J Comput Aided Mol Des (2010) 24:317–333
DOI 10.1007/s10822-010-9333-9

Prediction of SAMPL2 aqueous solvation free energies and tautomeric ratios using the SM8, SM8AD, and SMD solvation models

Raphael F. Ribeiro · Aleksandr V. Marenich · Christopher J. Cramer · Donald G. Truhlar

Received: 7 December 2009 / Accepted: 12 March 2010 / Published online: 1 April 2010

Abstract We applied the solvation models SM8, SM8AD, and SMD in combination with the Minnesota M06-2X density functional to predict vacuum-water transfer free energies (Task 1) and tautomeric ratios in aqueous solution (Task 2) for the SAMPL2 test set. The bulk-electrostatic contribution to the free energy of solvation is treated as follows: SM8 employs the generalized Born model with the Coulomb field approximation, SM8AD employs the generalized Born approximation with asymmetric descreening, and SMD solves the non-homogeneous Poisson equation. The non-bulk-electrostatic contribution arising from short-range interactions between the solute and solvent molecules in the first solvation shell is treated as a sum of terms that are products of geometry-dependent atomic surface tensions and solvent-accessible surface areas of the individual atoms of the solute. On average, three models tested in the present work perform similarly. In particular, we achieved mean unsigned errors of 1.3 (SM8), 2.0 (SM8AD), and 2.6 kcal/mol (SMD) for the aqueous free energies of 30 out of 31 compounds with known reference data involved in Task 1 and mean unsigned errors of 2.7 (SM8), 1.8 (SM8AD), and 2.4 kcal/mol (SMD) in the free energy differences (tautomeric ratios) for 21 tautomeric pairs in aqueous solution involved in Task 2.

Keywords Free energy · Generalized Born · Implicit solvation · Poisson equation · Solvation · Solvation modeling · Tautomerism

Introduction

The annual SAMPL challenges organized by OpenEye Scientific Software present the unique opportunity to make an assessment of various solvation protocols with respect to the prediction of solvation effects for very diverse solutes including many containing functionality of interest to the pharmaceutical industry. Accurate theoretical description of solvation is critically important in modeling because solvation effects are essential components of all liquid-state chemistry, and it is impossible to understand liquid-phase organic, biological, or inorganic chemistry without including them. For example, solvation effects influence solvent-dependent changes in reaction rates and reaction mechanisms, chromatographic retention behavior, interfacial transport effects, protein folding, ligand-receptor binding, and pharmaceutical bioavailability. Various explicit and implicit solvation models have been employed for prediction of pharmacokinetic and thermodynamic properties of drug-like compounds in biological fluids, for instance, in the evaluation of protein–ligand binding free energies using the free energy of solvation (or desolvation) of the ligand as a key component [1, 2]. In addition, accurate computational methods used in solvation modeling may be a viable alternative to experimental measurement, especially for highly polar solutes for which experimental determinations of solvation free energies are

Electronic supplementary material The online version of this article (doi:10.1007/s10822-010-9333-9) contains supplementary material, which is available to authorized users.

R. F. Ribeiro · A. V. Marenich · C. J. Cramer (✉) · D. G. Truhlar (✉)
Department of Chemistry and Supercomputing Institute, University of Minnesota, 207 Pleasant Street SE, Minneapolis, MN 55455-0431, USA
e-mail: cramer@umn.edu

D. G. Truhlar
e-mail: truhlar@umn.edu

difficult or inaccurate. Solvation models differ in their representation of the solvent, their adoption of a classical or quantum description of the solute, and their computational cost [3–10]. As a consequence, their accuracy for the prediction of free energies of solvation and their range of applicability are also different. The SAMPL initiative is especially useful because it allows for evaluation of current solvation methods in terms of their predictive abilities on mostly unpublished and obscure data sets that are unlikely to have been part of the training set of any tested protocol. The testing of different models against common test sets is a particularly valuable exercise that facilitates comparison of their relative strengths and weaknesses. The insight that one can gain with such an exercise will serve well for further improvement of existing solvation models and liquid-phase simulation techniques.

The present SAMPL2 challenge is focused on the prediction of vacuum-water transfer energies (Task 1) and the prediction of tautomer ratios (Task 2) for compounds in solution, many of which have been selected to be pharmaceutically relevant. The SAMPL2 data set includes compounds from three different categories: obscure, explanatory, and investigatory. Obscure compounds have known but not readily available experimental data. Explanatory compounds, which have unexpected or interesting experimental values, have been included in hopes that the solvation models involved in the SAMPL2 challenge might explain their unusual properties. Measured solvation free energies are not available for investigatory compounds. In the present paper, we evaluate the performance of our most recent implicit solvation models, namely, SM8, SM8AD, and SMD, over the SAMPL2 test set. Section 2 places the solvation models used in the present study in the context of recent advances in the field. Section 3 summarizes further computational details. Section 4 presents results and discussion. Section 5 summarizes our conclusions.

Review of computational methods

Computational methods that include the solvent implicitly have been recognized as a powerful alternative to explicit simulation techniques because they allow one to treat the solute with quantum mechanical calculations at the same high levels as those used in the gas phase [5–10]. Reliable calculations of solutes in solution must take account of electrostatics, cavitation, dispersion, and solvent structure. In the most complete implicit solvation models, the solute is polarized self-consistently by the reaction field, and it is the interaction of the mutually polarized solute and solvent subsystems that is called the electrostatic contribution. The electrostatic contribution can be evaluated by solving the Poisson equation for a nonhomogeneous dielectric medium with the free charge being that of the continuous charge density of the solute or by using alternative approaches, for instance, the generalized Born (GB) approximation [11–13], which represents the solute as a system of point charges located at the nuclear positions.

We have recently introduced three successively improved self-consistent reaction-field continuum solvation models, namely Solvation Model 8 (SM8) [14], Solvation Model 8 with Asymmetric Descreening (SM8AD) [15], and Solvation Model D (SMD) [16]. "Continuum" denotes that the solvent is not represented explicitly in any of these models, but rather it is treated as a dielectric medium with variable surface tension at the solute–solvent boundary. These models separate the observable solvation free energy into three components. One of these, $\Delta G^{\circ}_{\text{conc}}$ depends on solute concentration and vanishes for an ideal solution with the same concentration as the vapor. The other two are independent of concentration and correspond to a dilute solution. Thus the standard-state free energy of solvation is

$$\Delta G^{\circ}_{\text{S}} = \Delta G^{\circ}_{\text{conc}} + \Delta G_{\text{ENP}} + G_{\text{CDS}} \quad (1)$$

where ΔG_{ENP} is the bulk electrostatic contribution resulting from the interaction of a solute with its reaction field, which is the electric field produced by the polarized charge density that the solute induces in the solvent, and G_{CDS} is explained below. The solute cavity for the bulk electrostatic calculation is taken as a superposition of nuclear-centered spheres, and the solvent dielectric constant is taken as its bulk value right up to cavity boundary. The final component in Eq. 1 accounts for everything except concentration and that part of the electrostatic contribution that is modeled with the bulk dielectric constant. It is nominally associated with cavity formation, dispersion, and solvent-structure effects, but it also includes the deviation from the assumed bulk behavior of the solute electrostatic interaction with the first solvent shell. Since the third contribution is semiempirical it also makes up for any systematic errors such as the inexact treatment of solutes. The third contribution will be called the CDS term, denoting the nominal ingredients of cavitation, dispersion, and solvent structure. All the three models are used in the present study in their standard, published form, with no changes for the present applications.

The bulk electrostatic contribution is treated differently in SM8, SM8AD, and SMD. The SMD model is based on the polarized continuous quantum mechanical charge density of the solute (the "D" in the name stands for "density"). The SMD bulk electrostatic contribution to the free energy of solvation arises from a self-consistent reaction field treatment that involves solution of the nonhomogeneous Poisson equation (NPE) by the Integral-Equation-Formalism Polarizable Continuum Model (IEF-

PCM) algorithm [17]. In contrast to SMD, the SM8 and SM8AD models treat the bulk electrostatics using the GB approximation based on partial atomic charges, whose interaction with the solvent and with each other is dielectrically screened by the polarized solvent and descreened by other parts of the solute. The SM8 and SM8AD models treat dielectric descreening effects in terms of the so-called Born radii of individual atoms in the solute molecule. The SM8 model employs the Born radius based on the Coulomb field approximation of Still et al. [13]. The SM8AD model improves on the earlier SM8 model by using a new asymmetric descreening algorithm with a modified formula for the Born radius suggested by Grycuk [18]. Both the SM8 and SM8AD models employ class IV charge models [19], in particular, Charge Model 4 (CM4) [20] and Charge Model 4 M (CM4M) [21]. In the present work we use only CM4M.

The CDS contribution to the free energy of solvation computed by SM8, SM8AD, and SMD is a sum of terms that are proportional (with geometry-dependent proportionality constants called atomic surface tensions) to the solvent-accessible surface areas of the individual atoms of the solute. The CDS terms were parameterized [14–16] to include all of the deviations of the electrostatics from the assumed bulk model, such as the inexactness of the solvent permittivity model, including assumed values for intrinsic Coulomb radii, uncertainties in the treatment of solute charge outside the solute cavity in the case of SMD, and the inexactness of the solute charge model in the case of SM8 and SM8AD.

