J Comput Aided Mol Des (2010) 24:281–291
DOI 10.1007/s10822-010-9331-y

# Performance of the IEF-MST solvation continuum model in the SAMPL2 blind test prediction of hydration and tautomerization free energies

**Ignacio Soteras · Modesto Orozco · F. Javier Luque**

Received: 16 November 2009 / Accepted: 10 March 2010 / Published online: 19 March 2010


**Abstract** The IEF-MST continuum solvation model is used to predict the hydration free energies and tautomeric preferences of a set of multifunctional compounds compiled as a blind test for computational solvation methods in the SAMPL2 contest. Computations of hydration free energies was performed using both HF/6-31G(d) and B3LYP/6-31G(d) versions of the IEF-MST model. For tautomeric preferences, the IEF-MST data was combined with the gas phase free energy differences predicted at different levels of theory ranging from MP2/6-31+G(d) to MP2/CBS+[CCSD-MP2/6-31+G(d)] levels. Comparison with the experimental data provided for hydration free energies yields a root-mean square deviation (rmsd) close to 2.3 kcal/mol, which is quite remarkable, especially considering the reduced set of training compounds used in the parametrization of the IEF-MST method. With regard to tautomerism, the lowest error in the prediction of the relative stabilities between tautomers in solution is obtained by combining MP2/CBS+[CCSD-MP2/6-31+G(d)] results with IEF-MST hydration free energies, yielding a rmsd of *ca.* 3.4 kcal/mol. The results illustrate the delicate balance that must be kept between the intrinsic relative stabilities in the gas phase and the differential hydration preferences in order to obtain an accurate description of the prototropic tautomerism in bioorganic compounds.

**Keywords** Continuum solvation model · MST model · Hydration free energy · Tautomerism

I. Soteras · F. J. Luque (✉)
Departament de Fisicoquímica and Institut de Biomedicina (IBUB), Facultat de Farmàcia, Universitat de Barcelona, Av. Diagonal 643, 08028 Barcelona, Spain
e-mail: fjluque@ub.edu

M. Orozco
Molecular Modeling and Bioinformatics Unit, Institut de Recerca Biomèdica, Barcelona Scientific Park, Josep Samitier 1-6, 08028 Barcelona, Spain

M. Orozco
Department of Life Sciences, Barcelona Supercomputing Centre, Jordi Girona 29, 08034 Barcelona, Spain

M. Orozco
Departament de Bioquímica, Facultat de Biologia, Universitat de Barcelona, Av. Diagonal 647, Barcelona 08028, Spain

## Introduction

The accurate prediction of the solvation free energy ($\Delta G_{sol}$; [1]) is one of the current challenges in computational chemistry. This property not only gives information about the thermodynamics of transfer between gas phase and solution, but it is a key parameter to understand the role played by the solvent on chemical processes in solution. It is then not surprising the intense research effort paid to the development of computational strategies to estimate $\Delta G_{sol}$ with *chemical* accuracy. Among those strategies, quantum mechanical self-consistent reaction field (QM-SCRF) methods have become widely popular due to the combination of (1) well defined physical formalisms to couple the charge distribution of the solute with the reaction field of the solvent, (2) the simplicity of the underlying mathematical models, and (3) the low computational cost, which is only slightly greater than the required for a gas phase computation [2–5].

In the framework of QM-SCRF methods, the solute is treated at the QM level and the solvent is represented as a continuum polarizable medium characterized by suitable macroscopic properties. The definition of the boundary

between solute and solvent is one of the most delicate aspects for the computation of $\Delta G_{sol}$, which is estimated by combining electrostatic ($\Delta G_{ele}$) and non-electrostatic ($\Delta G_{n-ele}$) terms. Thus, QM-SCRF methods typically use a molecular-shaped cavity, but there are differences regarding the placement of the boundary (i.e., the set of atomic radii used to build up the cavity and the inclusion of scaling factors that multiply those radii), and the definition of either a single cavity or distinct cavities to evaluate the different contributions to the solvation free energy.

