Article

Solvation Free Energies for Aqueous and Nonaqueous Solutions Computed Using PM7 Atomic Charges

Sergei F. Vyboishchikov* and Alexander A. Voityuk*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: We describe a simple and accurate method, ESE-PM7, for calculating solvation free energies ΔG_{solv}° in aqueous and nonaqueous solutions. The method is based on a noniterative COSMO algorithm. Molecular geometries and atomic charges calculated using the semiempirical method PM7 are used to calculate ΔG_{solv}°. The method has been tested on 92 different solvents and 988 solutes. The mean absolute errors (MAEs) in ΔG_{solv}° in aqueous solutions estimated by the ESE-PM7 approach are found to be 1.62 kcal/mol for 389 neutral solutes and 3.06 kcal/mol for 139 ions. The MAEs for neutral molecules in organic solvents are 0.97, 0.74, and 0.51 kcal/mol in polar protic, polar aprotic, and nonpolar solvents, respectively. The developed method can be employed to quickly screen ΔG_{solv}° values of extended molecular systems including pharmaceutical and biological molecules.
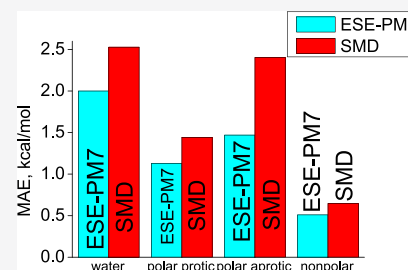
INTRODUCTION

The solvation free energy ΔG_{solv}° plays an important role in computational chemistry, since it can make a significant contribution to the total free energy of chemical reactions in solution. Most practical calculations of ΔG_{solv}° are based on the continuum solvation (CS) model. Usually, the computed ΔG_{solv}° is represented by the sum of the electrostatic energy E_{elst} and the correction term ΔG_{corr}, which mainly describes nonelectrostatic effects:

$$\Delta G_{solv}^{\circ} = E_{elst} + \Delta G_{corr}^{\circ} \tag{1}$$

The existing CS methods differ in the treatment of both E_{elst} and ΔG_{corr}°. The most popular CS approaches are the polarizable continuum model (PCM)[1−13] and the generalized Born (GB) method,[14,15] including SMx[16−20] and SMVLE[21] methods. The solvation methods in general[2] and PCM methods in particular[1] were reviewed in detail elsewhere.

The general idea of the PCM is that the solute placed in a cavity interacts with the solvent represented by a continuum with certain electrical properties. The polarization of the solvent by the solute is described by an electric charge distribution on the surface of the cavity. The charge distribution, in turn, is represented either by a continuous surface charge density $\sigma(\mathbf{r})$ or by discrete induced charges $\{q_i\}$. The electrostatic energy E_{elst} is calculated from the energy of the induced surface charge density $\sigma(\mathbf{r})$ in the electrostatic potential of the molecule $V(\mathbf{r})$:

$$E_{elst} = \frac{1}{2} \int V(\mathbf{r}) \, \sigma(\mathbf{r}) \, \mathrm{d}^2\mathbf{r} \tag{2}$$

where $\int \mathrm{d}^2\mathbf{r}$ is a surface integral over the molecular cavity. The electrostatic potential $V(\mathbf{r})$ is evaluated either from the electron density $\rho(\mathbf{r})$ and nuclear charges $\{Z_A\}$ located in nuclear positions $\{\mathbf{R}_A\}$:

$$V(\mathbf{r}) = -\int \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} \, \mathrm{d}^3\mathbf{r}' + \sum_A \frac{Z_A}{|\mathbf{R}_A - \mathbf{r}|} \tag{3}$$

or from atomic charges $\{Q_A\}$

$$V(\mathbf{r}) = \sum_A \frac{Q_A}{|\mathbf{R}_A - \mathbf{r}|} \tag{4}$$

Although the first approach (eq 3) is in principle more accurate than the discrete-charge formulation (eq 4), it suffers from the so-called outlying-charge problem.[22,23]

The cavity surface is typically constructed as a superposition of overlapping atom-centered spheres,[24] often smoothed near the seam of two spheres.[25,26] Then a set of surface grid points is generated and each atomic sphere is divided into spherical triangular tesserae.[24] Alternatively, the cavity can be bounded by an electron density isosurface corresponding to a certain ρ value.[27]

The nonelectrostatic effects that make an important contribution to ΔG_{solv}° include cavitation (formation of a cavity in the solvent), dispersion interactions, and specific interactions (e.g., the formation of strong hydrogen bonds in aqueous solution). Also, polarization of the solute by the solvent can be included in ΔG_{corr}°.

Among the most efficient PCM methods are COSMO, developed by Klamt and Schüürmann,[28−32] and the closely
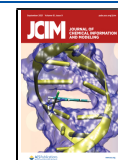
related C-PCM.[33,34] In our recent papers,[35−37] we presented a method for calculating solvation energies, dubbed uESE (universal easy solvation evaluation), based on the COSMO electrostatic term supplemented by an easily computed term ΔG_{corr}°. For aqueous solutions, uESE adopts the following form for the correction term:

$$\Delta G_{corr}^{\circ}(\text{uESE}) = \sum_A \kappa_A S_A + \sum_A p_A q_A{}^2 + \text{SRC} \tag{5}$$

where S_A are atomic surface areas; q_A are surface charges induced on these areas; κ_A and p_A are adjustable element-specific parameters. The first term, $\sum_A \kappa_A S_A$, corresponds to the cavitation energy plus solute−solvent dispersion interaction. This correction is similar to that used in many PCM[19,21,38] and GB[14,16,39,40] models. The $\sum_A p_A q_A{}^2$ term are electrostatics and polarization effects; SRC is the short-range correction that describes hydrogen bonds as well as a part of the Pauli repulsion. The same idea was employed for nonaqueous solutions, although with a slightly modified form of the ΔG_{corr}° term.

The uESE method is suitable for neutral and ionic solutes in many different solvents. The molecular charge distribution in the uESE method is represented by discrete nuclear-centered charges $\{Q_A\}$ computed using the DFT method. The best results were achieved using CM5 charges,[41] although other charge schemes give acceptable results as well.[35] Since the CM5 charges are derived using the Hirshfeld charge scheme,[42] the approach becomes time-consuming or even inapplicable to large-size systems. Thus, there is a need to develop a similar solvation method based on more readily available atomic charges for extended molecules including biopolymers. In this context, semiempirical methods that are about 3 orders of magnitude faster than DFT are ideal for calculating atomic charges in very large molecules.

Since the PM7 method[43] is widely regarded as the most advanced semiempirical approach, in this work we present a solvation energy scheme based on geometries and atomic charges derived from PM7 calculations.

Much effort has recently been made to elaborate fast and accurate CS methods for the computational treatment of solvent effects for large molecules. The ddCOSMO linear-scaling algorithm proposed by Lipparini et al.[44−46] enables very fast solution of the COSMO polarization equation. Its combination with the ESE-PM7 (easy solvation evaluation with PM7 charges) method will be of great benefit by treating extended systems.

■ METHOD

The total solvation free energy in the ESE-PM7 method is evaluated from the electrostatic E_{elst} and the correction ΔG_{corr}° terms according to eq 1.

