mol and a mean signed error of 0.00 kcal/mol. A preliminary fit to the 153 data points that use the PM3 Hamiltonian yields an rms error of 0.40 kcal/mol, a mean unsigned error of 0.29 kcal/mol, and a mean signed error of 0.00 kcal/mol. Not only are the rms and mean errors similar, the parameters obtained this way are also very similar. Thus, for the final fit, we combined these data points into a single set of 306 points. The resulting surface tensions are given in Table 2. Results for 41 representative solutes are given in Table 1 for discussion purposes. Results for the entire test set of compounds are given in Appendix C, which is contained in the supplementary material. The rms error over the 306 AM1 and PM3 data points is 0.41 kcal/mol. The mean unsigned error is 0.29 kcal/mol, and the mean signed error is 0.00 kcal/mol. Using the final fit, the rms error for the 153 data points that use the AM1 Hamiltonian is 0.40 kcal/mol, with a mean unsigned error of 0.29 kcal/mol and a mean signed error of −0.03 kcal/mol. The rms error for the 153 data points that use the PM3 Hamiltonian is 0.42 kcal/mol, while the mean unsigned and signed errors are 0.29 and 0.03 kcal/mol, respectively. For comparison, we note that the dispersion of the reference data is 2.05 kcal/mol, where the dispersion is calculated as

$$D = \left[\overline{(\Delta G_{ref}^\circ)^2} - \overline{(\Delta G_{ref}^\circ)}^2\right]^{1/2} \quad (30)$$

4. Discussion

Polarization. A common misconception is that polarization is not very important in nonpolar solvents. However, an

Table 3. Free Energies of Solvation (kcal/mol) and Partial Charges on Oxygen for Diethyl Ether and Butanal using PM3

	solute relaxation?	diethyl ether					butanal				
		ΔG_S°	ΔG_{ENP}	G_P	$G_P(O)$	$q(O)$	ΔG_S°	ΔG_{ENP}	G_P	$G_P(O)$	$q(O)$
expt		−2.75					−3.10				
SM4	none	−3.22	−0.47	−0.47	−0.69	−0.36	−3.23	−1.43	−1.43	−1.54	−0.41
SM4	partial[a]	−3.24	−0.50	−0.53	−0.75	−0.37	−3.42	−1.59	−1.79	−1.87	−0.45
SM4	full	−3.24	−0.50	−0.54	−0.76	−0.37	−3.44	−1.61	−1.81	−1.89	−0.45

[a] Electronic relaxation permitted but geometry frozen at the gas-phase minimum.

examination of eq 9 reveals that even a low dielectric solvent such as *n*-hexadecane (for which $\epsilon = 2.06$[6]) yields slightly more than one-half of the polarization energy that is obtained in water ($\epsilon = 78.3$[6]) for a fixed set of charges. A comparison between diethyl ether and butanal as calculated using our solvation model with the PM3 Hamiltonian illustrates this important point. Experimentally, butanal is solvated 0.4 kcal/mol better than diethyl ether. Both compounds are slightly oversolvated by our model compared to experiment (Table 1). If we only examine G_{CDS}°, it would appear that diethyl ether should be the better solvated compound for two reasons. First, diethyl ether has a larger surface area than butanal—346 Å^2 compared to 322 Å^2 when calculated using R_S^{CD}. The two additional (favorably solvated) hydrogens in diethyl ether should lead to a more negative free energy of solvation. In addition, the oxygen is buried in the ether, while in the aldehyde, it is quite exposed. Because the surface tension for oxygen is less negative than for carbon or for hydrogen, it is energetically more favorable for the oxygen to be buried and the hydrogen and carbon to be exposed. However, both experiment and our model show that butanal is better solvated than diethyl ether. Therefore, polarization must play an important role in the solvation of the aldehyde. Indeed, as shown in Table 3, G_P values are −1.81 kcal/mol for butanal and −0.54 kcal/mol for diethyl ether. Using these values, ΔG_{ENP} values are −1.61 kcal/mol for the aldehyde and −0.50 kcal/mol for the ether. We can assign the polarization energy contribution of a single atom, *k*, by[56]

$$G_P(k) = -\frac{1}{2}\left(1 - \frac{1}{\epsilon}\right)\left(q_k^2\gamma_{kk} + q_k \sum_{k' \neq k} q_{k'}\gamma_{kk'}\right) \qquad (31)$$