The magnitude of $\Delta G_{ele}$ strongly depends on the definition of the cavity, but also on the level of QM theory used to describe the solute, the dielectric response of the solvent, and the formalism adopted to couple the solute charge distribution and the reaction field induced in the surrounding continuum medium. The dielectric response of the polarizable medium is characterized by a single value of permittivity, which is taken to be equal to the bulk value of pure solvent, and $\Delta G_{ele}$ thus rather accounts for the *bulk-electrostatic* interaction between solute and solvent, neglecting then in principle the contribution of specific solute–solvent interactions. The non-electrostatic component includes a priori a variety of physical contributions, such as the energy cost associated with the creation of the solute cavity, and exchange-repulsion and dispersion interactions between solute and solvent, but in practice it also accounts for any deviation introduced upon assumption of the bulk-electrostatic interaction for $\Delta G_{ele}$. Finally, though both $\Delta G_{ele}$ and $\Delta G_{n-ele}$ should a priori be treated self-consistently [6–9], in practice most QM-SCRF models limit self-consistency to $\Delta G_{ele}$, while $\Delta G_{n-ele}$ depends on nuclear coordinates using either a single term for non-electrostatic contributions or by computing separately cavitation and repulsion-dispersion components [10]. Keeping in mind the preceding considerations, the computational protocol optimized for each QM-SCRF model, which also involves a careful parametrization against experimental data, must be kept for an accurate prediction of $\Delta G_{sol}$ [11, 12].

In order to calibrate the expected accuracy of solvation methods for predicting the hydration free energies of complex bioorganic compounds, Openeye Software recently organized a blind test challenge, named SAMPL1, involving a set of multifunctional molecules quite distinct to those generally considered in the parametrization of theoretical solvation methods [13]. In the current SAMPL2 edition, two blind tests were proposed to the participants (see opening paper in this special issue for further details). The first dataset involves the prediction of the hydration free energy for a series of bioorganic compounds, while the second one faces a new aspect as it is focused on the prediction of the tautomeric preferences in aqueous solution of a wide variety of compounds.

In this contribution we report the results obtained for the two sets of compounds considered in the SAMPL2 contest using the IEF-MST solvation model [14, 15], which relies on the Integral Equation Formalism of the Polarizable Continuum model [16, 17]. To this end, we first give a brief description of the formalism and parametrization adopted in the IEF-MST method. Then, the results are discussed separately for the sets of compounds included in the prediction of hydration free energies and tautomerism.

## Methods

### SAMPL2 dataset

The dataset proposed for the prediction of hydration free energies comprises three distinct subsets: (1) 8 compounds (*explanatory*) with known experimental hydration free energies at the beginning of the contest; (2) 23 compounds (*obscure*) with hydration free energies made available to the participants for 13 cases only after having completed the contest, and (3) finally 10 compounds (*investigatory*) with unknown hydration free energy. 5-Iodouracil was excluded from the set of obscure compounds due to the lack of optimized parameters in the IEF-MST model. Overall, the dataset involves includes 40 compounds and 21 experimentally available hydration free energies.

With regard to the prediction of tautomeric preferences, a wide variety of compounds were considered, including (1) 6-membered aromatic oxoheterocycles (2-pyridone and three related benzofused oxoheterocycles), (2) 4-pyrimidone, (3) 4-phtalazinone, (4) *cis* and *trans* α-diketones, (5) pyrazolones and isoxazolones, (6) xanthine, (7) cyclic *β*-diketones, (8) cyclic lactams and related compounds, (9) cyclic triketones, (10) nitrogen-nitrogen tautomerism in compounds with bridgehead nitrogen, (11) diazepines, and finally (12) five-membered ring 2-oxoheterocycles. Overall, the dataset includes 40 compounds and 110 tautomeric species. In cases (1)–(3) the answer was known quantitatively but not revealed to the participants at the beginning of the contest. In cases (4)–(6), the answer was known and given to the participants, while in the rest of cases quantitative data are mostly absent.

### IEF-MST model

The IEF-MST model computes $\Delta G_{sol}$ from the addition of electrostatic ($\Delta G_{ele},$) cavitation ($\Delta G_{cav}$) and van der Waals ($\Delta G_{vW}$) components. Within the IEF-MST framework, the solvent reaction field induced by the solute charge distribution is modelled as a set of point charges spread over the cavity surface, which are derived by solving Eq. 1.

$$CA^{-1}q = -g \tag{1}$$

where **q** denotes the vector that contains the point charges associated to each surface element after tesselation of the cavity surface into small elements (i.e., tesserae), **A** stands for a diagonal matrix containing the area of the tesserae, **C** is a square matrix that depends on the geometric features of the cavity and of the dielectric permittivity, and **g** is a vector that depends on the solute charge distribution.