Electrostatic Term. The calculation of E_{elst} for an N-atomic molecule in ESE-PM7 is very similar to that in our previous works[35−37] and follows closely Klamt and Schüürmann's original COSMO formulation.[28]

1. PM7 atomic charges $\{Q_A\}$ and coordinates $\{\mathbf{R}_A\}$ are read in from an output of an unperturbed (gas-phase) Gaussian[47] or MOPAC[48] calculation.

2. The van der Waals cavity surface is constructed as a superposition of spheres with modified Bondi radii[49] (*vide infra*) using the GEPOL93 algorithm[24] with 32 tesserae per atom. This gives the coordinates $\{\mathbf{r}_i\}$ for M surface points and the associated areas $\{S_i\}$.

3. An $M \times M$ matrix $\mathbf{A}$ and $M \times N$ matrix $\mathbf{B}$ are constructed:

$$A_{ij} = 1/|\mathbf{r}_i - \mathbf{r}_j| \quad \text{for } i \neq j$$

$$A_{ii} = 1.0694(4\pi/S_i)^{1/2} \tag{6}$$

$$B_{iA} = 1/|\mathbf{R}_A - \mathbf{r}_i|$$

4. The working equation $\mathbf{Aq} = -\mathbf{BQ}$ is solved with respect to the surface charges $\mathbf{q} = \{q_i\}$.

5. The resulting surface charges $\{q_i\}$ are used to compute the electrostatic energy:

$$E_{elst} = \frac{1}{2}(1 - 1/\varepsilon)\mathbf{qBQ} = \frac{1}{2}(1 - 1/\varepsilon)\sum_{i=1}^{M}\sum_{A=1}^{N} q_i B_{iA} Q_A \tag{7}$$

where ε is the dielectric constant of the solvent. In contrast to our previous paper,[37] we now use the same "Born-type"[50] factor $1 - 1/\varepsilon$ for all solvents.

As before, in ESE-PM7 we employ a noniterative implementation of COSMO, based on gas-phase atomic charges. The waiver of the iterative polarization calculation leads to a large saving in computing time. The effects of solute polarization are approximately taken into account by the correction term ΔG_{corr}°.

Correction (Nonelectrostatic) Term. In the present approach, we employ a ΔG_{corr}° form (eq 4) slightly different from that given by eq 5. First of all, we discarded the SRC term used previously[36,37] since it is not fully size consistent. Furthermore, the $\sum_A p_A q_A{}^2$ term is replaced by a linear term, $\sum_A g_A q_A$. The working ΔG_{corr}° expression for polar protic solvents is as follows:

$$\Delta G_{corr}^{\circ} = \sum_A (\kappa_A S_A + g_A q_A) + \zeta V \tag{8}$$

The ζV term implicitly describes the solute polarization and offers additional flexibility through the adjustable parameter ζ. Such a volume-dependent term was introduced by Andreussi et al.[38,51] According to Hille et al.,[52] the replacement of the volume by the solute polarizability leads to more accurate results.[52] The latter approach is elegant but hardly applicable to real models, since calculating the solute polarizability is fairly expensive.

All solvents are divided into four classes: (A) water; (B) nonaqueous polar protic solvents (mainly alcohols and phenols); (C) polar aprotic solvents; (D) nonpolar solvents, i.e., those with $\varepsilon < 9$. The same partition was employed in our previous work.[37] For polar protic solvents (classes A and B), ΔG_{corr}° is used in the form given by eq 8. To improve the flexibility of the model for solvent classes C and D, we have introduced an additional adjustable parameter—the solvent-dependent shift $\xi_{solv} S = \xi_{solv}\sum_A S_A$:

$$\Delta G_{corr}^{\circ} = \sum_A [(\xi_{solv} + \kappa_A)S_A + g_A q_A] + \zeta V \tag{9}$$

Note that the parameters κ_A and g_A are element-specific, whereas ξ_{solv} is solvent-specific but element-independent. In fact, ξ_{solv} is optional: omitting it will result in only a slight loss of accuracy.[37]

The atomic radii employed for cavity construction were modified. The same set of radii is used for all solvents, whereas κ_A, g_A, and ζ were fitted and tested separately for various classes of solvents. The nonlinear-least-squares fitting of van der Waals radii is described in the Supporting Information. Subsequently, Σ (eq 10) was minimized with respect to the adjustable parameters κ_A, g_A, and ζ:

Table 1. Van der Waals Radii R_A^{vdW} (Å) and Parameters ζ (kcal/mol·Å^{-3}), κ_A (kcal/mol·Å^{-2}), and g_A (kcal/mol) for Main-Group Elements for Various Classes of Solvents[a]

	H	C	N	O	F	S	Cl	Br	I
R_A^{vdW}	0.90	1.80	1.96	1.52	1.47	2.34	1.75	1.85	1.98
Solvent Class A: Water; $\zeta = 0$									
κ_A	0.124	0.057	−0.036	0.070	0.069	0.038	−0.016	−0.047	−0.066
g_A	10.0	−4.47	6.28	1.26	−3.67	−7.00	14.0	22.3	26.9
Solvent Class B: Polar Protic; $\zeta = -0.076$									
κ_A	0.037	0.097	0.017	−0.014	0.075	0.076	0.008	−0.034	−0.031
g_A	3.45	−6.09	6.71	13.2	4.87	−13.4	16.3	23.3	17.7
Solvent Class C: Polar Aprotic; $\zeta = -0.178$									
κ_A	−0.119	0.165	0.148	−0.074	0.053	0.129	0.010	0.012	
g_A	−7.99	8.88	11.0	32.8	18.9	18.5	32.0	33.0	
Solvent Class D: Nonpolar; $\zeta = -0.129$									
κ_A	−0.133	0.098	0.082	−0.065	0.057	0.106	0.061	0.048	0.052
g_A	−39.4	−22.6	−17.2	−6.22	−10.3	−31.5	−21.8	−22.2	−29.3

[a]The ξ_{solv} values are given in the Supporting Information.

Table 2. Mean Signed Error (MSE), Mean Absolute Error (MAE), and Standard Deviation (SD) of the Hydration Free Energy in kcal/mol for Various Data Sets by the ESE-PM7 Method with Respect to Reference Values in Comparison with the PM7/COSMO2 Method[a]

	ESE-PM7			ESE-PM7(SN)			PM7/COSMO2
solutes	MSE	MAE	SD	MSE	MAE	SD	SD
MNSol (528)[b]	0.19	2.00	2.79	0.17	1.90	2.62	
neutrals (389)	−0.04	1.62	2.21	−0.10	1.48	1.96	
cations (59)	1.25	3.13	3.91	1.38	3.37	4.20	
anions (80)	0.52	3.01	4.03	0.58	2.85	3.72	
MNSol* (464)[c]	0.35	1.91	2.64	0.30	1.82	2.53	2.62[d]
neutrals (330)	0.08	1.46	1.90	−0.04	1.33	1.72	2.24[d]
cations (59)	1.25	3.13	3.91	1.38	3.37	4.20	2.87[d]
anions (75)	0.86	2.91	3.91	0.93	2.73	3.56	3.69[d]
Mobley141 (141)[c]	−0.43	1.25	1.72	−0.36	1.19	1.65	2.54[e]
Blind (63)[c]	−0.83	2.53	3.49	−0.53	2.30	2.94	
SAMPL1 (53)[c]	−0.24	2.35	3.50	−0.15	2.30	2.91	3.73[d]
SAMPL4 (42)[c]	−0.29	1.28	1.60	−0.25	1.34	1.59	1.92[d]
C10 (10)[c]	1.00	1.65	2.22	1.08	1.77	2.31	2.28[d]

[a]The complete lists of solutes, the calculated hydration free energies, and the reference values are given in the Supporting Information. [b]Fitting set; for an explanation, see text. [c]Test set; for an explanation, see text. [d]Data from ref 54. [e]Data from ref 54, "Mobley266" data set.