The SM8, SM8AD, and SMD models have been recently tested using 26 combinations of various basis sets and density functionals over a set of 2,892 solvation data including 345 free energies of solvation for neutral solutes in water, 2,072 free energies of solvation for neutral solutes in 90 nonaqueous solvents, 143 transfer free energies for neutral solutes between water and 15 organic solvents, and 332 free energies of solvation for ions in acetonitrile, dimethyl sulfoxide, methanol, and water [15]. The number of solvation energy calculations performed in this testing totals 75,192 for each of the three models. The mean unsigned error averaged over 26 theoretical levels for 2,560 solvation data for neutral solutes is 0.6, 0.7, and 0.8 kcal/mol for SM8AD, SM8, and SMD, respectively. The mean unsigned error averaged over 26 theoretical levels for 332 free energies of solvation for ions is 4.0 (SM8AD), 4.4 (SM8), and 4.3 kcal/mol (SMD). Although all the three models perform nearly equally well, one should consider that the SM8 and SM8AD models employ partial atomic charges, and their accuracy for a particular level of electronic structure theory may depend on whether consistent partial charges can be computed for that level of theory, but the ability to compute such partial atomic charges is not guaranteed for all possible theory levels and basis sets. For this reason, solvation models that solve the NPE such as the density-based model SMD are deemed to be less sensitive to the choice of basis set, although basis sets containing diffuse functions can lead to errors due to charge lying outside the solute cavity [22]. On the other hand, the GB models such as SM8 and SM8AD can be favored for their lower computational cost compared to the cost of NPE solvers; one SCF iteration typically takes 1.5–2 times longer with SMD than with SM8 or SM8AD. Comparing the two GB approaches, we note that the SM8AD model based on the asymmetric descreening approximation [15, 18] is expected to be more realistic than the SM8 model, which is based on the Coulomb field approximation of Still et al. [13], particularly for cases when the individual partial atomic charges are asymmetrically situated in the molecule, i.e., located near the dielectric boundary rather than at the center of the molecular surface [15, 18].

Computational details

All the solvation free energies are given for the gas-phase solute having a standard state of an ideal gas at a gas-phase concentration of 1 mol/L and for the liquid-phase solute being dissolved in an ideal solution at a liquid-phase concentration of 1 mol/L, and the temperature is 298 K. In this case $\Delta G^{\circ}_{\text{conc}}$in Eq. 1 becomes $\Delta G^{\circ}_{\text{conc}} \equiv 0$. Free energies that employ this standard state definition will be denoted by the superscript "*". When one uses the same geometry in the gas phase and in solution (as we do here), ΔG_{ENP} reduces to ΔG_{EP}.

Following our earlier work on the SAMPL1 challenge [23] all free energies of solvation calculated for Task 1 (Transfer Energies) were computed using the Minnesota solvation models outlined above with the density functional M06-2X [24], and the 6-31G(d) [25, 26] basis set. We used the geometries included in the instructions for SAMPL2, except (as discussed below) for glycerol.

In order to gather further insights into our results, we also computed the free energy of solvation for the molecules present in the explanatory subset using solvent–solute clusters composed of the solute and one or two water molecules in chemically intuitive positions, i.e., interacting via hydrogen bonding or dipole–dipole forces. Gas-phase structures of the water-solute clusters were optimized using M06-2X [24]/MG3S [27].

For the solvent–solute clusters with a single water molecule, the free energy of solvation of a given solute $\Delta G^{*}_{\text{S}}(\text{M})$ is evaluated using the corresponding thermochemical cycle that relates $\Delta G^{*}_{\text{S}}(\text{M})$ to the calculated free energy of solvation of the water-solute cluster, $\Delta G^{*}_{\text{S}}(\text{H}_2\text{O} \cdot \text{M})$, as follows

$$\Delta G_S^*(M) = \Delta G_g^o(B.E.) - \Delta G^{o\rightarrow *} + \Delta G_S^*(H_2O \cdot M) - \Delta G_S^*(H_2O) - RT\ln[H_2O] \quad (2)$$

In Eq. 2, $\Delta G_g^o(B.E.)$ is the computed gas-phase binding free energy of the solute-water cluster that corresponds to the ideal-gas standard state of 1 atm denoted by the superscript "o", $\Delta G^{o\rightarrow *} \equiv RT\ln(24.46)$ refers to the free energy change between one mole of an ideal gas taken at 1 atm (24.46 mol/L) and one mole taken at the concentration of 1 mol/L, $\Delta G_S^*(H_2O)$ is the experimental free energy of solvation of water, and $RT\ln[H_2O] \equiv RT\ln(54.34)$ refers to the free energy change between one mole of H_2O ideal gas taken at the concentration of H_2O in liquid water (55.34 mol/L) and one mole taken at the concentration of 1 mol/L [16, 28]. A procedure for calculation of $\Delta G_S^*(M)$ using the solvent–solute clusters with more than one solvent molecule is described elsewhere [28].

Task 2 (Tautomeric Teasers) involved the calculation of the free energy difference between two tautomeric forms in solution. The free energy of each tautomer in solution is calculated as a sum of the corresponding gas-phase free energy and the aqueous free energy of solvation for a given tautomer. For all the compounds involved in Task 2, the gas-phase free energies at 298 K were calculated at the M06-2X/MG3S level of theory using the gas-phase molecular geometries optimized at the same level. The corresponding aqueous free energies of solvation were computed at the M06-2X/6-31G(d) level of theory with the M06-2X/MG3S gas-phase geometries using the SM8, SM8AD, and SMD continuum solvation models. To examine the accuracy of the M06-2X gas-phase free energies, we compared them with those calculated using the BMC-CCSD multicoefficient correlation method [29] for selected cases in Task 2. In these computations, noted further as BMC-CCSD//M06-2X/MG3S, we used the BMC-CCSD total electronic energies calculated at the M06-2X/MG3S gas-phase geometries with the 298 K thermal free energy corrections calculated at the M06-2X/MG3S level.

All calculations in this work were performed using two locally modified versions of the *Gaussian03* [30] electronic structure package, namely, Minnesota Gaussian Functional Module (*MN-GFM*) [31] and Minnesota Gaussian Solvation Module (*MN-GSM*) [32].

Results and discussion

Task 1: aqueous free energies of solvation

Task 1 involved the prediction of aqueous free energies of solvation (air–water transfer energies) for 41 polyfunctional molecules selected by Guthrie (see introduction to SAMPL2 challenge in this issue of the journal) and divided into three groups: (a) obscure compounds, (b) explanatory compounds, and (c) investigatory compounds. With respect to the explanatory compounds, the corresponding experimental aqueous free energies of solvation were provided in advance, so that unusual trends in the relevant properties of these compounds could be reproduced and explained. There were no experimental data provided for the investigatory group, and to the best of our knowledge, there are no experimental solvation energies available for these compounds in the literature.

Table 1 compares aqueous free energies of solvation calculated by SM8, SM8AD, and SMD for 30 out of 31 compounds originally present in the obscure and explanatory subsets to the corresponding experimental values. Note that we excluded 1-iodouracil from consideration here because the tested solvation models were not parameterized for iodine-containing compounds. The biggest outlier among the obscure compounds is diflunisal. Apart from diflunisal, the accuracy of SM8, SM8AD, and SMD depends on the nature of a tested compound. For example, SM8AD gives an error of 5.6 kcal/mol for ketoprofen whereas SM8 and SMD show much smaller deviations of 1.8 (SM8) and 2.1 kcal/mol (SMD), respectively, for the same compound. In addition, SM8AD overestimates the aqueous free energy of solvation of acetylsalicylic acid by 4.1 kcal/mol whereas the error for SM8 and SMD is only −1.6 kcal/mol on average. SM8AD is more accurate than SMD and SM8 in the case of sulfolane apparently due to a better representation of compounds with oxidized sulfur functionalities in the SM8AD training set [15].

In the case of D-xylose and D-glucose, the SMD model is less accurate than SM8 and SM8AD. However, caution should be exercised before consideration of the experimental data given in the SAMPL2 database for these carbohydrates because the experimental procedures for estimating the free energies of solvation of highly polar molecules are prone to a variety of problems, such as the decomposition of the solute at the high temperatures needed to achieve a reasonable vapor pressure, and inherent numerical errors in the extrapolation of the high-temperature results to the corresponding room-temperature values. On the other hand, the substantially more negative solvation free energies predicted for the two sugars by SMD may reflect enhanced outlying charge errors in these molecules heavily functionalized with oxygen lone pairs near the cavity boundary. Overall, for the set of obscure compounds, SM8 is slightly more accurate (within the uncertainty of the experimental data) than SM8AD and SMD, with 50% of its predictions giving errors of 1 kcal/mol or less, and only four molecules having errors greater than 2 kcal/mol.