$\Delta G_{\text{ele}}$ is determined as shown in Eq. 2, where the index *sol* means that the perturbation operator $\hat{V}_R^{\text{sol}}$ (Eq. 3) is adapted to the fully polarized charge distribution of the solute in solution (given by the wavefunction $\Psi^{\text{sol}}$), and the index *o* stands for the gas phase environment.

$$\Delta G_{\text{ele}} = \left\langle \Psi^{\text{sol}} \middle| \hat{H}^o + \hat{V}_R^{\text{sol}} \middle| \Psi^{\text{sol}} \right\rangle - \frac{1}{2} \left\langle \Psi^{\text{sol}} \middle| \hat{V}_R^{sol} \middle| \Psi^{\text{sol}} \right\rangle - \left\langle \Psi^o \middle| \hat{H}^o \middle| \Psi^o \right\rangle \tag{2}$$

$$\hat{\mathrm{V}}_R^{\text{sol}} = \sum_{j=1}^{M} \frac{q_j^{\text{sol}}}{|r_j - r|} \tag{3}$$

where $M$ stands for the total number of tesserae, each bearing a point charge $q_j^{\text{sol}}$ located at $r_j$.

$\Delta G_{\text{cav}}$ is computed following Claverie-Pierotti's scaled particle theory (Eq. 4) [18, 19]. Finally, $\Delta G_{\text{vW}}$ is computed using a linear relationship to the solvent-exposed surface of each atom (Eq. 5) [20–24], where the atomic surface tensions are determined by fitting to a set of experimental free energies of solvation

$$\Delta G_{\text{cav}} = \sum_{i=1}^{N} w_i \Delta G_{\text{cav},i}^P \tag{4}$$

where $N$ is the number of atoms, $w_i$ is a weighting factor determined from the ratio between the solvent-exposed surface of atom $i$ and the total surface of such an atom, and $\Delta G_{\text{cav},i}^P$ is the cavitation free energy of the sphere associated to atom $i$.

$$\Delta G_{\text{vW}} = \sum_{i=1}^{N} \xi_i S_i \tag{5}$$

where $\xi_i$ and $S_i$ stand for the atomic surface tension and surface of atom $i$.

The IEF-MST model has been implemented at both HF and B3LYP levels using the 6-31G(d) basis set and gas phase optimized geometries [14, 15]. A dual molecular-shaped cavity is used for the calculation of $\Delta G_{\text{ele}}$ and $\Delta G_{n-\text{ele}}$. Thus, the electrostatic term is determined by using a solvent-exposed surface created by scaling the atomic radii by a solvent-dependent factor, which adopts values of 1.25, 1.50, 1.60 and 1.80 for the solvation of neutral compounds in water, octanol, chloroform and carbon tetrachloride, whereas no scaling is used for the non-electrostatic term. Pauling's atomic radii are used for all atoms, except for polar hydrogens, which are treated with an effective radius of 0.9 Å. For the specific case of water as solvent, specific radius (in Å) were assigned to the following atom types: N in –N– (1.4) and $–NH_2$ (1.55); O in –OH (1.45), and S in –SH (1.85). The atomic surface tensions in Eq. 5 were refined by fitting to experimental free energies of hydration for 72 simple, monofunctional compounds.

## Computational details

Present calculations have been performed using a locally modified version of Gaussian-03 [25] without any alteration or refinement of the standard parameters used in the HF and B3LYP IEF-MST models. Additional computations were carried out using readjusted surface tensions for fluorine and chlorine, as noted explicitly in the text (see below). For the set of compounds included the prediction of the hydration free energies, calculations were performed using the geometries distributed by the organizers in order to facilitate comparison with the values provided by other methods. In addition, following the parametrization of the IEF-MST model, computations were also performed using the geometries optimized in the gas phase at both HF/6-31G(d) and B3LYP/6-31G(d) levels.

In the case of those compounds considered for tautomerism, the geometries were fully optimized at the MP2/6-31+G(d) level, and the minimum energy nature of the stationary points was verified by inspection of the vibrational frequencies, which were positive in all cases. Then, the relative energies between tautomers were estimated from single-point computations performed at the MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ, which were in turn used to estimate the energy difference by extrapolation to complete basis set (CBS) using the formalism reported by Park et al. [26]. Finally, higher-order electron correlation effects were accounted for through the difference between CCSD and MP2 calculations performed with the 6-31+G(d) basis. The best gas phase estimate of the energy difference was then derived by combining the MP2/CBS results with the higher-order electron correlation correction (MP2/CBS+[CCSD-MP2/6-31+G(d)]). In addition, ZPE, thermal and entropic corrections were determined using the rigid rotor-harmonic oscillator model [27] using the standard procedure implemented in Gaussian03 in order to obtain the relative stabilities between tautomers in the gas phase. Finally, the tautomerization free energy differences in water (at 298 K) were determined by adding the relative hydration free energies computed using the HF/6-31G(d) and B3LYP/6-31G(d) implementations of the IEF-MST model.