Table 3. Mean Signed Error (MSE), Mean Absolute Error (MAE), and Standard Deviation (SD) of the Hydration Free Energy in kcal/mol for Various Data Sets by the ESE-PM7 Method with Respect to Reference Values in Comparison with the uESE, SMD, and SMD/PM3 Methods[a]

	ESE-PM7			SMD/DFT		uESE/DFT		SMD/PM3	
solutes	MSE	MAE	SD	MSE	MAE	MSE	MAE	MSE	MAE
MNSol (528)	0.19	2.00	2.79	2.06[b]	2.53[b]	−0.06[c]	1.47[c]	1.3[d]	2.3[d]
neutrals (389)	−0.04	1.62	2.21	0.57[b]	1.15[b]	−0.14[c]	0.99[c]	1.1[d]	1.5[d]
cations (59)	1.25	3.13	3.91	3.44[b]	3.76[b]	0.12[c]	2.73[c]	5.3[d]	5.5[d]
anions (80)	0.52	3.01	4.03	8.92[b]	8.92[b]	0.19[c]	2.83[c]	−1.1[d]	3.2[d]

[a]The complete lists of solutes and the calculated hydration free energies and the reference values are given in the Supporting Information. [b]Data from ref 36. [c]Data from ref 37. [d]Data from ref 58 (Table III).

$$\Sigma(\{\kappa_A\}, \{g_A\}, \zeta)$$
$$= \sum_k w_k \sum_i \left(E_{elst}(k, i) + \Delta G_{corr}^{\circ}(\{\kappa_A\}, \{g_A\}, \zeta) - \Delta^{ref} G_{solv}^{\circ}(k, i) \right)^2 \qquad (10)$$

where the index i runs over the molecules within a given subset and k runs over subsets; w_k is a suitable weighting factor. Each subset corresponds to a specific solvent. Also, different weighting factors were used for ions and neutral solutes. Since eqs 8 and 9

are linear with respect to the adjustable parameters κ_A, g_A, and ζ, the fitting is equivalent to solving a system of linear equations. Thus, the fitting problem is reduced to a single matrix inversion. Finally, the solvent-dependent parameter ξ_{solv} was found for each solvent by a separate least-squares fit.

Experimental values of ΔG_{solv}° from the Minnesota Solvation Database[53] were used as a reference. The PM7 charges were calculated for molecular geometries optimized at the PM7 level

using the Gaussian 16 program.[47] The final parameters κ_A, g_A, ζ, and ξ_{solv} for each of the solvent classes A−D are given in Table 1.

■ RESULTS AND DISCUSSION

The method was tested on several independent data sets that include both neutral and ionic solutes.
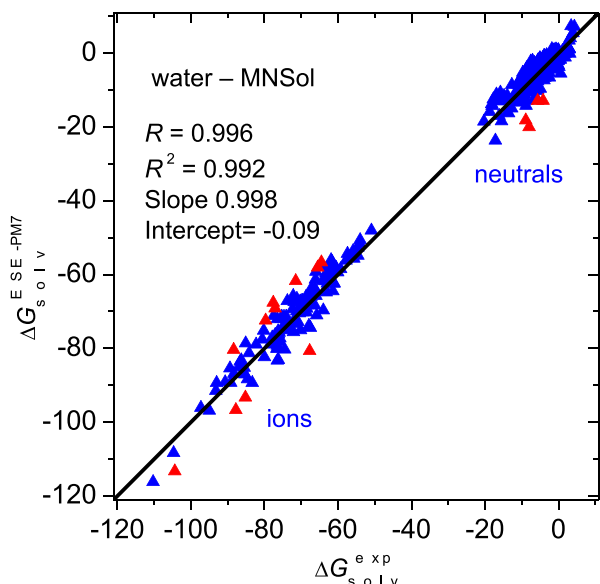

Figure 1. Hydration free energies (in kcal/mol) for 528 molecules and ions calculated by ESE-PM7 method versus experimental values. Red points denote outliers with a deviation greater than 7 kcal/mol.

Aqueous Solutions. In Table 2 we show a statistical evaluation of ESE-PM7 for aqueous solutions. Our test set of hydration energies $\Delta^{ref}G^{\circ}_{solv}$ ("MNSol") includes all suitable entries presented in the Minnesota Solvation Database:[53] 389 neutral molecules, 59 cations, and 80 anions. Five structures, for which PM7 yields qualitatively wrong geometries, were
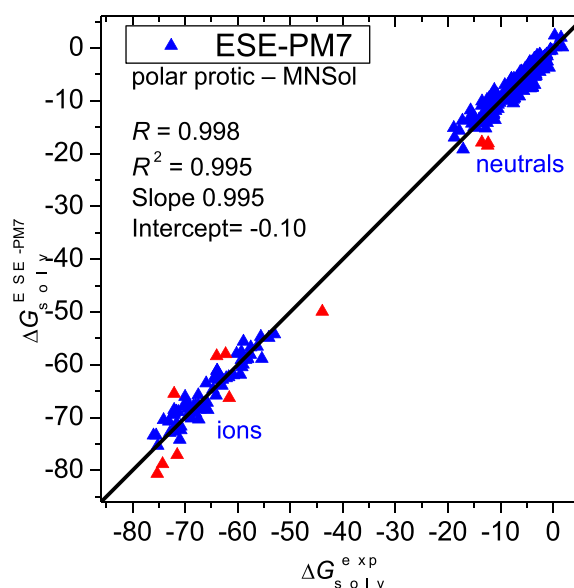

Figure 2. Solvation free energies (in kcal/mol) in nonaqueous protic solvents (class B) for 467 molecules and ions calculated by our ESE-PM7 method versus experimental values. Red points denote outliers with a deviation greater than 4 kcal/mol.

discarded. Also several other sets were tested: the data set of 464 solutes used by Kříž and Řezáč;[54] the subset of 141 solutes from Mobley et al.'s data set[55] ("Mobley"); Guthrie's SAMPL1 "blind challenge" data set ("Blind") containing 63 neutral pharmacologically important molecules;[56] reduced Guthrie's data set (53 molecules)[56] used by Kříž and Řezáč[54] ("SAMPL1"); reduced Guthrie's SAMPL4 data set[57] used by Kříž and Řezáč[54] ("SAMPL4"); ionic data set (six cations and four anions) by Kříž and Řezáč[54] ("C10").