Table 1 indicates that over the set of explanatory compounds all of our models did poorly in predicting the

Table 1 Aqueous free energies of solvation (kcal/mol) for the obscure and explanatory SAMPL2 sets (Experimental aqueous free energies of solvation (or vacuum-water free energies of transfer) were provided by OpenEye Scientific Software)

Compound	SM8	SM8AD	SMD	Exp
Obscure subset				
Uracil	−16.76	−16.56	−14.15	−16.59 ± 0.28
5-bromouracil	−16.39	−16.74	−14.09	−18.17 ± 0.55
5-chlorouracil	−16.59	−14.09	−14.34	−17.74 ± 0.78
5-fluorouracil	−17.38	−16.69	−13.88	−16.92 ± 0.88
5-trifluoromethyluracil	−16.52	−16.30	−13.46	−15.46 ± 0.16
6-chlorouracil	−14.24	−14.09	−11.32	−15.83 ± 1.22
Cyanuric acid	−20.32	−17.90	−14.77	−18.26 ± 0.27
Caffeine	−11.90	−12.36	−10.34	−12.64 ± 0.74
Methyl paraben	−9.85	−11.38	−9.09	−9.51 ± 0.26
Ethyl paraben	−9.50	−10.92	−9.17	−9.20 ± 0.30
Propyl paraben	−9.24	−10.60	−9.01	−9.37 ± 0.22
Butyl paraben	−8.42	−9.54	−8.36	−8.72 ± 0.27
Acetylsalicylic acid	−11.93	−13.98	−11.12	−9.94 ± 0.18
Diflunisal	−13.96	−16.41	−14.38	−9.40 ± 0.20
Flurbirprofen	−9.03	−10.46	−9.26	−8.42 ± 0.16
Ibuprofen	−6.88	−7.54	−7.42	−7.00 ± 0.64
Ketoprofen	−12.56	−16.42	−12.91	−10.78 ± 0.18
Naproxen	−10.90	−12.61	−11.29	−10.21 ± 0.18
Phthalimide	−11.99	−11.75	−8.61	−9.61 ± 0.50
Sulfolane	−13.26	−7.10	−12.29	−8.61 ± 0.31
D-glucose	−25.29	−29.27	−36.00	−25.47 ± 0.22
D-xylose	−21.71	−24.98	−28.31	−20.52 ± 0.27
MUE	1.28	2.17	2.73	
Explanatory subset				
Hexachlorobenzene	−0.23	−1.01	1.39	−2.30 ± 1.16
Hexachloroethane	−0.04	−0.70	1.24	−1.41 ± 0.10
Octafluorocyclobutane	3.60	3.82	4.42	3.01 ± 0.03
Trimethyl *o*-trifluoroacetate	−1.67	−3.61	−0.67	−0.80 ± 0.20
Trimethyl phosphate	−7.46	−10.18	−7.45	−8.70 ± 0.10
4-nitroaniline	−10.48	−12.00	−9.30	−10.01 ± 0.11
Glycerol[a]	−11.71	−12.20	−11.65	−13.40 ± 1.00
Pentachloronitrobenzene	−1.73	−2.08	1.41	−5.22 ± 0.25
MUE	1.47	1.68	2.28	
RMSE[b]	1.78	2.61	3.54	
MUE[b]	1.33	2.04	2.61	
MSE[b]	−0.24	−0.95	0.43	

Theoretical (SM8, SM8AD, and SMD) aqueous free energies were calculated using the Cartesian geometries provided by OpenEye Scientific Software unless noted otherwise. RMSE, MUE, and MSE refer to root mean squared error, mean unsigned error, and mean signed error, respectively. Note that we excluded 1-iodouracil originally present in the obscure subset from consideration here because the tested solvation models were not parameterized for iodine-containing compounds

[a] Calculated using a re-optimized geometry as discussed in the text; the aqueous free energies of solvation calculated using OpenEye's Cartesian geometries are −17.0 (SM8), −18.8 (SM8AD), and −21.0 (SMD) kcal/mol

[b] Calculated by summation over the 30 compounds of the obscure and explanatory subsets (see also a comment on 1-iodouracil in footnote a)

experimental solvation free energy of glycerol, with errors ranging from 4 to 8 kcal/mol when we used the geometry of glycerol as provided by OpenEye Scientific Software without any further structural optimization (and thus with only one conformation of glycerol being taken into account; see footnote a in Table 1). To further investigate the problem with glycerol, we performed a conformational search for the global minimum of glycerol using the MMFF94 [33] force field with the GMMX algorithm present in the PCModel software [34], and then we re-optimized the two lowest energy conformers at the mPW1PW [35]/MIDI! [36] level. The aqueous free energies of solvation for the two conformations were calculated using the SM8, SM8AD, and SMD solvation models. The lowest energy conformer is preferred over the next higher one by 3.5 kcal/mol in the gas phase and 2.9 kcal/mol in solution (complete details of solvation free energies with SM8, SM8AD, and SMD are included in the Electronic Supplementary Material along with the Cartesian coordinates for both glycerol structures). The resulting SM8, SM8AD, and SMD aqueous solvation free energies calculated for the lowest conformer in solution are reported in Table 1. These values are in much better

agreement with experiment (the average deviation is 1.6 kcal/mol) than those calculated using the SAMPL2 original geometry for glycerol (the average deviation is 5.8 kcal/mol). Over all the explanatory set, SM8 is more accurate than SM8AD and SMD for predicting the aqueous free energy of solvation of 3 out of the 8 molecules in this group with the accuracy of 1 kcal/mol or less whereas SM8AD and SMD show such an accuracy for only two molecules of the given subset.

Table 1 also shows mean signed and mean unsigned errors (MSE and MUE, respectively) as well as root mean squared errors (RMSE) in the predicted energies with respect to experiment. We conclude that all our models perform reasonably well on either obscure or explanatory set of compounds in the SAMPL2 challenge, with the MUE ranging from 1.4 kcal/mol (SM8) to 2.6 kcal/mol (SMD).

Table 2 shows the SM8, SM8AD, and SMD results for the investigatory compounds, for which no experimental data were given. The investigatory group may be subdivided into three classes: (1) phosphorous oxidized compounds, (2) sulfur oxidized compounds, and (3) azoles. For the 25 organophosphorus compounds tested in previous work [15], the mean unsigned errors (relative to the corresponding experimental data) in the SM8, SM8AD, and SMD aqueous free energies of solvation calculated at the M06-2X/6-31G(d) level of electronic structure theory were equal to 1.5 (SM8), 1.3 (SM8AD), and 1.7 kcal/mol (SMD) [15]. These errors are comparable with the absolute deviations in the SM8, SM8AD, and SMD predictions on hexamethylphosphoramide, dimethyl methylphosphonate, and methyl dimethylphosphinate, which are between 0.3 and 3 kcal/mol (Table 2). The mean signed errors in the aqueous free energies of solvation calculated by SM8, SM8AD, and SMD for the five sulfonylureas tested in previous work [15] were equal to −10.8 (SM8), −5.5 (SM8AD), and −7.4 kcal/mol (SMD) [15]. These compounds contain the S–O functionality; therefore, they structurally resemble the sulfones and sulfoxides presented in Table 2. Thus, for the four sulfones and sulfoxides from the investigatory set, the SM8AD model should provide the most accurate predictions whereas the SM8 model (along with SMD for two cases out of four) is likely to overestimate the corresponding solvation free energies. For oxazole, thiazole, and isothiazole, the absolute deviations in the corresponding solvation free energies predicted by any two of SM8, SM8AD, and SMD are between 0.1 and 2 kcal/mol. These deviations are comparable with those obtained in previous work [15] for imidazole (0.3–1.5 kcal/mol) using SM8, SM8AD, and SMD with M06-2X and 6-31G(d).

Table 2 Aqueous free energies of solvation (kcal/mol) for the investigatory SAMPL2 set (SM8, SM8AD, and SMD aqueous free energies of solvation (or vacuum-water free energies of transfer) were calculated using the Cartesian geometries provided by OpenEye Scientific Software)

Compound	SM8	SM8AD	SMD
Hexamethylphosphoramide	−7.47	−9.91	−10.19
Dimethyl methylphosphonate	−9.53	−11.22	−10.09
Methyl dimethylphosphinate	−9.79	−12.34	−12.81
Methyl phenyl sulfoxide	−11.53	−10.24	−9.21
Trifluoromethyl phenyl sulfoxide	−7.63	−5.52	−4.76
Methyl phenyl sulfone	−13.55	−8.54	−11.79
Trifluoromethyl phenyl sulfone	−8.32	−2.11	−5.20
Oxazole	−5.94	−4.58	−3.88
Thiazole	−4.86	−4.84	−4.15
Isothiazole	−6.22	−5.92	−3.98

Sometimes adding one or more solvent molecules to the solute molecule explicitly can substantially improve the performance of implicit solvation models [37, 38]. We have examined such an approach in the present work by clustering all of the molecules in the explanatory SAMPL2 test set, except glycerol. We added one water molecule to each tested solute molecule except 4-nitroaniline, for which we added two explicit water molecules. The resulting molecular structures of the water-solute clusters are presented in Fig. 1. Table 3 shows SM8, SM8AD, and SMD aqueous free energies of solvation of unclustered solutes calculated by Eq. 1 with the use of the corresponding solvation free energies of the clusters.