## Results and discussion

### Hydration free energies

The hydration free energies ($\Delta G_{hyd}$) determined from IEF-MST computations at both HF and B3LYP levels using the geometries provided by the organizers of the contest for the whole set of 40 compounds are given in Table 1. There is a slight overestimation of the HF values compared to the B3LYP ones, as noted in a mean-signed error (mse) of 0.5 kcal/mol and in the regression equation $\Delta G_{B3LYP} = 0.93\Delta G_{HF}$ ($r = 0.99$; $p < 0.0001$). However, the differences between HF and B3LYP estimates are quite small, as noted in a root-mean square deviation (rmsd) of 1.1 kcal/mol, which is within the range of uncertainty expected from most experimental measures.

The statistical analysis of the comparison between predicted and experimental data for the subset of 30 compounds is given in Table 2. On average, the IEF-MST results deviate from the experimental values by 0.3 and 0.9 kcal/mol at the HF and B3LYP levels, respectively. Nevertheless, there is not a systematic deviation, as noted in the regression equations $\Delta G_{exp} = 0.94\Delta G_{HF}$ ($r = 0.96$; $p < 0.0001$) and $\Delta G_{exp} = 1.00\Delta G_{B3LYP}$ ($r = 0.96$; $p < 0.0001$) shown in Fig. 1. The root-mean square deviation (rmsd) amounts to 2.5 and 2.3 kcal/mol for the IEF-MST models parametrized at the HF and B3LYP levels, respectively. The rmsd for the subset of *obscure* compounds is smaller (2.1 and 2.0 kcal/mol at the HF and B3LYP levels, respectively), whereas a larger deviation is found for the subset of *explanatory* compounds (3.6 and 3.1 kcal/mol at the HF and B3LYP levels, respectively).

The larger deviation observed for the *explanatory* subset can be mainly attributed to the hydration free energies predicted for glycerol and polyhalogenated compounds. The prediction of the hydration free energy of glycerol was made considering the single geometry provided by the organizers, where the hydroxyl groups are well exposed to the solvent and which was not further explored. For this conformation, the estimated $\Delta G_{hyd}$ values are 4–5 kcal/mol more negative than the experimental value. Nevertheless, formation of a single intramolecular hydrogen-bond has a dramatic effect on the solvation properties, as the hydration free energy is reduced to around −14 kcal/mol (a further 3 kcal/mol reduction in $\Delta G_{hyd}$ is predicted for a conformation containing two intramolecular hydrogen-bonds). Clearly, adoption of a single structure for a flexible compound such as glycerol cannot be representative of the conformational distribution in aqueous solution and can therefore lead to a large error in the estimated hydration free energy, as one should expect a delicate equilibrium between different conformations due to the balance between intramolecular and intermolecular hydrogen bonds [28–30]. Similar considerations could be made to the hydration free energy reported in Table 1 for D-glucose and D-xylose.

A more subtle effect concerns the predicted hydration free energy of polyhalogenated compounds (hexachlorobenzene, hexachloroethane, octafluorocyclobutane and pentachlronitrobenzene). Whereas the IEF-MST $\Delta G_{hyd}$ values of 5-fluorouracil, 5-chlorouracil and 6-chlorouracil are in good agreement with the experimental data, specially for IEF-MST HF/6-31G(d) computations, the difference between predicted and experimental $\Delta G_{hyd}$ values for those polyhalogenated compounds exhibit larger deviations ranging from 2.2 up to 4.7 kcal/mol. This finding can likely be attributed to an inadequacy of the atomic surface tensions of fluorine and chlorine atoms, as the limited number of halogenated compounds considered in the parametrization of the IEF-MST model typically contain a lower number of halogens. In fact, an apparently minor adjustment of the surface tensions parametrized in the IEF-MST model for fluorine ($\xi_F$; from −0.066 to −0.085 kcal mol$^{-1}$ Å$^2$ at the HF level, and from −0.078 to −0.102 kcal mol$^{-1}$ Å$^2$ at the B3LYP level) and chlorine ($\xi_{Cl}$; from −0.112 to −0.130 kcal mol$^{-1}$ Å$^2$ and from −0.112 to −0.131 kcal mol$^{-1}$ Å$^2$ at the HF and B3LYP levels, respectively) suffices to correct the deviations found between estimated and experimental values for polyhalogenated compounds, which in turn stresses the need to consider a set of compounds covering the diversity in chemical space during the refinement of the non-electrostatic components in QM-SCRF continuum models