Note that this equation includes the diagonal term for atom *k* and one-half the sum of the interaction energies of all atoms, *k'*, with atom *k*. Using eq 31, we can calculate the polarization energy associated with the oxygen in each compound. Specifically, oxygen accounts for −1.89 kcal/mol polarization energy in the aldehyde and −0.76 kcal/mol in the ether. Upon dissolution, the charge on the oxygen increases from −0.41 to −0.45 in the aldehyde and from −0.36 to −0.37 in the ether. Thus, even in *n*-hexadecane, it is important to allow the wave function to polarize. If we do not allow the wave function to polarize, i.e. if the solution-phase wave function remains the same as the gas-phase wave function, G_P values are −1.43 kcal/mol for the aldehyde and −0.47 kcal/mol for the ether. The oxygen contributions are −1.54 and −0.69 kcal/mol, respectively. Evidently, polarization is approximately 10 times more important for the aldehyde than for the ether.

Surface Tensions. Figure 2 is a plot of G_{CDS}° as calculated by eq 12 for all 306 data points in the test set versus solvent-accessible surface area calculated with a solvent radius of 2.0 Å. A similar plot results from using a larger or smaller solvent radius provided the radius is in a physically reasonable range. As can be seen, G_{CDS}° when calculated in this manner does not

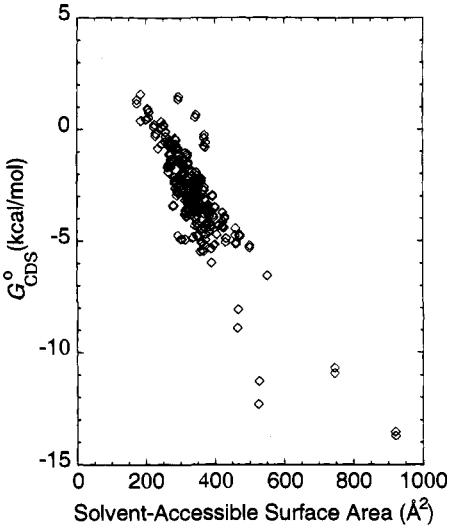

Figure 2. Plot of G_{CDS}° calculated by eq 12 versus total solvent-accessible surface area as calculated with a solvent radius of 2.0 Å.

tend toward zero when *extrapolated* from molecular data to zero solvent-accessible surface area. In fact, it appears that a linear equation for G_{CDS}° with a single solvent radius, such as has been used in our previous models,[2-5] would require an intercept of about +4.4 kcal/mol to properly fit the data. Although some workers have advocated the use of a non-zero intercept for surface tension terms,[57-59] this creates problems. If an intercept were to be added to eq 26, a molecule with zero surface area would have a CDS energy associated with it. Moreover, non-zero intercepts are not size extensive; in the case where the number of entities changes during the calculation of ΔG_S° (e.g., along a reaction path for dissociation or association), an intercept causes a nonphysical change in the total free energy of the system. We also considered adding a term that was proportional to the square root[58] of the solvent-accessible surface area. This suffers from much the same problem as the intercept; two separate fragments would have a different CDS energy if treated as dissociated than if treated as one weakly interacting fragment (supermolecule approach) with the same total solvent-accessible surface area. This is why the separation of short-range and long-range forces is a critical improvement in this model. As shown in Figure 3, this allows us to accurately calculate G_{CDS}° with two linear parameters that preserve size extensiveness. Figure 3 also shows that there is no size-dependent error in the model; that is, larger molecules are calculated as accurately as smaller molecules, at least over the size range represented in the test set.