In general, the use of explicit solvent molecules for calculation of the solvation free energy (microsolvation) should improve on only using CDS terms to account for such effects as charge transfer between the solute molecule and the first solvation shell and the partial covalent character of strong hydrogen bonds. Table 3 indicates that on average the use of water-solute clusters has little effect on the performance of SM8, degrades the performance of SM8AD, and improves the performance of SMD. When the continuum approximation is justified, that is, when no "special" first-shell interactions mandate the inclusion of specific first-shell solvent molecules, the inclusion of one or more explicit solvent molecules should not degrade the performance of a continuum model (except for the issue that when one includes explicit solvent molecules, one should average properly, in a free energy sense, over their locations, but this is not done here), and this is the case for SM8. In the SMD model, improvement is observed after clustering, but derives almost entirely from the two chlorocarbons, suggesting that the atomic surface tension for Cl may be somewhat too hydrophobic in SMD (so that burying some of the exposed Cl surface with a clustering solvent molecule leads to an improved solvation free energy). The reason for the poor effect of clustering on the

Table 3 Aqueous free energies of solvation (kcal/mol) for selected compounds calculated with and without adding explicit water molecules (Experimental aqueous free energies (exp) of solvation (or vacuum-water free energies of transfer) were provided by OpenEye Scientific Software)

Compound	SM8	SM8/C	SM8AD	SM8AD/C	SMD	SMD/C	Exp
Hexachlorobenzene	−0.23	−0.78	−1.01	−1.29	1.39	0.81	−2.30
Hexachloroethane	−0.04	−1.65	−0.70	−2.03	1.24	−0.64	−1.41
Octafluorocyclobutane	3.60	4.55	3.82	4.52	4.42	4.05	3.01
Trimethyl *o*-trifluoroacetate	−1.67	−1.50	−3.61	−3.23	−0.67	−1.34	−0.80
Trimethyl phosphate	−7.46	−7.15	−10.18	−11.17	−7.45	−8.66	−8.70
4-nitroaniline	−10.48	−12.81	−12.00	−14.77	−9.30	−10.62	−10.01
Pentachloronitrobenzene	−1.73	−0.70	−2.08	−0.86	1.41	1.90	−5.22
MUE	1.47	1.61	1.68	2.45	2.28	1.65	

The SM8, SM8AD, and SMD aqueous free energies of solvation for unclustered solutes were calculated without adding explicit water molecules. The SM8/C, SM8AD/C, and SMD/C aqueous free energies of solvation for unclustered solutes were calculated by Eq. 2 using the calculated free energies of solvation of the corresponding water-solute clusters. MUE refers to mean unsigned error

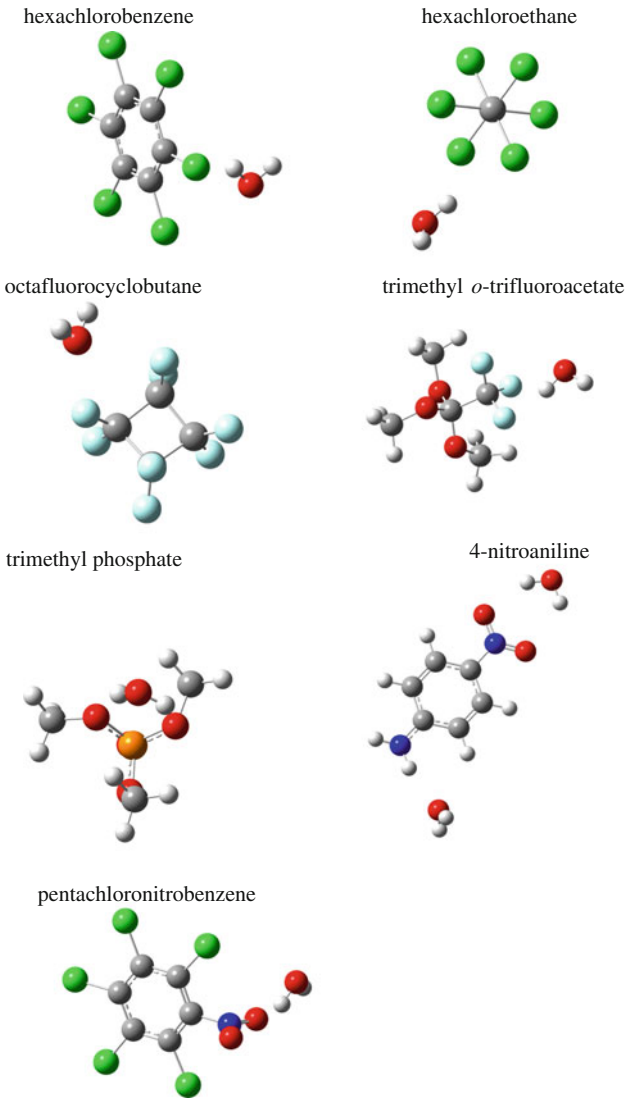


Fig. 1 Water-solute clusters for selected solutes involved in Task 1

predicted SM8AD solvation free energies is not clear, but it should be kept in mind that experimental uncertainties are high here.

Task 2: tautomeric teasers

The prediction of tautomeric equilibria presents a number of challenges from a modeling standpoint. For example many of the tautomeric forms explicitly provided as part of the challenge can themselves exist in multiple rotameric forms, and to complicate matters further the lowest-energy rotamer of a given tautomer may be different in the gas phase than in aqueous solution. In addition, the proximity of multiple hydrophilic functional groups in several of the tautomers suggests that consideration of explicit first-shell water molecules may be important in certain instances. A careful consideration of all of these issues, together with proper statistical averaging over equilibrium populations, should be done in order to make meaningful comparisons to experiment. Unfortunately, such careful attention devoted to each of the many tautomeric equilibria included in this challenge was not possible within the time frame of the exercise; instead, we elected to begin from the molecular structures provided by the SAMPL2 organizers and not to consider any others. This allows for a comparison of the results from the SM8, SM8AD, and SMD models to those from solvation models surveyed by other groups as part of this overall effort, but suggests that it may be dangerous to go beyond such internal comparison to consider additionally the experimental situation. Certainly we think that it will be worthwhile to return to many of these tautomeric equilibria in the future and explore all relevant conformational and methodological issues in detail, but the results presented below should be recognized as having restricted value beyond method-to-method comparison.