A comparison with the PM7/COSMO2 method[54] done on the MNSol* data set of Kříž and Řezáč shows the same average performance of the methods with a standard deviation (SD;

Table 4. MSE and MAE of the Solvation Free Energy in kcal/mol for 14 Polar Protic Solvents Computed Using the ESE-PM7 Model in Comparison with uESE-CM5 and SMD (Total of 467 Entries)

solvent[a]	uESE-CM5[b]		SMD[c]		ESE-PM7		
	MSE	MAE	MSE	MAE	MSE	MAE	SD
octanol (247)	0.04	0.75	0.62	1.24	−0.03	1.02	1.40
heptanol (12)	0.25	0.47	0.72	0.74	0.20	0.85	0.95
m-cresol (7)	0.54	0.68	1.56	1.56	−0.09	1.13	1.33
benzyl alcohol (10)	0.09	0.35	0.42	0.66	−0.34	0.67	1.00
hexanol (14)	0.16	0.44	0.69	0.77	0.12	0.82	0.93
pentanol (22)	0.29	0.66	0.40	0.72	0.21	0.97	1.17
sec-butanol (9)	0.26	0.44	−0.23	0.53	−0.16	0.39	0.55
isobutanol (17)	0.09	0.63	−0.01	0.56	0.60	0.75	1.00
methoxyethanol (6)	−0.15	0.44	0.20	0.83	−0.99	0.99	1.21
butanol (21)	0.38	0.68	0.12	0.64	0.07	0.94	1.33
isopropanol (7)	−0.31	0.63	−0.91	1.02	−1.03	1.03	1.53
propanol (7)	−0.27	0.57	−0.66	0.81	−0.97	0.99	1.50
ethanol (8)	−0.60	0.82	−1.21	1.29	−1.33	1.33	1.65
methanol: cations (29)	−0.62	2.77	−0.32	2.44	0.14	2.18	2.86
methanol: anions (51)	−0.07	1.73	3.50	3.70	0.09	1.69	2.27
methanol: all ions (80)	−0.35	2.13	2.11	3.24	0.11	1.87	2.50
all polar protic (467)	0.02	0.93	0.74	1.44	−0.03	1.13	1.59

[a]The number of entries in the data set is given in parentheses. [b]Data from ref 37. [c]Data from ref 37 calculated using the Gaussian program package.[47]

Table 5. MSE and MAE of the Solvation Free Energy in kcal/mol for 20 Polar Aprotic Solvents Computed Using the ESE-PM7 Model in Comparison with uESE-CM5 and SMD (Total of 338 Entries)

solvent[a]	uESE-CM5[b]		SMD[c]		ESE-PM7		
	MSE	MAE	MSE	MAE	MSE	MAE	SD
bromoethane (7)	−0.04	0.61	−0.67	0.80	−0.09	0.72	1.05
2-methylpyridine (6)	0.09	0.53	−0.03	0.60	0.11	0.54	0.71
o-dichlorobenzene (11)	−0.06	0.41	−0.85	0.85	0.00	0.80	1.05
dichloroethane (39)	0.07	0.51	−0.18	0.49	0.06	0.57	0.77
4-methyl-2-pentanone (MIBK) (13)	0.21	1.01	−0.02	0.77	0.22	0.94	1.21
pyridine (7)	0.07	0.48	−0.19	0.65	0.00	0.61	0.91
cyclohexanone (10)	0.31	1.05	0.28	0.96	0.18	0.88	1.28
acetophenone (9)	0.20	0.69	−0.26	0.61	0.06	0.58	0.87
butanone (13)	0.00	0.81	−0.64	0.96	−0.09	0.90	1.16
benzonitrile (PhCN) (7)	−0.01	0.56	−0.64	0.77	−0.13	0.82	1.13
o-nitrotoluene (6)	0.01	0.16	−0.14	0.51	−0.06	0.41	0.60
nitroethane (7)	0.03	0.30	−0.37	0.58	−0.07	0.68	0.84
nitrobenzene (15)	0.01	0.22	−0.43	0.62	−0.07	0.54	0.73
acetonitrile (MeCN)							
neutral solutes (7)	0.65	0.69	−0.56	0.74	0.24	0.92	1.21
cations (39)	−0.28	2.01	7.65	7.93	−0.53	2.41	4.01
anions (30)	0.25	1.82	−2.63	3.01	0.60	1.60	1.96
all ions (69)	−0.18	1.97	3.18	5.79	−0.04	2.06	3.28
nitromethane (MeNO$_2$) (7)	0.06	0.53	−0.46	0.85	−0.01	0.78	0.94
dimethylformamide (DMF) (7)	0.12	0.49	0.07	0.65	0.02	0.67	0.90
dimethylacetamide (DMA) (7)	0.11	0.54	−0.01	0.76	0.03	0.64	0.89
sulfolane (7)	0.12	0.54	1.42	1.48	0.04	0.86	1.04
dimethyl sulfoxide (DMSO)							
neutral solutes (7)	0.12	0.75	0.53	0.88	−1.03	1.62	2.59
cations (4)	0.15	2.25	8.31	8.31	0.30	1.85	2.53
anions (66)	−0.14	2.01	−1.69	3.55	−0.39	3.00	3.95
methylformamide (NMF) (7)	0.17	0.85	0.23	0.86	0.07	0.96	1.15
polar aprotic							
neutrals (199)	0.10	0.59	−0.18	0.73	−0.01	0.74	1.07
cations (43)	−0.24	2.03	7.71	7.96	−0.46	2.36	3.90
anions (96)	−0.02	1.95	−1.99	3.38	−0.08	2.57	3.46
all polar aprotic (338)	0.02	1.16	0.31	2.40	−0.08	1.47	2.45

[a]The number of entries in the data set is given in parentheses. [b]Calculated using the Gaussian program package.[47] [c]Data from ref 37 calculated using the Gaussian program package.[47]

root-mean-square error) of about 2.6 kcal/mol. As seen, ESE-PM7 is clearly more advantageous for the neutral solutes but less accurate for ions. Also, ESE-PM7 is markedly more accurate when applied to the SAMPL1, SAMPL4, and Mobley data sets. For the C10 set of ions, ESE-PM7 and PM7/COSMO2 show similar performances.

A comparison with other solvation methods is provided in Table 3. With respect to the DFT-based SMD[19] method, which has become a method of choice for routine ΔG°_{solv} computations,[47] our noniterative ESE-PM7 method yields a substantially higher accuracy on the MNSol database (2.00 kcal/mol vs 2.53 kcal/mol). Although for the neutral solutes ESE-PM7 is not as accurate as SMD, it shows a much better performance for ions. With a mean absolute error (MAE) of about 3 kcal/mol, ESE-PM7 approaches the experimental accuracy. When tested on the "blind challenge" data set, the performance of ESE-PM7 is similar to that of SMD/DFT. Note that our DFT-based uESE method[37] yields substantially more accurate results in all categories.

In comparison with another semiempirical-based method, SMD/PM3[58] (Table 3), ESE-PM7 gives a better overall performance on the MNSol data set (MAE of 2.00 kcal/mol vs 2.3 kcal/mol), with a substantial advantage for ions and a slightly

higher MAE (1.62 kcal/mol vs 1.5 kcal/mol) for the neutral solutes.

In order to better evaluate the validity of our approach, Figure 1 shows hydration energies calculated for the MNSol data set, with failures ($|\Delta^{calc}G^\circ_{solv} - \Delta^{ref}G^\circ_{solv}| > 7$ kcal/mol) indicated in red. There are four problematic neutral solutes as well as 11 ionic solutes. To sum up, the ESE-PM7 method yields more accurate estimates of hydration free energies than the PM7/COSMO2 method, but it is less accurate than the best DFT-based methods, in particular for neutral species. Nevertheless, an MAE of about 1.6 kcal/mol for the neutrals provides sufficient confidence for most practical cases.