In preliminary work, we considered various values for the CS solvent radius; in particular, we experimented with numerous large radii up to 10 Å. The rms error, to within 0.05 kcal/mol, was fairly independent of the actual radius for values larger

(56) Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1992**, *198*, 74; **1993**, *202*, 567(E).

(57) Sharp, K. A.; Nicholls, A.; Friedman, R.; Honig, B. *Biochemistry* **1991**, *30*, 9686.

(58) Hermann, R. B. *J. Comput. Chem.* **1993**, *14*, 741.

(59) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978.

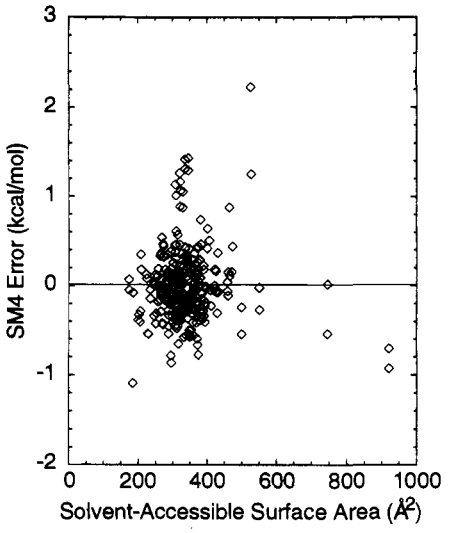

Figure 3. Plot of the error in the SM4 *n*-hexadecane model versus the total solvent-accessible surface area as calculated with a solvent radius of 2.0 Å.

than 4 Å. We have chosen the radius of a sphere that contains the same volume as the solvent molecule as calculated from the density of the pure substance at 298 K.[17]

Examination of G°_{CDS}. Abraham and co-workers[9] provided values for free energies of transfer from gas to *n*-hexadecane for the noble gases. Even though neither AM1 nor PM3 contain any parameters for these elements, we can calculate their free energies of solvation using our model. In particular, ΔG°_{S} must equal G°_{CDS} for these solutes as ΔG_{ENP} is identically zero within the present model for these compounds. This yields

$$\Delta G^{\circ}_{S(expt)}(He,Ne,Ar) = G_{CS} + G_{CD} \qquad (32)$$

Furthermore, since the only semiempirical parameter, σ^{CS}, in G°_{CS} is already determined, we can immediately calculate G°_{CS} (given the rare gas radius). This yields G°_{CD} by substitution into eq 32, and then eq 3 yields a σ_k^{CD} value for each rare gas. We shall call G°_{CD} calculated in this manner G°_{CD}(SM4), and the resulting σ_k^{CD} values are listed in Table 2. Because these gases were guaranteed to come out with the correct energy, they were not included in the error calculations for the model.

The above calculations on rare gases might seem to be a fruitless exercise, but they actually provide insight into the physical reasonableness of the present scheme. This is because we can estimate G°_{CD} independently by combining previous estimates[14] of the positive entropic contributions to cavity formation with estimates of the dispersion contribution from molecular beam interaction well depths.[16,60,61] As stated in the introduction, these two terms—dispersion and the positive entropic contribution to cavity formation—are considered to be the main negative free energy contributors to G°_{CD}. We shall refer to G°_{CD} estimated as the sum of these two terms as G°_{CD}(est). A comparison of G°_{CD}(SM4) and G°_{CD}(est) allows us to check if the partition obtained by using the experimental ΔG°_{S} and our universal σ^{CS} agrees with the independent estimate.