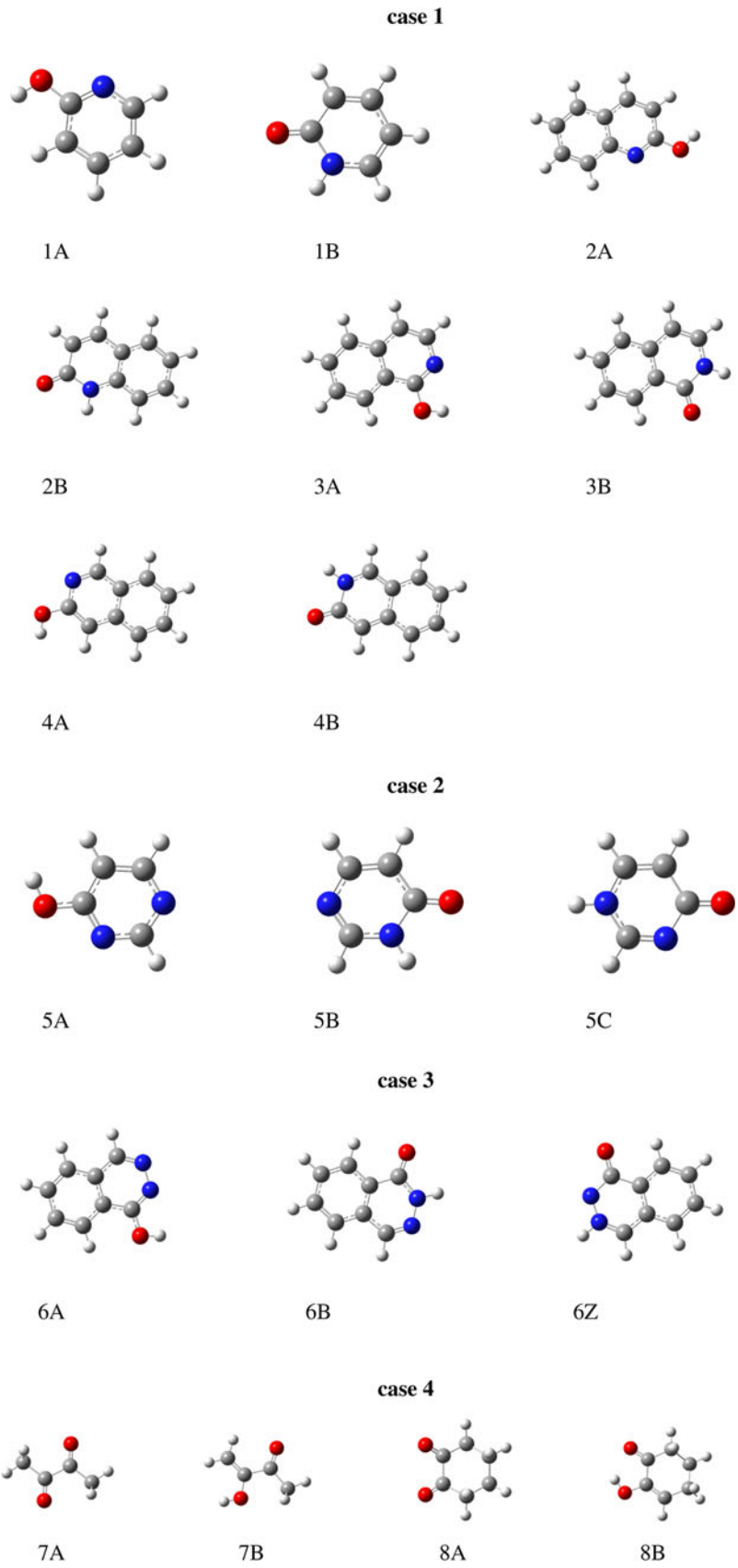


Fig. 2 Molecular structures for solutes involved in Task 2

Fig. 2 continued

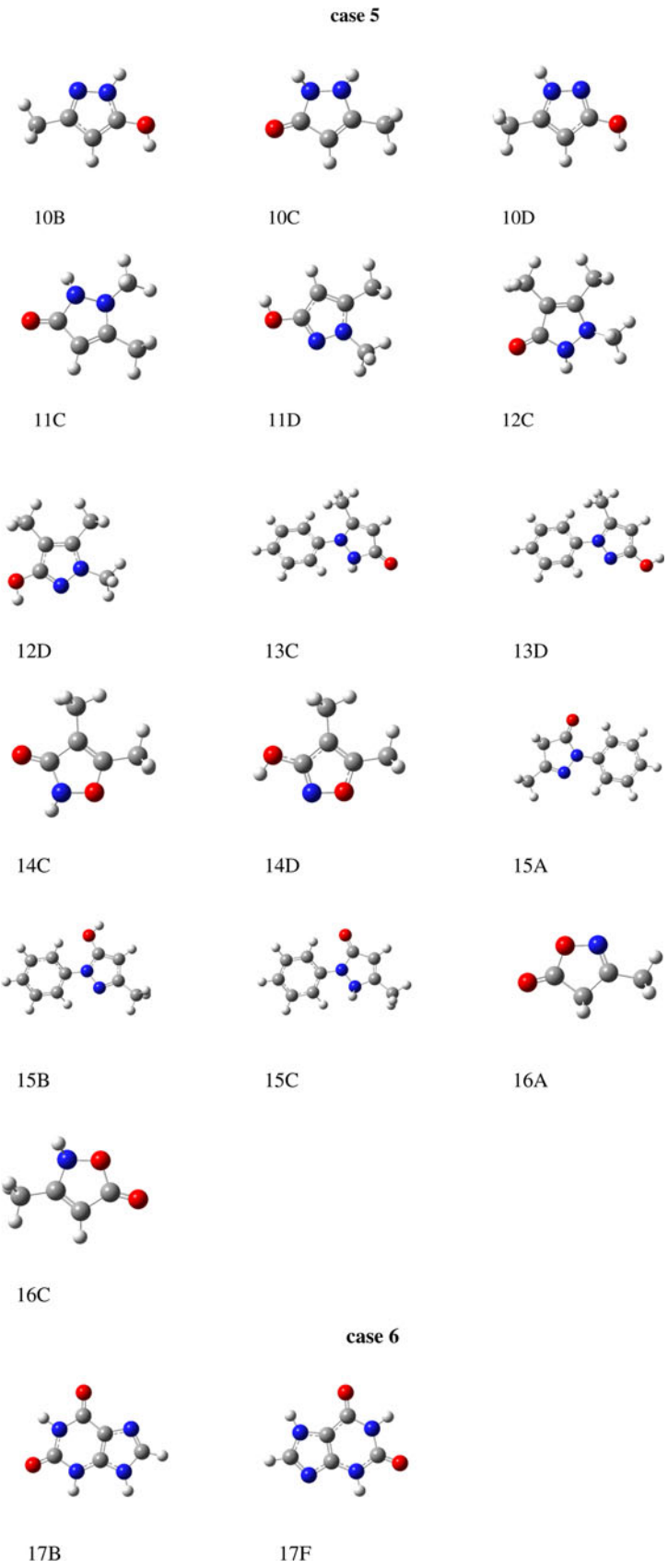


Fig. 2 continued

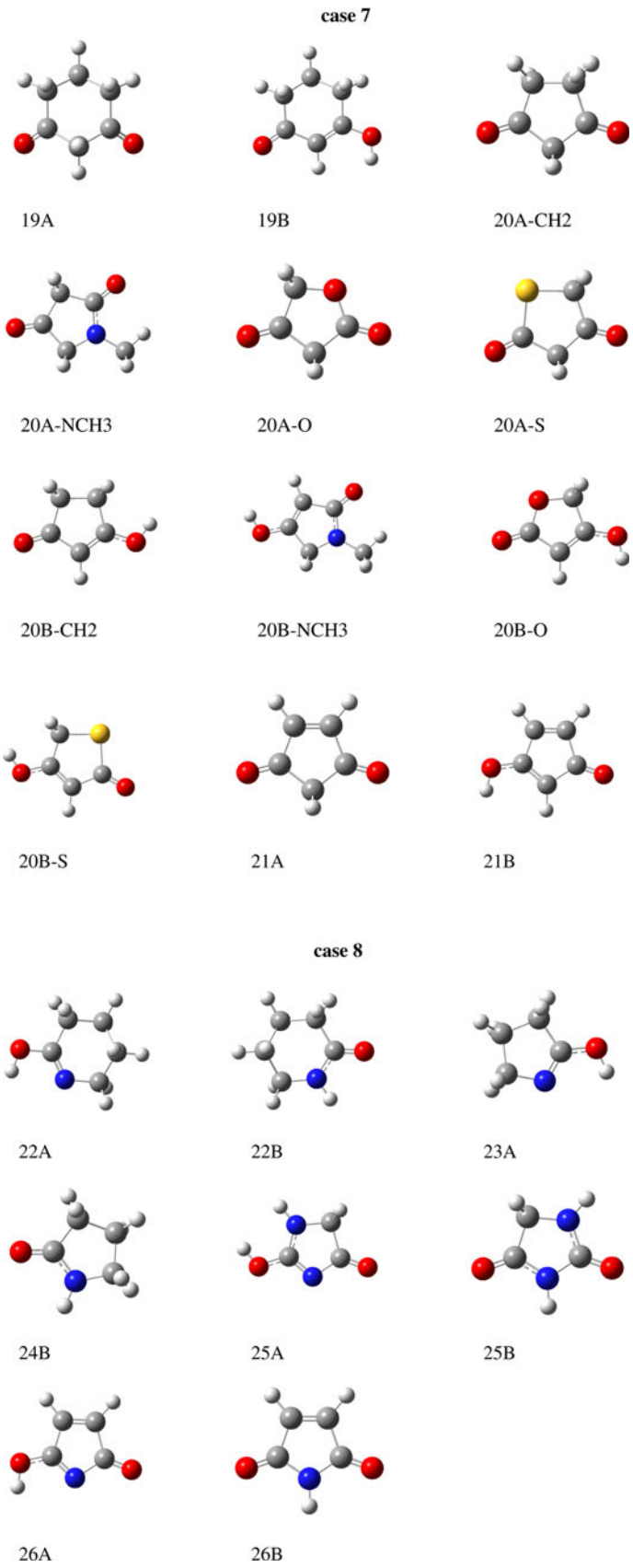


Fig. 2 continued

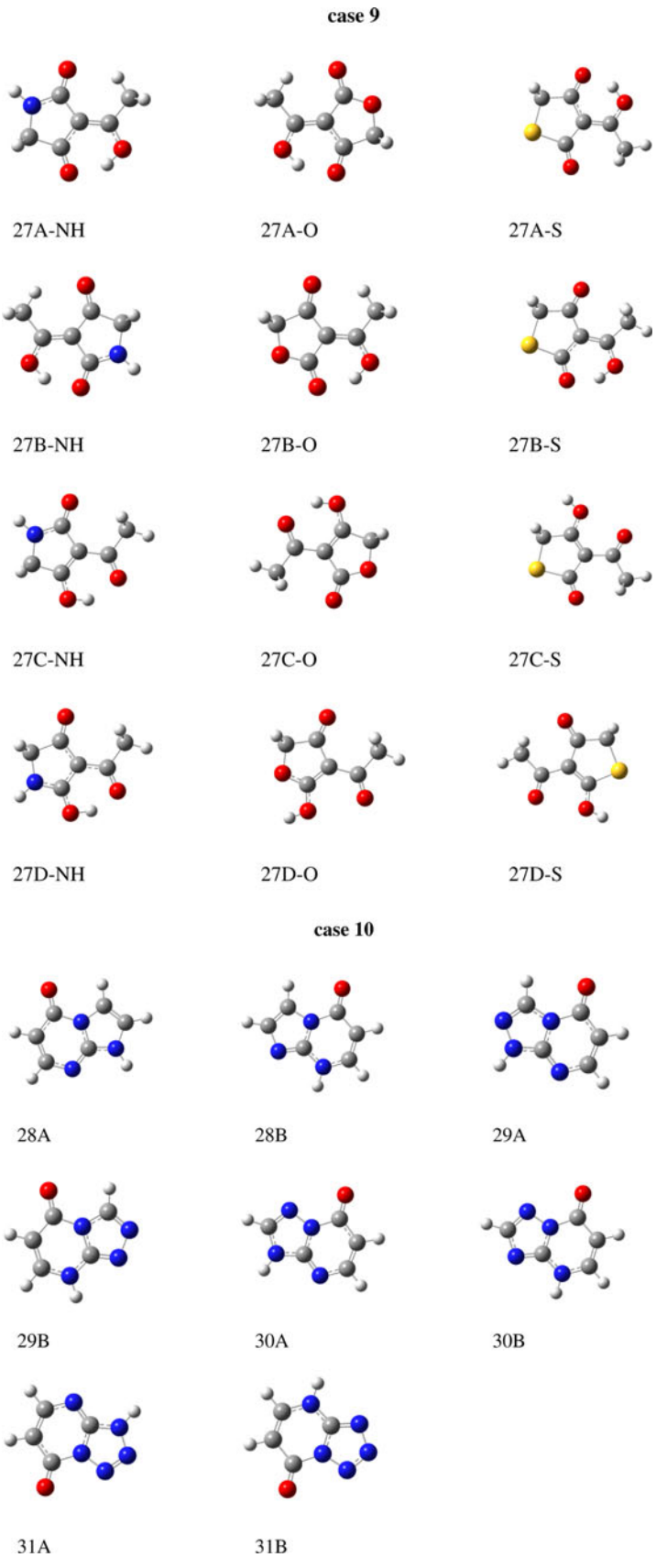
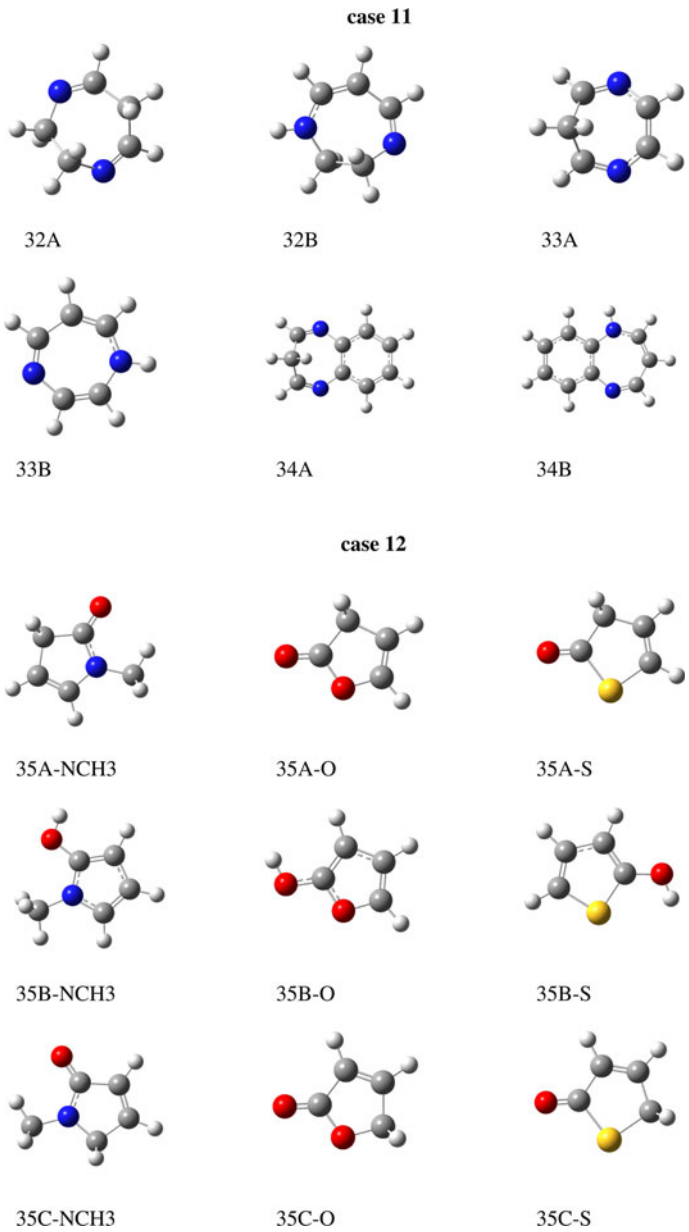


Fig. 2 continued



Task 2 involves the prediction of the free energy difference between two tautomers in aqueous media for twelve cases selected by Taylor and divided into three categories: (a) obscure, (b) explanatory, and (c) investigatory subsets. The obscure subset comprises cases 1–3, the explanatory subset comprises cases 4–6, and the investigatory subset comprises cases 7–12. The experimental values for the explanatory compounds were given along with the SAMPL2 assignment while the obscure subset constitutes a blind test for which the answer was known but not revealed at the time when the challenge was given. For the investigatory subset, the reference data are either absent or evaluated only qualitatively, and the purpose of this exercise is to predict them quantitatively using existing theoretical techniques. Figure 2 shows the molecular structures of all compounds involved in Task 2.

Table 4 compares the difference between the free energies of the two tautomers for each compound in the obscure and explanatory subsets, as calculated in the gas phase and in aqueous solution using the SM8, SM8AD, and SMD continuum solvation models, to the corresponding reference data in aqueous solution. The SM8AD model achieves the lowest mean unsigned errors in all three cases of the obscure subset with respect to the corresponding experimental data (Table 4). The SM8, SM8AD, and SMD models predict the dominant tautomer correctly for six