Since in certain cases an accuracy improvement may be still desirable, we have considered the possibility of introducing separate parameters for sulfur and nitrogen in high oxidation states (the ESE-PM7(SN) model). Indeed, this extra flexibility results in markedly better estimates (see Table 2) for the neutral solutes and anions. A substantial improvement (by 0.22 kcal/mol in MAE) achieved for the "Blind" data set shows that the ESE-PM7 model allows for a more flexible adjustment when needed. However, we do not consider this scheme as a general approach, since the explicit dependence on the coordination of
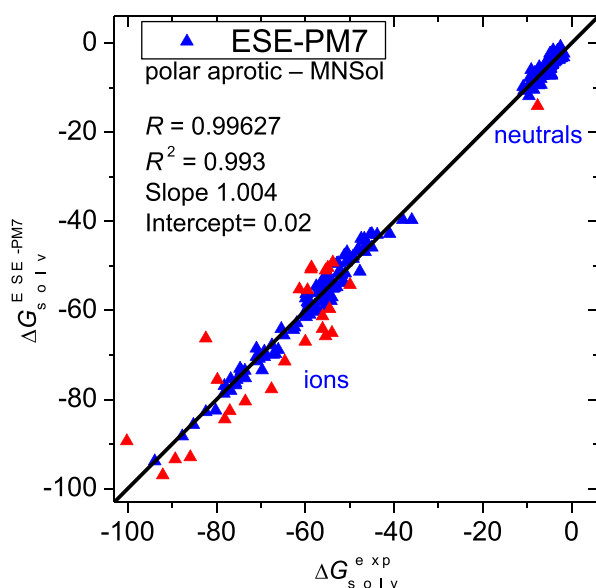
Figure 3. Solvation free energies (in kcal/mol) in polar aprotic solvents (class C) for 339 molecules and ions calculated by the ESE-PM7 method versus experimental values. Red points denote outliers with a deviation greater than 4 kcal/mol.

atoms may lead to wrong estimates for the system with untypical molecular configurations.

Nonaqueous Solutions. The statistics for the ESE-PM7 method for nonaqueous solvents (solvent classes B−D) are listed in Tables 4−6. Table 4 shows the results for polar protic solvents in comparison with two established DFT-based solvation schemes, uESE-CM5 and SMD. The overall performance of the ESE-PM7 method is better than that of the standard SMD method, with an MAE 0.17 kcal/mol lower. In particular, a better performance of ESE-PM7 is found for ΔG_{solv}° in octanol, *m*-cresol, *sec*-butanol, and methanol. The advantage of ESE-PM7 is particularly pronounced for ionic solutes in methanol, with an MAE about 1.3 kcal/mol lower than that of SMD. For hexanol and isopropanol ESE-PM7 has roughly the same performance as SMD, whereas it is slightly less accurate for the other eight solvents but still gives good estimates of ΔG_{solv}°, mostly within 1 kcal/mol.

Compared to the DFT-based uESE-CM5 method, ESE-PM7 does not achieve quite the same accuracy for neutral solutes, except for *sec*-butanol. However, for ions in methanol ESE-PM7 is more accurate, with an MAE about 0.25 kcal/mol lower. Overall, the ESE-PM7 results for protic solvents appear to be very convincing.

In comparison with the ΔG_{solv}° computed for methanol by Kromann et al. using the SMD approach in combination with several semiempirical methods,[58] ESE-PM7 with an SD of 2.86 kcal/mol is superior to DFTB/SMD[†], which shows the best performance yielding an SD of 4.7 kcal/mol.

A good agreement between the experimental and ESE-PM7 calculated solvation free energies is illustrated in Figure 2. For only 12 solutes out of 467 is the deviation beyond 4 kcal/mol, and none deviates more than by 7 kcal/mol. The most problematic cases are ions i191 (2,4-dinitrophenolate), i047 (NH_4^+), and i148 (hydrazinium $N_2H_5^+$). The found deviations are possibly due to the formation of strong hydrogen bonds.

Statistics for class C (polar aprotic) solvents in comparison with SMD and uESE-CM5 are presented in Table 5. According to these data, the general performance of ESE-PM7 is

significantly better than that of SMD, with an MAE nearly 1 kcal/mol smaller. The advantage of the ESE-PM7 method is particularly pronounced for ionic solutes. For neutral solutes, both ESE-PM7 and SMD yield virtually the same error. Taken individually, ESE-PM7 is more accurate for neutral solutes in 12 solvents, while for the other eight (dichloroethane, MIBK, PhCN, $MeNO_2$, MeCN, DMF, DMSO, NMF), SMD is somewhat superior. The accurate ESE-CM5 method is superior to ESE-PM7 for most polar aprotic solvents, with the exception of cyclohexanone, PhC(O)Me, MeCN (in the case of anions), and DMSO (in the case of cations). The ESE-PM7 method performs significantly better than all semiempirical methods considered by Kromann et al.:[58] the standard deviations of DFTB/SMD showing the best performance are 12.5 kcal/mol for DMSO and 10.9 kcal/mol for acetonitrile, which are much larger than the SD of ESE-PM7. For other solvents, the best SD found in ref 58 is 2.7 kcal/mol (PM6/COSMO), which is still larger than that yielded by ESE-PM7.

The situation for individual solutes in polar aprotic solvents is visualized in Figure 3. The only strongly deviating neutral solute is dimethyl sulfoxide (0503dim), belonging to the solutes (containing S−O bond) that give a large error for aqueous solutions as well (*vide supra*). Also various ions have significant errors (14 of 140 ions have an error beyond 6 kcal/mol). Still, 99 ions have a deviation below 3 kcal/mol, and 80 of them even have a deviation below 2 kcal/mol.

The comparative statistics for class D (nonpolar) solvents are presented in Table 6. Both DFT-based reference methods, SMD and uESE-CM5, are quite accurate, with an MAE well below 0.7 kcal/mol. Our ESE-PM7 method shows a similar performance, with a total MAE of 0.77 kcal/mol. For 46 of 57 nonpolar solvents tested, ESE-PM7 yields a smaller error than SMD. Only for seven solvents does SMD work better, while for four solvents ESE-PM7 and SMD provide virtually the same performance (MAE).

Overall, ESE-PM7 is a reliable method with an MAE below 1 kcal/mol for all nonpolar solvents, with the exception of nonanol. In terms of the SD, there are just 10 solvents out of 57 that are just slightly beyond 1 kcal/mol.

A graphical illustration of the solvation energies in nonpolar solutes is given in Figure 4. A strong deviation (>4 kcal/mol) is registered for 0400hyd (dihydrogen), 0447pho (diethyl-4-nitrophenylthiophosphonate), and 0441pho (dimethyl-4-nitro-phenylthiophosphonate). The latter molecules are probably difficult due to a terminal phosphorus-coordinated sulfur, while H_2 lacks an electrostatic contribution due to zero charges. Still, the overall performance of ESE-PM7 for nonpolar solvents is convincing: 1336 of 1554 molecules show an error below 1 kcal/mol, 1512 show an error within 2 kcal/mol, and only for 10 of them does the deviation exceed 3 kcal/mol. The corresponding numbers for the DFT-based SMD method are as follows: 1262, 1480, and 15 solutes, respectively.