For the calculation of G_{CD}(SM4), the only solute-dependent parameter needed is R_k. For R_k, we used the molecular beam data of Buck *et al.*[16,60,61] Their data give the equilibrium distance between a noble gas and a methane molecule for He, Ne, and Ar. According to ref 16, in van der Waals systems that contain a noble gas, it is common for the more polarizable partner to determine the equilibrium interaction distance, while the type of noble gas determines the interaction well depth. Using a methane radius of 2.0 Å taken from another source,[15] we determined the effective radii of He, Ne, and Ar in an alkane solution to be 1.84, 1.83, and 1.88 Å, respectively. With R_k determined, we may now calculate G_{CD}(SM4), which is given in Table 4 for each gas. We now consider the estimation of G_{CD}(est). To a first approximation, we consider dispersion to be a sum of pairwise energies of interaction for the noble gas atom with individual methane molecules located at the pairwise equilibrium separation. For the three noble gases studied here,[16,60,61] the pairwise energies of interaction are −0.05, −0.11, and −0.33 kcal/mol for He, Ne, and Ar, respectively. In solution, a noble gas atom will be surrounded by more than one methane. Many spherical or nearly spherical nonassociating liquids pack so that each molecule tends to have about 12 nearest neighbors.[62] Furthermore, as calculated from the density at 298 K, the volume of a *n*-hexadecane molecule is 489 Å^3.[17] Multiplication by 0.75 gives 366 Å^3 as the approximate volume of 12 methylene groups, implying that, in the pure liquid, a sphere of 4.4 Å completely contains, on average, about 12 methylene groups. A sphere of this size is consistent with the sums of the noble gas radii, R_k and R_S^{CD}. In addition, the methylene groups are probably more polarizable than methane since they are part of a larger molecule; this would lead to larger favorable dispersion terms. Nevertheless, we will use the energetics of a methane−noble gas dispersion interaction for this rough approximation. Finally, we are able to calculate the positive entropic contribution to cavity formation from information in ref 14.

Table 4. Comparison of Dispersive and Positive Entropic Solvation Contributions As Calculated by SM4 *n*-Hexadecane and in Section 4[a]

	1.[b] dispersion	2.[c] $-T\Delta S_{(Cavity)}$	G_{CD}(est)[d]	G_{CD}(SM4)[e]
argon	−4.0	−5.3	−9.3	−9.0
neon	−1.3	−5.0	−6.3	−7.7
helium	−0.6	−4.3	−4.9	−7.5
methane	−4.8	−5.3	−10.1	−10.2

[a] All energies are given in kcal/mol. [b] Calculated from data in refs 16, 60, and 61. [c] Entropy of cavity formation calculated from ref 14. T = 298 K. [d] Sum of 1 and 2. [e] This work.

G_{CD}(SM4) for helium is −7.5 kcal/mol. The entropic contribution from cavity formation is −4.3 kcal/mol, 12 dispersion interactions total −0.6 kcal/mol, and the sum of these two, G_{CD}(est), is −4.9 kcal/mol. This number is low compared to G_{CD}(SM4), but the agreement is not bad for such a small, quantum mechanically complex solute. For argon, G_{CD}(SM4) is −9.0 kcal/mol. As shown in Table 4, argon cavity formation contributes −5.3 kcal/mol and 12 argon-methane dispersion interactions contribute −4.0 kcal/mol to G_{CD}(est). Thus, for argon, the value of G_{CD}(est), −9.3, is in quite good agreement with G_{CD}(SM4). It is interesting to look at an approximately spherical molecule, such as methane, that we can calculate using AM1. For methane, G_{CD}(SM4) is −10.2 kcal/mol using either the AM1 or PM3 Hamiltonian. If we assume roughly an argon-sized cavity for methane, the cavity

(60) Buck, U.; Kohl, K. H.; Kohlhase, A.; Faubel, M.; Staemmler, V. *Mol. Phys.* **1985**, *55*, 1255.
(61) Buck, U.; Schleusuer, J.; Malik, D. J.; Secrest, D. *J. Chem. Phys.* **1981**, *74*, 1707.
(62) Chandler, D. *Introduction to Modern Statistical Mechanics*; Oxford University Press: New York, NY, 1987; pp 214−218.