Table 4 Free energy difference (kcal/mol) for tested tautomeric pairs in the obscure and explanatory subsets (Notations used in the table for tautomeric pairs refer to the notations given in the SAMPL2 manual)

Tautomeric pair	Gas	SM8	SM8AD	SMD	Exp[a]
Obscure subset					
Case 1					
1A_1B	−3.5	−1.7	−3.0	−1.9	−4.8 ± 0.3
2A_2B	−8.3	−5.1	−6.1	−5.7	−6.1 ± 0.3
3A_3B	−4.4	−5.7	−6.7	−6.6	−7.2 ± 0.3
4A_4B	3.2	2.3	0.8	2.3	−2.3 ± 0.4
MUE		2.6	1.3	2.1	
Case 2					
5A_5B	−4.7	−3.4	−4.4	−2.8	−4.8 ± 0.5
5B_5C	8.7	1.0	1.5	1.9	0.5 ± 0.2
MUE		0.9	0.7	1.7	
Case 3					
6A_6Z	10.7	2.0	0.8	2.0	−2.4 ± 0.3
6A_6B	−10.7	−8.4	−9.7	−9.5	−9.2 ± 0.4
MUE		2.6	1.9	2.3	
Explanatory subset					
Case 4					
7A_7B	9.3	7.1	6.5	6.8	7.0 ± 1.5
8A_8B	−4.2	−2.9	−2.9	−2.9	−3.0 ± 3.0
MUE		0.1	0.3	0.1	
Case 5					
10B_10C	2.5	1.1	0.0	0.9	−2.9 ± 0.4
10D_10C	3.1	4.5	2.6	4.1	−1.2 ± 0.2
11D_11C	4.9	5.8	4.0	4.5	−0.5 ± 0.2
12D_12C	6.9	5.3	3.1	3.7	−1.8 ± 0.7
13D_13C	3.2	4.8	3.3	3.5	0.1 ± 0.1
14D_14C	3.1	2.0	0.8	2.1	0.3 ± 0.3
15A_15B	4.4	3.1	3.6	3.0	0.9 ± 0.3
15A_15C	5.3	2.9	2.3	2.0	−1.2 ± 0.3
15B_15C	0.9	−0.2	−1.2	−1.0	−2.2 ± 0.3
16A_16C	3.1	0.4	0.4	1.8	0.5 ± 0.1
MUE		3.8	2.7	3.3	
Case 6					
17F_17B	8.1	4.3	3.6	2.3	3.4
RMSE[b]		3.4	2.4	3.0	
MUE[b]		2.7	1.8	2.4	
MSE[b]		2.7	1.7	2.3	

For example, the notation M1_M2 refers to the free energy of the M1 → M2 reaction defined as the difference between the free energy of tautomer M2 and the free energy of tautomer M1. The SM8, SM8AD, and SMD free energy differences between two tautomeric forms in aqueous solution were calculated using the SM8, SM8AD, and SMD continuum solvation models, respectively. The free energy difference between two tautomeric forms in the gas phase was calculated as well ("gas"). MUE refers to mean unsigned error

[a] SAMPL2 reference data evaluated from the corresponding experimental tautomeric ratios in aqueous solution and provided by OpenEye Scientific Software