■ CONCLUSIONS

We have developed a simple and accurate method, ESE-PM7, for calculating the solvation free energies in aqueous and non-aqueous solutions. The very fast estimation of ΔG_{solv}° is due to (1) the use of geometries and atomic charges of solutes calculated with the semiempirical method PM7 and (2) the noniterative COSMO algorithm employed to calculate the solute−solvent electrostatic interaction. The parametrization of the nonelectrostatic term ΔG_{corr}° for 92 solvents divided into four classes (water, polar protic, polar aprotic, and nonpolar solvents)

Table 6. MSE and MAE of the Solvation Free Energy in kcal/mol for 57 Nonpolar Solvents Computed Using the ESE-PM7 Model in Comparison with uESE-CM5 and SMD (Total of 1554 Entries)

| solvent[a]	uESE-CM5		SMD[b]		ESE-PM7		
	MSE	MAE	MSE	MAE	MSE	MAE	SD
pentane (26)	−0.01	0.28	−0.22	0.35	−0.01	0.34	0.50
hexane (59)	−0.01	0.36	−0.01	0.52	0.03	0.48	0.65
heptane (69)	0.02	0.37	0.22	0.55	0.07	0.42	0.60
isooctane (32)	−0.03	0.39	−0.32	0.45	−0.02	0.39	0.55
octane (38)	−0.01	0.28	0.06	0.43	0.01	0.36	0.50
nonane (26)	−0.02	0.22	0.04	0.37	0.01	0.18	0.22
decane (39)	−0.04	0.30	0.02	0.43	−0.02	0.29	0.47
undecane (13)	0.02	0.33	0.27	0.49	0.05	0.40	0.46
dodecane (8)	−0.12	0.34	0.06	0.30	0.02	0.18	0.21
cyclohexane (92)	−0.04	0.46	0.32	0.60	−0.01	0.49	0.68
perfluorobenzene (15)	0.06	0.36	0.49	0.56	0.03	0.38	0.46
pentadecane (9)	−0.12	0.31	0.39	0.48	0.04	0.13	0.16
hexadecane (198)	−0.05	0.45	0.32	0.68	−0.04	0.45	0.71
decalin (27)	0.02	0.30	0.57	0.67	0.02	0.37	0.51
carbon tetrachloride (79)	0.01	0.35	0.02	0.53	0.02	0.45	0.60
isopropyltoluene (6)	0.02	0.22	0.49	0.49	0.01	0.15	0.17
mesitylene (7)	0.02	0.25	0.09	0.54	0.05	0.39	0.50
tetrachloroethene (10)	0.02	0.26	0.57	0.74	0.01	0.18	0.21
benzene (75)	0.14	0.51	0.58	0.81	0.20	0.71	1.05
sec-butylbenzene (5)	0.04	0.16	0.12	0.25	0.04	0.19	0.21
tert-butylbenzene (14)	0.03	0.25	0.09	0.40	0.04	0.29	0.44
butylbenzene (10)	0.03	0.26	0.34	0.50	0.05	0.32	0.45
trimethylbenzene (11)	0.01	0.21	0.23	0.46	0.02	0.20	0.28
isopropylbenzene (19)	−0.02	0.25	−0.05	0.39	−0.04	0.28	0.46
toluene (51)	0.01	0.30	0.35	0.58	−0.01	0.39	0.52
triethylamine (7)	−0.03	0.58	0.77	0.98	−0.01	0.60	0.82
xylene (48)	0.03	0.34	0.45	0.60	0.01	0.39	0.53
ethylbenzene (29)	−0.03	0.30	0.17	0.47	−0.04	0.35	0.46
carbon disulfide (15)	−0.13	0.43	0.12	0.65	−0.14	0.71	1.16
tetralin (9)	−0.31	0.71	−1.08	1.30	−0.35	0.71	1.17
dibutyl ether (15)	0.06	0.60	0.61	0.79	0.06	0.58	0.86
diisopropyl ether (22)	0.03	0.91	0.41	0.76	−0.01	1.00	1.23
hexadecyl iodide (9)	−0.03	0.23	−0.07	0.42	−0.02	0.17	0.22
phenyl ether (6)	−0.09	0.35	−1.05	1.05	−0.11	0.48	0.76
fluorooctane (6)	−0.01	0.06	−0.48	0.48	−0.02	0.14	0.18
ethoxybenzene (7)	−0.09	0.34	0.08	0.45	−0.13	0.47	0.74
anisole (8)	−0.05	0.27	0.31	0.51	−0.08	0.46	0.75
diethyl ether (72)	−0.01	0.78	0.26	0.82	0.03	0.89	1.13
bromoform (12)	−0.04	0.24	0.72	0.72	−0.05	0.26	0.44
iodobenzene (20)	−0.07	0.45	−0.20	0.34	−0.06	0.50	0.75
chloroform (109)	0.00	0.64	0.28	0.79	0.06	0.81	1.15
dibromoethane (10)	−0.10	0.34	0.63	0.70	−0.08	0.34	0.47
butyl acetate (22)	0.08	0.56	1.08	1.15	0.08	0.66	0.92
bromooctane (5)	−0.04	0.16	−0.88	0.88	−0.05	0.27	0.32
bromobenzene (27)	−0.07	0.39	−0.48	0.51	−0.05	0.51	0.70
fluorobenzene (7)	−0.13	0.51	−0.79	0.83	−0.16	0.69	0.96
chlorobenzene (38)	−0.07	0.39	−0.63	0.65	−0.03	0.47	0.66
chlorohexane (11)	−0.01	0.17	−1.10	1.10	−0.02	0.30	0.40
ethyl acetate (24)	0.10	0.85	0.99	1.10	0.01	0.97	1.34
acetic acid (7)	0.06	0.46	2.37	2.37	−0.01	0.73	0.98
aniline (10)	0.22	0.70	0.76	0.78	0.20	0.84	1.23
dimethylpyridine (6)	−0.05	0.59	0.13	0.69	0.00	0.52	0.62
tetrahydrofuran (7)	−0.13	0.58	0.33	0.77	−0.17	0.72	0.97
decanol (11)	0.06	0.62	1.17	1.22	0.11	0.91	1.00
tributyl phosphate (16)	0.04	0.54	0.38	0.56	0.02	0.43	0.52
nonanol (10)	0.03	0.83	0.67	0.73	0.09	1.33	1.44
dichloromethane (11)	−0.18	0.67	−0.61	0.65	−0.29	0.83	1.14
all nonpolar (1554)	−0.01	0.44	0.22	0.65	0.01	0.51	0.77