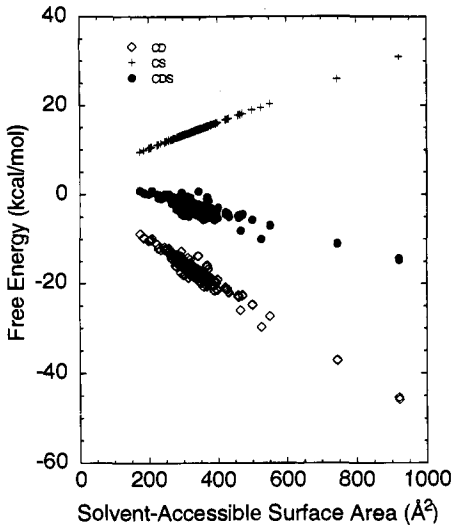

Figure 4. Examination of the magnitudes of the G°_{CDS}, G_{CD}, and G_{CS} terms as functions of the solvent-accessible surface area of the molecule calculated using R_S^{CD}. Note that the G_{CS} term is actually calculated using the CS surface area, resulting in an *apparent* non-zero intercept at zero solvent-accessible surface area for G_{CS} in this plot.

entropy term is equal to −5.3 kcal/mol. From ref 15, it appears that the pairwise energy of interaction for two methane molecules is about −0.4 kcal/mol. Dispersion is then worth −4.8 kcal/mol, making G_{CD}(est) equal to −10.1 kcal/mol. Again, this can be considered very good agreement considering the crudeness of this comparison and the fact that we are ignoring any other negative contributions. A summary of all results from this comparison can be found in Table 4.

It can be argued that because all molecules in the previous comparison form roughly the same sized cavity in *n*-hexadecane that we have not truly tested the limits of this model for larger solutes. Unfortunately, the above comparison cannot be done in the same manner for larger or nonspherical solutes as ref 14 does not contain data for these types of molecules that would allow us to calculate the entropic cavity-formation contribution to G_{CD}(est). However, Figure 4 examines how the magnitudes of the CD and CS terms vary as a function of the size of the molecule. As an example, we will compare the results of our model for ethane and *n*-hexadecane using the AM1 Hamiltonian (as shown in Table 1, nearly identical results are obtained using the PM3 Hamiltonian), two compounds listed in Table 1. First we should note that the CD solvent-accessible surface area for *n*-hexadecane is 3.0 times greater than that of ethane, while the CS surface areas differ by a factor of 2.2. Because of this, using AM1, the more size-sensitive CD term changes from −12.2 to −36.6 kcal/mol on going from ethane to *n*-hexadecane, while the CS term only changes from +11.7 to +25.7 kcal/mol. The net result is that *n*-hexadecane is calculated to be 9.9 kcal/mol better solvated than ethane. Experimentally, *n*-hexadecane is 9.9 kcal/mol better solvated than ethane, demonstrating once again that the model does not have any size-dependent errors.

Other Group IV Elements. Although the compilation of Abraham and co-workers[9] contains data for silicon and tin, we did not fully parameterize these elements. To add a new element to the model, we need R_k, $\varrho_k^{(0)}$, $\varrho_k^{(1)}$, and $q_k^{(0)}$ ($q_k^{(1)}$ is always set at 0.1). In this model, R_k is set to the Bondi radius[51] for that element. The Bondi radius of silicon is 2.10 Å. However, Bondi does not list a radius for tin.[51] Using the value of the van der Waals radius as estimated by Pauling,[63] ref 51 does estimate that the van der Waals radius of tin is 0.2 Å greater than that of silicon. This agrees with ref 64, in which the atomic, covalent, and 4− ion radii of tin are listed to be 0.2 Å greater than the respective silicon radii.[64] Thus, we chose R_k for tin to be 2.30 Å. The value of the Coulomb radius, $\varrho_k^{(0)}$, for silicon was determined by considering the difference in R_k for carbon and silicon. Because R_k for silicon is 0.50 Å greater than that for carbon, $\varrho_k^{(0)}$ for silicon was chosen to be 2.28 Å, which is 0.50 Å greater than $\varrho_k^{(0)}$ for carbon. Using the same method, $\varrho_k^{(0)}$ for tin was chosen to be 2.48 Å. Finally, following the procedure used for carbon, $\varrho_k^{(1)}$ was set to zero for silicon and tin and $q_k^{(0)}$was not used as a parameter. Experimental solvation data are available only for the tetramethyl- and tetraethyl-substituted elements in both cases.[9] These molecules resulted in a zero surface area for both silicon and tin, making parameterization of the surface tension impossible. However, because the surface areas for silicon and tin are zero, it is possible to calculate ΔG_S° for these compounds using only the carbon and hydrogen surface tensions. The results are listed in Table 1. Using AM1, in the case of silicon, we oversolvate tetramethylsilane by 1.0 kcal/mol and undersolvate tetraethylsilane by 0.2 kcal/mol. Tetramethyltin is oversolvated by 0.9 kcal/mol, and the tetraethyl analog is undersolvated by 0.4 kcal/mol. PM3 does not contain parameters for tin, so results are only available for the silanes. Our PM3 model oversolvates tetramethylsilane by 1.0 kcal/mol and undersolvates tetraethylsilane by 0.5 kcal/mol. It should be noted that these results were obtained using parameters that were not optimized for these molecules and they were not included in the test set error calculations.