[b] RMSE, MUE, and MSE refer to root mean squared error, mean unsigned error, and mean signed error, respectively, calculated over all 21 data

tautomeric pairs in the obscure subset but fail to predict the dominant tautomer qualitatively for the tautomer pair 4A and 4B and for 6A and 6Z. Concerning the latter pair, the failure of our models can be related to the zwitterionic nature of 6Z which cannot be described adequately by the methods used in the present study. According to the SAMPL2 manual for Task 2 (Tautomeric Teasers), the corresponding equilibrium constant in the case of 5B and 5C is known to be very sensitive to solvent variation. Indeed, all the models tested in this study confirm that the difference in the free energies of 5B and 5C in aqueous solution can differ from that in the gas phase by factors of 5–9 (Table 4). All of our models predict the reference data for case 4 (*cis* and *trans*-α-diketones) and case 6 (xanthine tautomerism) in the explanatory subset quantitatively correctly (within the given uncertainties of the corresponding experimental energies). By contrast, the tested models fail to reproduce qualitatively the dominant tautomer for five out of ten tautomeric pairs in case 5 (pyrazolones and isoxazolones). We investigated the failure of our models in

Table 5 Comparison of the free energy difference (kcal/mol) for selected tautomeric pairs calculated using the M06-2X/MG3S and BMC-CCSD//M06-2X/MG3S methods (see the first footnote for Table 4 and the main text for more detail)

Tautomeric pair	M06-2X/MG3S				BMC-CCSD//M06-2X/MG3S				Exp[a]
	Gas	SM8	SM8AD	SMD	Gas	SM8	SM8AD	SMD	
Case 3									
6A_6Z	10.7	2.0	0.8	2.0	13.5	4.7	3.6	4.7	−2.4 ± 0.3
6A_6B	−10.7	−8.4	−9.7	−9.5	−15.9	−13.5	−14.8	−14.6	−9.2 ± 0.4
MUE		2.6	1.9	2.3		5.8	5.7	6.3	
Case 5									
10B_10C	2.5	1.1	0.0	0.9	2.4	1.0	−0.1	0.8	−2.9 ± 0.4
10D_10C	3.1	4.5	2.6	4.1	3.3	4.6	2.8	4.3	−1.2 ± 0.2
11D_11C	4.9	5.8	4.0	4.5	5.0	5.9	4.2	4.6	−0.5 ± 0.2
12D_12C	6.9	5.3	3.1	3.7	6.8	5.2	3.1	3.6	−1.8 ± 0.7
13D_13C	3.2	4.8	3.3	3.5	3.4	5.0	3.5	3.7	0.1 ± 0.1
14D_14C	3.1	2.0	0.8	2.1	3.8	2.7	1.5	2.8	0.3 ± 0.3
15A_15B	4.4	3.1	3.6	3.0	5.2	3.9	4.4	3.8	0.9 ± 0.3
15A_15C	5.3	2.9	2.3	2.0	6.2	3.8	3.3	3.0	−1.2 ± 0.3
15B_15C	0.9	−0.2	−1.2	−1.0	1.0	0.0	−1.1	−0.9	−2.2 ± 0.3
16A_16C	3.1	0.4	0.4	1.8	4.5	1.8	1.8	3.3	0.5 ± 0.1
MUE		3.8	2.7	3.3		4.2	3.1	3.7	

[a] See footnote a to Table 4

regard to case 5 as well as in regard to the tautomeric pair 6A and 6Z in view of the possible error in the M06-2X/MG3S gas-phase free energies for the corresponding molecules. Table 5 compares the free energy differences for the problematic tautomeric pairs calculated using the M06-2X/MG3S//M06-2X/MG3S gas-phase free energies to those calculated using the BMC-CCSD//M06-2X/MG3S gas-phase free energies. In general, the BMC-CCSD method does not improve the resulting free energy differences for the given tautomeric pairs in solution, producing even larger errors as compared to M06-2X/MG3S. Thus we conclude that the failure of our methods to correctly predict the reference data for the given tautomeric pairs in aqueous solution can be attributed to the inaccuracy of our solvation models for these particular cases. However, some of the reference data were evaluated using empirical linear structure energy relationship techniques involving the use of empirical solvent parameters to derive solute-specific equations for equilibrium constants. Therefore, one can speculate that the uncertainty in the corresponding reference data can be larger than previously estimated (0.1–0.7 kcal/mol in Table 5). On the other hand, we notice very good agreement with the experimental value for the tautomeric pair 16A and 16C (Table 4) obtained by direct UV observation [39] which is the most accurate datum for case 5. It would also be interesting to examine solvent clustering in these instances, but we did not undertake this task.

Table 6 presents the free energy difference predicted for the corresponding tautomers in the investigatory subset for which the experimental tautomeric ratios are known at least qualitatively. Tetronic acids (20A/B.o) are believed to be enolized in solution, and this assumption is confirmed by SM8 and SM8AD. The SMD model predicts that the diketone form is slightly dominating. Substitution of oxygen by sulfur results in a stronger preference for the enolized form, whereas the addition of an N-methyl or a CH_2 group results in the contrary trend with all of the tested models. Case 8 comprises cyclic lactams and similar compounds. All our models correctly predict that the oxo-form should be dominant in aqueous solution. Case 9 comprises cyclic triketones for many of which our models predict that their tautomeric ratios in aqueous solution are close to those in the gas phase in accord with a suggestion given in the SAMPL2 manual that the cyclic triketones can be highly polar and their polarity is likely to remain the same between tautomers and, therefore, their tautomeric ratios may not be affected by solvation. Case 10 involves the nitrogen-nitrogen tautomerism in compounds with bridgehead nitrogen. Our results indicate that for the tautomeric pair 29A and 29B and for 31A and 31B the solvation effects favor the enolized (B) form whereas for the tautomeric pair 28A and 28B and 30A and 30B the solvation effects only slightly influence the tautomeric ratio as well as for diazepines (case 11). For five membered ring 2-

Table 6 Free energy difference (kcal/mol) for tested tautomeric pairs in the investigatory subset (see the first footnote of Table 4 and the main text for more detail)

Tautomeric pair	Gas	SM8	SM8AD	SMD
Case 7				
19A.h_19B.h	1.5	0.5	0.4	0.1
20A.ch2_20B.ch2	4.9	1.2	0.7	0.2
20A.o_20B.o	2.8	−0.2	−0.7	0.2
20A.s_20B.s	2.3	−1.8	−2.4	−1.8
20A.nch3_20B.nch3	4.7	1.8	1.5	2.9
21A_21B	13.9	13.0	12.7	12.7
Case 8				
22A_22B	−10.1	−10.0	−11.5	−11.3
23A_23B	−11.7	−10.9	−12.5	−11.8
24A_24B	−21.8	−12.8	−14.1	−13.5
25A_25B	−25.5	−15.1	−16.7	−14.2
26A_26B	−19.8	−15.6	−17.1	−16.2
Case 9				
27A.o_27B.o	−0.9	−1.0	−0.1	−0.2
27A.o_27C.o	0.4	−1.0	−0.3	−0.2
27A.o_27D.o	13.4	7.1	7.8	7.2
27A.s_27B.s	0.2	0.3	0.2	2.1
27A.s_27C.s	14.1	6.4	4.8	6.5
27A.s_27D.s	16.5	10.9	9.4	11.3
27A.nh_27B.nh	−0.8	−0.4	−0.2	−0.4
27A.nh_27C.nh	1.5	0.6	1.0	0.6
27A.nh_27D.nh	1.8	1.5	1.5	0.5
Case 10				
28A_28B	0.1	0.4	0.6	0.5
29A_29B	3.7	−0.8	−0.1	0.1
30A_30B	−4.0	−4.0	−3.0	−3.5
31A_31B	−2.3	−6.9	−5.8	−5.9
Case 11				
32A_32B	−9.3	−9.6	−8.9	−8.6
33A_33B	11.2	10.0	10.5	10.6
34A_34B	5.1	4.5	5.0	4.7
Case 12				
35C.nch3_35A.nch3	−0.3	2.4	2.8	1.5
35A.nch3_35B.nch3	6.4	5.9	6.4	5.4
35C.o_35A.o	1.8	4.8	5.0	4.2
35A.o_35B.o	9.0	8.9	8.8	8.1
35C.s_35A.s	2.8	4.7	4.9	4.4
35A.s_35B.s	3.2	2.6	2.3	1.8

oxoheterocycles involved in case 12, all three models reasonably agree with one another and predict correctly (in accord with the SAMPL2 manual) that tautomer 35C is dominant, regardless of the substituent (z = N, O, or N–CH_3). With respect to the 35A.x_35B.x tautomerism, our results indicate that solvation is not the major factor in the process, and the ketone form is predicted to be dominant in both the gas phase and aqueous solution, regardless of the nature of the heteroatom z.

Conclusion

The quantum mechanical continuum solvent models SM8, SM8AD, and SMD predict experimental aqueous free energies of solvation (vacuum-water transfer free energies) for 30 out of 31 polyfunctional compounds involved in the SAMPL2 test set with mean unsigned errors of 1.3 (SM8), 2.0 (SM8AD), and 2.6 kcal/mol (SMD). Mean unsigned errors in the free energy differences (tautomeric ratios) for 21 tautomeric pairs in aqueous solution tested in the present study are 2.7 (SM8), 1.8 (SM8AD), and 2.4 kcal/mol (SMD). Given the complex character of the compounds involved in the SAMPL2 challenge as well as in view of the possibility of larger than reported uncertainties in several reference data (especially, for the tautomer ratios), all three models tested in the present work perform well. Nevertheless, further improvement of these models will merit consideration, especially with respect to improving the representation of zwitterionic functionality in existing training sets.