Table 6. continued
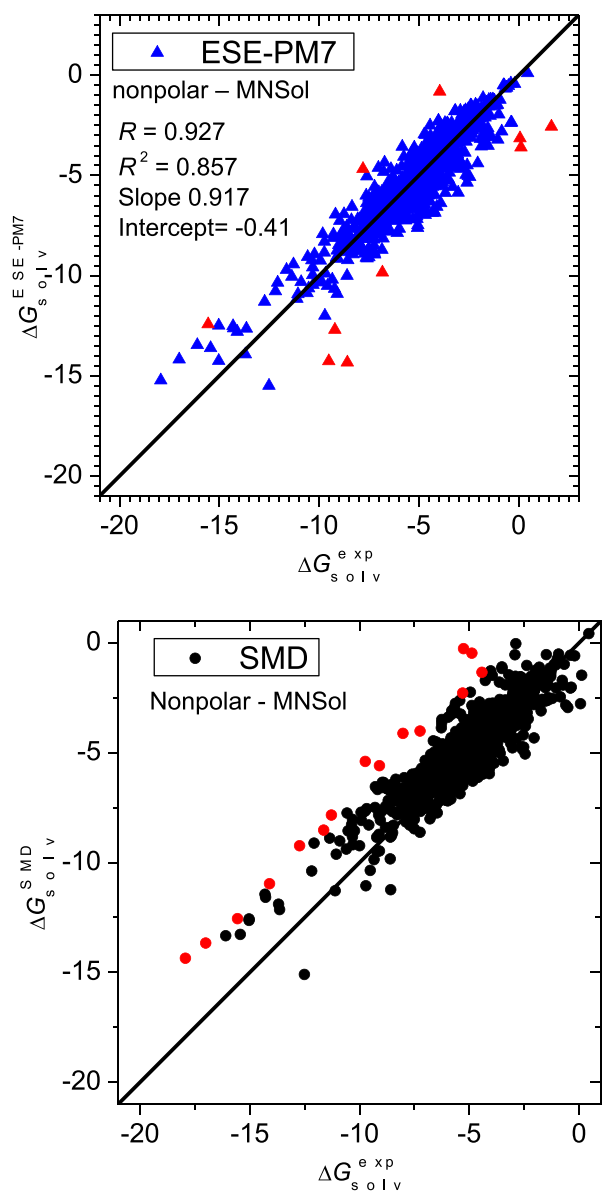
Figure 4. Solvation free energies (in kcal/mol) in nonpolar solvents (class D) for 1554 molecules calculated by the ESE-PM7 method versus experimental values. Red points denote outliers with a deviation greater than 3 kcal/mol. SMD results are given for comparison.

allows quite accurate predictions of $\Delta G^{\circ}_{\text{solv}}$. The MAE of the ESE-PM7 method in aqueous solutions is found to be 1.62 kcal/mol for 389 neutral solutes and 3.06 kcal/mol for 139 ions. For nonaqueous solutions, ESE-PM7 provides more accurate estimates of $\Delta G^{\circ}_{\text{solv}}$, with MAEs for neutral solutes of 0.97, 0.74, and 0.51 kcal/mol in polar protic, polar aprotic, and nonpolar solvents, respectively.

The ESE-PM7 method is intended for a standalone application and can be used to quickly estimate $\Delta G^{\circ}_{\text{solv}}$ of extended molecular systems and to screen multiple data on solvation in drug design. Furthermore, it can be used in combination with PM7 for molecular dynamics simulations over a nanosecond scale.

■ DATA AND SOFTWARE AVAILABILITY

The ESE-PM7 program executables and instructions are openly available for download at http://iqc.udg.es/~vybo/ESE-PM7.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.1c00885.

Least-squares fitting of atomic radii; parameters ξ_{solv} for solvent classes C and D; ESE-PM7-calculated $\Delta G^{\circ}_{\text{solv}}$ in various solvents (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

Sergei F. Vyboishchikov − *Institut de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, 17003 Girona, Spain;* ⓞ orcid.org/0000-0003-1338-3437; Email: vyboishchikov@googlemail.com

Alexander A. Voityuk − *Institut de Química Computacional i Catàlisi and Departament de Química, Universitat de Girona, 17003 Girona, Spain; Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain;* ⓞ orcid.org/0000-0001-6620-4362; Email: alexander.voityuk@gmail.com

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jcim.1c00885

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

■ REFERENCES

(1) Tomasi, J.; Mennucci, B.; Cammi, R. Quantum mechanical continuum solvation models. *Chem. Rev.* **2005**, *105*, 2999−3094.

(2) Skyner, R. E.; McDonagh, J. L.; Groom, C. R.; van Mourik, T.; Mitchell, J. B. O. A review of methods for the calculation of solution free energies and the modelling of systems in solution. *Phys. Chem. Chem. Phys.* **2015**, *17*, 6174−6191.

(3) Barone, V.; Cossi, M.; Tomasi, J. A new definition of cavities for the computation of solvation free energies by the polarizable continuum model. *J. Chem. Phys.* **1997**, *107*, 3210−3221.

(4) Cancès, E.; Mennucci, B.; Tomasi, J. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anistropic dielectrics. *J. Chem. Phys.* **1997**, *107*, 3032−3041.

(5) Mennucci, B.; Cancès, E.; Tomasi, J. Evaluation of solvent effects in isotropic and anisotropic dielectrics, and in ionic solutions with a unified integral equation method: theoretical bases, computational implementation and numerical applications. *J. Phys. Chem. B* **1997**, *101*, 10506−10517.

(6) Cammi, R.; Mennucci, B. Linear response theory for the polarizable continuum model. *J. Chem. Phys.* **1999**, *110*, 9877−9886.

(7) Cossi, M.; Barone, V.; Robb, M. A. A direct procedure for the evaluation of solvent effects in MC-SCF calculations. *J. Chem. Phys.* **1999**, *111*, 5295−5302.

(8) Tomasi, J.; Mennucci, B.; Cancès, E. The IEF version of the PCM solvation method: An overview of a new method addressed to study molecular solutes at the QM *ab initio* level. *J. Mol. Struct.: THEOCHEM* **1999**, *464*, 211−226.

(9) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Polarizable dielectric model of solvation with inclusion of charge penetration effects. *J. Chem. Phys.* **2001**, *114*, 5691−5701.

(10) Cossi, M.; Scalmani, G.; Rega, N.; Barone, V. New developments in the polarizable continuum model for quantum mechanical and classical calculations on molecules in solution. *J. Chem. Phys.* **2002**, *117*, 43−54.

(11) Lipparini, F.; Scalmani, G.; Mennucci, B.; Cancès, E.; Caricato, M.; Frisch, M. J. A variational formulation of the polarizable continuum model. *J. Chem. Phys.* **2010**, *133*, 014106.

(12) Pomogaeva, A.; Chipman, D. M. Hydration energy from a composite method for implicit representation of solvent. *J. Chem. Theory Comput.* **2014**, *10* (1), 211−219.

(13) Pomogaeva, A.; Chipman, D. M. Composite method for implicit representation of solvent in dimethyl sulfoxide and acetonitrile. *J. Phys. Chem. A* **2015**, *119*, 5173−5180.

(14) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendrickson, T. J. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.

(15) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Generalized Born Solvation Model SM12. *J. Chem. Theory Comput.* **2013**, *9*, 609−620.

(16) Cramer, C. J.; Truhlar, D. G. General parameterized SCF model for free energies of solvation in aqueous solution. *J. Am. Chem. Soc.* **1991**, *113*, 8305−8311.

(17) Cramer, C. J.; Truhlar, D. G. An SCF solvation model for the hydrophobic effect and absolute free energies of aqueous solvation. *Science* **1992**, *256*, 213−217.

(18) Cramer, C. J.; Truhlar, D. G. A universal approach to solvation modeling. *Acc. Chem. Res.* **2008**, *41*, 760−768.

(19) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **2009**, *113*, 6378−6396.

(20) Marenich, A. V.; Cramer, C. J.; Truhlar, D. G. Uniform treatment of solute−solvent dispersion in the ground and excited electronic states of the solute based on a solvation model with state-specific polarizability. *J. Chem. Theory Comput.* **2013**, *9*, 3649−3659.

(21) Liu, J.; Kelly, C. P.; Goren, A. C.; Marenich, A. V.; Cramer, C. J.; Truhlar, D. G.; Zhan, C.-G. Free energies of solvation with surface, volume, and local electrostatic effects and atomic surface tensions to represent the first solvation shell. *J. Chem. Theory Comput.* **2010**, *6*, 1109−1117.