Nitro Compounds. As a whole, nitro-containing compounds are not treated well by the general model presented here. Some of the difficulty lies with the disparity in the semiempirical wave function treatment of these compounds. With respect to gas-phase heats of formation, AM1 does poorly with nitro compounds, while PM3 does quite well.[65] It is not recommended that AM1-SM4 or PM3-SM4 be used with nitro-containing compounds; therefore, nitro-containing compounds were not included in the parameterization test set or in the error calculations. Instead, the recommended treatment for nitro-containing compounds is to use only the PM3 Hamiltonian with an oxygen surface tension of −66.96 cal mol^{-1} Å^{-2} in the nitro group (and the other parameters equal to the standard SM4 parameters). All other parameters remain the same. With this treatment, the rms error over the six compounds is 0.55 kcal/mol with nitrotoluene being a significant outlier. Results of this analysis are presented in Appendix C of the supplementary material.

Subsets Containing Various Elements or Functional Groups. By examining specific subsets of the test set, we can see that the model generally does well for all elements parameterized. For example, using AM1, hydrocarbons have an rms error of 0.45 kcal/mol with a dispersion in the experimental data of 2.64 kcal/mol. Examining Table 5, we also see that, overall, the AM1 and PM3 models perform similarly for most functional groups. The facts that a single set of solvation parameters works equally well with charges obtained by mapping from AM1 and PM3 wave functions and that these two models yield such qualitatively similar results are a significant testimony to the physical nature of the CM1 and SM4 models. Thus, whereas for SM1, SM2, and SM3,[2-4]

(63) Pauling, L. *The Nature of the Chemical Bond,* 3rd ed.; Cornell University Press: Ithaca, NY, 1939; pp 221−224.
(64) Emsley, J. *The Elements*; Oxford University Press: New York, NY, 1989; pp 172, 196.
(65) Stewert, J. J. P. In *Reviews in Computational Chemistry*; Boyd, D. B., Lipkowitz, K. B., Eds.; VCH: New York; Vol. 1, pp 45−81.

Table 5. Summary of Errors (kcal/mol) for Each Type of Functional Group or Combination of Functional Groups in the Data Set Used for Parameterization