Given our comparison of SM8, SM8AD, and SMD, it is natural to inquire as to whether one is clearly to be preferred over the others. The significant variation in mean unsigned errors over the molecular and tautomeric test sets noted above does not provide a single endorsement, however—SM8 clearly outperforms SM8AD and SMD for the former set while SM8AD does best for the latter. We have already noted that we consider uncertainty in the experimental data to remain too high to consider the observed differences to be decisive. Moreover, the test sets themselves are not especially diverse, so that model differences may be associated more with individual functionalities (e.g., the better performance of SM8AD for oxidized sulfur functionality, which derives from its training on a test set that had better representation of such functional groups compared to SM8 and SMD). In addition, however, there may be other reasons to choose one solvation model over another that are not readily tested by the SAMPL2 data. For example, the SM8 and SM8AD models rely on underlying charge models that are used in the computation of generalized Born electrostatic free energies of solvation, and such models have been defined primarily for a set of balanced, medium-sized basis sets; by contrast, SMD does not depend on a charge model, and thus it may be used with more extended basis sets or with basis sets for which a charge model just does not happen be available. Similarly, by construction, we expect SM8AD to be more accurate for large solutes with complex shapes having very asymmetric charge distributions. And, of course, individual users may have empirically observed

one model to be more effective than another for certain classes of problems, or such observations may have been made in the literature. Thus, rather than recommend a single model as the "best" for all situations, we encourage users to consider all of the above factors when making a choice of SM*x* model and proceed accordingly.

Acknowledgements This work was supported by the Office of Naval Research under Grant N 00014-05-01-0538, the Army Research Office under Grant US ARMY RES LAB/W911NF09-1-0377, and the National Science Foundation (Grant CHE06-10183 and Grant CHE07-04974). Computational resources were provided by the Minnesota Supercomputing Institute.

References

1. Bamborough P, Cohen FE (1996) Modeling protein-ligand complexes. Curr Opin Struct Biol 6(2):236–241
2. Pei J, Wang Q, Zhou J, Lai L (2004) Estimating protein-ligand binding free energy: atomic solvation parameters for partition coefficient and solvation free energy calculation. Proteins 54(4):651–664
3. Schiffer CA, Caldwell JW, Stroud RM, Kollman PA (1992) Inclusion of solvation free energy with molecular mechanics: alanyl dipeptide as a test case. Protein Sci 1(3):396–400
4. Kollman P (1993) Free energy calculations: applications to chemical and biochemical phenomena. Chem Rev 93(7):2395–2417
5. Rivail J-L, Rinaldi DL, Ruiz-Lopez MF (1991) The self-consistent reaction field model for molecular computations in solution. In: Formosinho SJ, Arnaut L, Csizmadia I (eds) Theoretical and computational models for organic chemistry. Kluwer Academic Publishers, Dordrecht, pp 79–92
6. Tomasi J, Persico M (1994) Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent. Chem Rev 94(7):2027–2094
7. Hawkins GD, Zhu T, Li J, Chambers CC, Giesen DJ, Liotard DA, Cramer CJ, Truhlar DG (1998) Universal solvation models. In: Gao J, Thompson MA (eds) Combined quantum mechanical and molecular mechanical methods, American Chemical Society, Symposium Series, vol 712, Washington, pp. 201–219
8. Cramer CJ, Truhlar DG (1999) Implicit solvation models: equilibria, structure, spectra, and dynamics. Chem Rev 99(8): 2161–2200
9. Tomasi J, Mennucci B, Cammi R (2005) Quantum mechanical continuum solvation models. Chem Rev 105(8):2999–3094
10. Mennucci B, Cammi R (eds) (2008) Continuum solvation models in chemical physics: from theory to applications. Wiley, New York
11. Hoijtink GJ, de Boer E, van der Meij PH, Weijland WP (1956) Reduction potentials of various aromatic hydrocarbons and their univalent anions. Recueil des Travaux Chimiques des Pays-Bas et de la Belgique 75:487–503
12. Tucker SC, Truhlar DG (1989) Generalized born fragment charge model for solvation effects as a function of reaction coordinate. Chem Phys Lett 157(1–2):164–170
13. Still WC, Tempczyk A, Hawley RC, Hendrickson T (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. J Am Chem Soc 112(16):6127–6129
14. Marenich AV, Olson RM, Kelly CP, Cramer CJ, Truhlar DG (2007) Self-consistent reaction field model for aqueous and nonaqueous solutions based on accurate polarized partial charges. J Chem Theory Comput 3(6):2011–2033
15. Marenich AV, Cramer CJ, Truhlar DG (2009) Universal solvation model based on the generalized Born approximation with asymmetric descreening. J Chem Theory Comput 5(9): 2447–2464
16. Marenich AV, Cramer CJ, Truhlar DG (2009) Universal solvation model based on solute electron density and a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. J Phys Chem B 113(18):6378–6396
17. Tomasi J, Mennucci B, Cancès E (1999) 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) 464(1):211–226
18. Grycuk T (2003) Deficiency of the coulomb-field approximation in the generalized Born model: an improved formula for Born radii evaluation. J Chem Phys 119(9):4817–4826
19. Storer JW, Giesen DJ, Cramer CJ, Truhlar DG (1995) Class IV charge models: a new semiempirical approach in quantum chemistry. J Comp-Aided Mol Des 9:87–110
20. Kelly CP, Cramer CJ, Truhlar DG (2005) SM6: a density functional theory continuum solvation model for calculating aqueous solvation free energies of neutrals, ions, and solute-water clusters. J Chem Theory Comput 1(6):1133–1152
21. Olson RM, Marenich AV, Cramer CJ, Truhlar DG (2007) Charge Model 4 and intramolecular charge polarization. J Chem Theory Comput 3(6):2046–2054
22. Baldridge K, Klamt A (1997) First principles implementation of solvent effects without outlying charge error. J Chem Phys 106: 6622–6633
23. Marenich AV, Cramer CJ, Truhlar DG (2009) Performance of SM6, SM8, and SMD on the SAMPL1 test set for the prediction of small-molecule solvation free energies. J Phys Chem B 113(14):4538–4543
24. Zhao Y, Truhlar DG (2008) The M06 suite of density functionals for main group thermochemistry, kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06 functionals and twelve other functionals. Theor Chem Acc 120:215–241
25. Francl MM, Pietro WJ, Hehre WJ, Binkley JS, Gordon MS, DeFrees DJ, Pople JA (1982) Self-consistent molecular orbital methods. XXIII. A polarization-type basis set for second-row elements. J Chem Phys 77:3654–3666
26. Hariharan PC, Pople JA (1973) The influence of polarization functions on molecular orbital hydrogenation energies. Theoret Chimica Acta 28(3):213–222
27. Lynch BJ, Zhao Y, Truhlar DG (2003) Effectiveness of diffuse basis functions for calculating relative energies by density functional theory. J Phys Chem A 107(9):1384–1388
28. Bryantsev VS, Diallo MS, Goddard WA (2008) Calculation of solvation free energies of charged solutes using mixed cluster/continuum models. J Phys Chem B 112:9709–9719
29. Lynch BJ, Zhao Y, Truhlar DG (2005) The 6-31B(d) basis set and the BMC-QCISD and BMC-CCSD multicoefficient correlation methods. J Phys Chem A 109(8):1643–1649
30. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Bakken V, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P,

Martin KomaromiI, RL FoxDJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2003) Gaussian03, revision E.01. Gaussian Inc., Pittsburgh
31. MN-GFM (2008) Minnesota gaussian functional module, version 4.1. University of Minnesota, Minneapolis
32. MN-GSM (2009) Minnesota gaussian solvation module, version 2009. University of Minnesota, Minneapolis
33. Halgren TA (1996) Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. J Comp Chem 17(5):490–519
34. PCModel, version 9.1 for Windows, 2006, Serena Software, Bloomington, 47402
35. Adamo C, Barone V (1998) Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: the mPW and mPW1PW models. J Chem Phys 108(2):664–676
36. Easton RE, Giesen DJ, Welch A, Cramer CJ, Truhlar DG (1996) The MIDI! basis set for quantum mechanical calculations of molecular geometries and partial charges. Theor Chim Acta 93(5):281–301
37. Kelly CP, Cramer CJ, Truhlar DG (2006) Adding explicit solvent molecules to continuum solvent calculations for the calculation of aqueous acid dissociation constants. J Phys Chem A 110(7): 2493–2499
38. Ribeiro RF, Marenich AV, Cramer CJ, Truhlar DG (2009) Solvent dependence of ^{14}N nuclear magnetic resonance chemical shielding constants as a test of the accuracy of the computed polarization of solute electronic densities by the solvent. J Chem Theory Comput 5(9):2284–2300
39. Katritzky AR, Øksne S, Boulton AJ (1962) The tautomerism of heteroaromatic compounds with five-membered rings—III: further isoxazol-5-ones. Tetrahedron 18(6):777–790