(22) Klamt, A.; Jonas, V. Treatment of the outlying charge in continuum solvation models. *J. Chem. Phys.* **1996**, *105*, 9972−9981.

(23) Baldridge, K.; Klamt, A. First principles implementation of solvent effects without outlying charge error. *J. Chem. Phys.* **1997**, *106*, 6622−6633.

(24) Pascual-Ahuir, J. L.; Silla, E.; Tuñón, I. GEPOL: An improved description of molecular-surfaces. 3. A new algorithm for the computation of a solvent-excluding surface. *J. Comput. Chem.* **1994**, *15*, 1127−1138.

(25) Scalmani, G.; Frisch, M. J. Continuous surface charge polarizable continuum models of solvation. I. General formalism. *J. Chem. Phys.* **2010**, *132*, 114110.

(26) York, D. M.; Karplus, M. A smooth solvation potential based on the conductor-like screening model. *J. Phys. Chem. A* **1999**, *103*, 11060−11079.

(27) Foresman, J. B.; Keith, T. A.; Wiberg, K. B.; Snoonian, J.; Frisch, M. J. Solvent effects 5. The influence of cavity shape, truncation of electrostatics, and electron correlation on *ab initio* reaction field calculations. *J. Phys. Chem.* **1996**, *100*, 16098−16104.

(28) Klamt, A.; Schüürmann, G. COSMO: a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc., Perkin Trans. 2* **1993**, *2*, 799−805.

(29) Klamt, A. COSMO and COSMO-RS. In *Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Ed.; Wiley: Chichester, U.K., 1998; Vol. *1*, pp 604−616.

(30) Klamt, A.; Diedenhofen, M. A refined cavity construction algorithm for the conductor-like screening model. *J. Comput. Chem.* **2018**, *39*, 1648−1655.

(31) Andzelm, J.; Kölmel, C.; Klamt, A. Incorporation of solvent effects into density functional calculations of molecular energies and geometries. *J. Chem. Phys.* **1995**, *103*, 9312−9320.

(32) Klamt, A. The COSMO and COSMO-RS solvation models. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 699−709.

(33) Barone, V.; Cossi, M. Quantum calculation of molecular energies and energy gradients in solution by a conductor solvent model. *J. Phys. Chem. A* **1998**, *102*, 1995−2001.

(34) Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669−681.

(35) Voityuk, A. A.; Vyboishchikov, S. F. A simple COSMO-based method for calculation of hydration energies of neutral molecules. *Phys. Chem. Chem. Phys.* **2019**, *21*, 18706−18713.

(36) Voityuk, A. A.; Vyboishchikov, S. F. Fast and accurate calculation of hydration energies of molecules and ions. *Phys. Chem. Chem. Phys.* **2020**, *22*, 14591−14598.

(37) Vyboishchikov, S. F.; Voityuk, A. A. Fast non-iterative calculation of solvation energies for water and non-aqueous solvents. *J. Comput. Chem.* **2021**, *42*, 1184−1194.

(38) Dupont, C.; Andreussi, O.; Marzari, N. Self-consistent continuum solvation (SCCS): the case of charged systems. *J. Chem. Phys.* **2013**, *139*, 214110.

(39) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **1996**, *100*, 19824−19839.

(40) Cramer, C. J.; Truhlar, D. G. SMx continuum models for condensed phases. *Lecture Series on Quantum Chemistry* **2006**, *1*, 112.

(41) Marenich, A. V.; Jerome, S. V.; Cramer, C. J.; Truhlar, D. G. Charge Model 5: An extension of Hirshfeld population analysis for the accurate description of molecular interactions in gaseous and condensed phases. *J. Chem. Theory Comput.* **2012**, *8*, 527−541.

(42) Hirshfeld, F. L. *Theor. Chim. Acta* **1977**, *44*, 129−138.

(43) Stewart, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *J. Mol. Model.* **2013**, *19*, 1−32.

(44) Lipparini, F.; Stamm, B.; Cancès, E.; Maday, Y.; Mennucci. Fast domain decomposition algorithm for continuum solvation models: energy and first derivatives, B. *J. Chem. Theory Comput.* **2013**, *9*, 3637−3648.

(45) Lipparini, F.; Scalmani, G.; Lagardère, L.; Stamm, B.; Cancès, E.; Maday, Y.; Piquemal, J.-P.; Frisch, M. J.; Mennucci. Quantum, classical, and hybrid QM/MM calculations in solution: General implementation of the ddCOSMO linear scaling strategy, B. *J. Chem. Phys.* **2014**, *141*, 184108.

(46) Lipparini, F.; Lagardère, L.; Scalmani, G.; Stamm, B.; Cancès, E.; Maday, Y.; Piquemal, J.-P.; Frisch, M. J.; Mennucci, B. Quantum calculations in solution for large to very large molecules: a new linear scaling QM/continuum approach, B. *J. Phys. Chem. Lett.* **2014**, *5*, 953−958.

(47) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.;

Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. *Gaussian 16*, rev. A.03; Gaussian, Inc.: Wallingford, CT, 2016.

(48) Stewart, J. J. P. *MOPAC2016*; Stewart Computational Chemistry: Colorado Springs, CO, USA, 2016.

(49) Bondi, A. Van der Waals volumes and radii. *J. Phys. Chem.* **1964**, *68*, 441−451.

(50) Born, M. Volumen und Hydratationswärme der Ionen. *Eur. Phys. J. A* **1920**, *1*, 45−48.

(51) Andreussi, O.; Dabo, I.; Marzari, N. Revised self-consistent continuum solvation in electronic-structure calculations. *J. Chem. Phys.* **2012**, *136*, 064102.

(52) Hille, C.; Ringe, S.; Deimel, M.; Kunkel, C.; Acree, W. E.; Reuter, K.; Oberhofer, H. Generalized molecular solvation in non-aqueous solutions by a single parameter implicit solvation scheme. *J. Chem. Phys.* **2019**, *150*, 041710.

(53) Marenich, A. V.; Kelly, C. P.; Thompson, J. D.; Hawkins, G. D.; Chambers, C. C.; Giesen, D. J.; Winget, P.; Cramer, C. J.; Truhlar, D. G. *Minnesota Solvation Database*, ver. 2012. University of Minnesota, November 26, 2012. https://conservancy.umn.edu/bitstream/handle/11299/213300/MNSolDatabase_v2012.zip (accessed 2019-05-17).

(54) Kříž, K.; Řezáč, J. Reparametrization of the COSMO solvent model for semiempirical methods PM6 and PM7. *J. Chem. Inf. Model.* **2019**, *59*, 229−235.

(55) Mobley, D. L.; Dill, K. A.; Chodera, J. D. Treating entropy and conformational changes in implicit solvent simulations of small molecules. *J. Phys. Chem. B* **2008**, *112*, 938−946.

(56) Guthrie, J. P. Blind challenge for computational solvation free energies: introduction and overview. *J. Phys. Chem. B* **2009**, *113*, 4501−4507.

(57) Guthrie, J. P. SAMPL4, a blind challenge for computational solvation free energies: the compounds considered. *J. Comput.-Aided Mol. Des.* **2014**, *28*, 151−168.

(58) Kromann, J. C.; Steinmann, C.; Jensen, J. H. Improving solvation energy predictions using the SMD solvation method and semiempirical electronic structure methods. *J. Chem. Phys.* **2018**, *149*, 104102.