			AM1-SM4			PM3-SM4		
class of compds	no.[a]	dispersion of ref data	rms error	mean unsigned error	mean signed error	rms error	mean unsigned error	mean signed error
				CH Compounds				
normal alkanes	9	3.02	0.18	0.16	0.09	0.23	0.14	−0.10
branched alkanes	6	0.96	0.23	0.06	0.06	0.28	0.25	−0.04
cyclic alkanes	3	0.95	1.09	0.99	0.99	0.98	0.90	0.90
alkenes	5	1.08	0.39	0.38	−0.38	0.42	0.42	−0.42
alkynes	5	1.11	0.19	0.19	−0.12	0.19	0.14	0.00
aromatics	9	3.15	0.56	0.45	−0.15	0.81	0.47	0.31
subtotal—CH	37	2.64	0.45	0.34	−0.10	0.54	0.29	0.06
				CHO Compounds				
alcohols	11	1.68	0.52	0.35	0.10	0.44	0.28	0.10
aldehydes	6	1.45	0.28	0.23	−0.23	0.33	0.33	−0.33
ketones	9	1.23	0.44	0.34	−0.11	0.48	0.44	−0.24
esters	9	3.02	0.31	0.21	−0.03	0.51	0.41	0.21
ethers	5	1.33	0.68	0.62	0.14	0.79	0.63	0.37
subtotal—CHO	40	2.16	0.45	0.33	−0.02	0.51	0.40	0.02
				CHN Compounds				
amines	10	1.26	0.30	0.28	0.12	0.32	0.28	0.23
nitriles	4	1.25	0.16	0.14	0.14	0.21	0.20	−0.13
pyridines	6	0.49	0.31	0.26	−0.26	0.17	0.15	0.10
subtotal—CHN	20	1.23	0.28	0.24	0.01	0.26	0.22	0.12
				Others				
fluorides	5	1.68	0.37	0.29	0.29	0.23	0.18	0.18
chloroalkanes	9	0.80	0.18	0.12	−0.07	0.21	0.17	−0.16
chloroalkenes	5	0.72	0.20	0.18	−0.01	0.27	0.22	0.01
chloroaromatics	3	0.49	0.22	0.15	−0.13	0.25	0.21	0.20
bromohydrocarbons	9	0.87	0.26	0.20	−0.17	0.27	0.24	−0.10
iodohydrocarbons	7	0.88	0.27	0.19	−0.11	0.19	0.14	0.02
4 or more elements	9	0.67	0.52	0.47	−0.15	0.45	0.41	−0.02
thiols	3	1.13	0.35	0.28	−0.10	0.14	0.14	0.04
sulfides	4	0.94	0.29	0.26	0.26	0.18	0.15	−0.08
water, ammonia	2	0.29	0.77	0.57	−0.57	0.07	0.07	0.00
subtotal—other	56	1.41	0.29	0.21	−0.05	0.21	0.18	−0.02
total	153	2.05	0.40	0.29	−0.03	0.42	0.29	0.03
combined AM1/PM3	306	2.05	0.41	0.29	0.00			

[a] Number of unique compounds in the functional group. Each unique compound was calculated by AM1 and PM3, creating an overall test set of 306 data points.

the surface tensions were clearly making up for systematic deficiencies in the AM1 and PM3 charge distributions, with the SM4 model, this does not appear to be the case.

There are a few classes of solutes where the use of one Hamiltonian over the other is suggested. Aromatic molecules form a class where the rms errors in the two models differ significantly; the error in the PM3 model is dominated by chrysene. Removal of chrysene drops the error in both models to 0.35 kcal/mol over eight aromatic hydrocarbons. Although fluorine- and sulfur-containing molecules are treated reasonably well in both models, PM3 is significantly better than AM1 at predicting free energies of solvation for these. In addition, AM1 oversolvates ammonia by 1.1 kcal/mol while the error using PM3 is 0.1 kcal/mol, well within experimental error. For those wishing more information on choosing the proper Hamiltonian, full results are given in the supplementary material. Despite their differences, both models perform acceptably well over nearly the entire test set. In particular, nitrogen-, sulfur-, and halogen-containing compounds are treated extremely accurately, with rms errors that approach experimental uncertainty.

There are also two groups of compounds where neither AM1-SM4 nor PM3-SM4 does very well, namely carboxylic acids and saturated five- and six-membered rings. The carboxylic acids were left out of the parameterization set because of uncertainties in the meaning of the experimental data. The experimental data may be complicated by artifacts due to (i) carboxylic acid dimerization in hydrocarbon solvents and (ii) the presence of Lewis basic sites in the experimental gas-phase chromatography systems. Table 6 compares our predictions to literature data for three carboxylic acids, and we do indeed see much larger discrepancies than for, say, esters or alcohols. We note that both the effects mentioned above would tend to increase the apparent solubility of carboxylic acids in *n*-hexadecane, and that is consistent with the calculated free energies of solvation being systematically less negative than the reported experimental values.

Table 6. AM1-SM4 and PM3-SM4 Free Energies of Solvation (kcal/mol) for Carboxylic Acids

	ΔG_S°(AM1-SM4)	ΔG_S°(PM3-SM4)	ΔG_S°(expt)
acetic acid	−1.66	−1.02	−2.39
propanoic acid	−2.16	−1.52	−3.12
butanoic acid	−2.78	−2.19	−3.86

The second class of compounds where both models do poorly is saturated five- and six-membered ring systems. The models undersolvate any such system by about 1 kcal/mol. The most reasonable explanation for this problem, as pointed out by Abraham and co-workers,[66] is that saturated rings are more polarizable than acyclic alkanes, as indicated by their refractive indices. Making the hydrogen surface tension, $\sigma_{Hk'}^{CD}$, on satu-

(66) Abraham, M. H.; Chadha, H. S.; Whiting, G. S.; Mitchell, R. C. *J. Pharm. Sci.* **1994**, *83*, 1085.
(67) Carbo, R.; Hernandez, J. A.; Sarn, F. *Chem. Phys. Lett.* **1977**, *47*, 581.
(68) Mitin, A. V. *J. Comput. Chem.* **1988**, *9*, 107.

rated rings about 8% more negative would account for the systematic discrepancy, but making $\sigma_{Hk'}^{CD}$ a function of electronic properties beyond the bond order, such as polarizability or orbital energy, is beyond the scope of the present parameterization. Since we had the same systematic error for saturated rings in the SM2 and SM3 models for aqueous solvation, it will mostly cancel if one calculates free energies of transfer for aqueous to nonpolar phases.

5. Concluding Remarks

We have developed a quantum mechanical solvation model for free energies of solvation in the nonpolar solvent *n*-hexadecane. The solvation model consists of transferable parameters that are Hamiltonian-independent. In particular, the parameters presented here apply to both the AM1 and PM3 Hamiltonians (with CM1A and CM1P charge models, respectively) and should yield accurate free energies of solvation when used in conjunction with any Hamiltonian that provides reasonably high-quality geometries and partial charges. This model has two new features, use of the CM1 charge models and a set of both short- and long-range surface tensions. The new nonpolar solvation model is parameterized for compounds containing H, C, N, O, F, S, Cl, Br, and I. For 306 compounds, the RMS error is 0.41 kcal/mol, the mean unsigned error is 0.29 kcal/mol, and the mean signed error is 0.00 kcal/mol over a wide range of functional groups.

In addition to the practical importance of having a solvation model for large, nonpolar solvents, *the present study provides new insight into the physical factors governing the solvation process*. In particular, we learned that the size of the primary solvation shell depends on the character of the solvation effect under consideration, with dispersion interactions extending only a short way into the solvent but solvent structural perturbations extending further into solution. The physical reasonableness of this partitioning is corroborated by our analysis of the solvation of noble gases.

In future work we will extend SM4 to other solvents, including *inter alia* other alkanes, alcohols, chloroform, and water. Because the model is very physical, extension to other solvents should be straightforward. For example, we already have preliminary data indicating that solvation in all other alkanes can be accommodated in a single framework identical to the present one with only very slight modifications based on physical data.

Acknowledgment. The authors are grateful to Gregory Hawkins for assistance with the calculations and to Professor Michael Abraham and Professor Peter Carr for providing their experimental data and for helpful discussions. This work was supported in part by the National Science Foundation.

Supplementary Material Available: Text and tables giving extra details (23 pages). Included are the expression for the derivative of a CM1 charge with respect to the density matrix, listings of all values of R_k, $Q_k^{(0)}$, $Q_k^{(1)}$, $q_k^{(0)}$, $d_{kk'}^{(1)}$, $d_{kk'}^{(2)}$, $r_{kk'}^{(1)}$, and $r_{kk'}^{(2)}$, and complete results for all 306 data points in the parameterization and all other compounds discussed in the text. This material is contained in many libraries on microfiche, immediately follows this article in the microfilm version of the journal, and can be ordered from the ACS; see any current masthead page for ordering information.

JA942997+