Based on the preceding considerations, we recomputed the hydration free energies using the geometries of the molecules optimized in the gas phase at the HF and B3LYP levels with the 6-31G(d) basis, thus following the standard protocol adopted in the parametrization of the IEF-MST model (see "Methods"), in conjunction with the readjusted surface tensions for fluorine and chlorine atoms (see Table 1). Upon exclusion of glycerol, D-glucose and D-xylose, the hydration free energies determined for the optimized geometries are generally close to the values calculated for the geometries provided by Openeye Software (see Table 1), as noted in mean-signed deviations of 0.7 and 0.1 kcal/mol at the HF and B3LYP levels, respectively. Moreover, the deviation between theoretical and experimental values amounts to 2.4 (HF) and 2.1 (B3LYP) kcal/mol, which are similar to the values determined for the original geometries (see above and Table 2). These findings indicate a modest effect of the molecular geometry, which agrees with the trends previously found for the set of compounds included in the SAMPL1 dataset [31]. As expected, the largest difference is found for the explanatory compounds, thus reflecting the impact of the readjusted surface tensions for polyhalogenated compounds (see Table 1).

**Table 1** Hydration free energies (kcal/mol) determined from IEF-MST computations at HF/6-31G(d) and B3LYP/6-31G(d) levels

| Dataset | Exp. | IEF-MST[a] | |
|---|---|---|---|
| | | HF | B3LYP |
| Explanatory | | | |
| Hexachlorobenzene | −2.3 | 0.8 (−2.6) | 1.0 (−2.4) |
| Hexachloroethane | −1.4 | 0.8 (−2.6) | 0.9 (−2.6) |
| Octafluorocyclobutane | 3.0 | 7.3 (5.5) | 6.7 (4.8) |
| Trimethyl orthotrifluoroacetate | −0.8 | 1.0 (4.6) | 0.1 (0.1) |
| Trimethyl phosphate | −8.7 | −7.6 (−6.8) | −7.2 (−6.6) |
| 4-Nitroaniline | −10.0 | −10.6 (−9.6) | −11.6 (−11.1) |
| Glycerol | −13.4 | −18.8 (−17.7) | −17.2 (−16.8) |
| Pentachloronitrobenzene | −5.2 | −0.5 (−2.7) | −0.9 (−3.1) |
| Obscure | | | |
| Uracil | −16.6 | −17.0 (−15.2) | −15.4 (−14.3) |
| 5-Bromouracil | −18.2 | −16.4 (−14.5) | −15.1 (−14.2) |
| 5-Chlorouracil | −17.7 | −17.6 (−17.2) | −15.7 (−16.1) |
| 5-Fluorouracil | −16.9 | −16.9 (−14.9) | −14.9 (−14.3) |
| 5-Trifluoromethyluracil | −15.5 | −15.7 (−14.2) | −13.7 (−14.7) |
| 6-Chlorouracil | −15.8 | −14.1 (−14.2) | −12.9 (−14.0) |
| Cyanuric acid | −18.1 | −19.8 (−17.8) | −18.0 (−17.0) |
| Caffeine | −12.6 | −6.4 (−6.7) | −6.9 (−7.3) |
| Methyl paraben | −9.5 | −8.5 (−7.8) | −8.8 (−8.6) |
| Ethyl paraben | −9.2 | −8.2 (−7.6) | −8.6 (−8.4) |
| Propyl paraben | −9.4 | −8.0 (−7.4) | −8.4 (−8.2) |
| Butyl paraben | −8.7 | −7.6 (−7.3) | −8.0 (−8.1) |
| Acetylsalicylic acid | −9.9 | −10.2 (−8.0) | −10.1 (−8.5) |
| Diflunisal | −9.4 | −12.8 (−11.6) | −12.4 (−12.2) |
| Flurbiprofen (racemic) | −8.4 | −8.5 (−6.6) | −8.3 (−5.5) |
| Ibuprofen (racemic) | −7.0 | −7.8 (−6.0) | −7.7 (−6.3) |
| Ketoprofen (racemic) | −10.8 | −11.3 (−8.6) | −11.0 (−9.2) |
| Naproxen | −10.2 | −10.3 (−7.8) | −10.1 (−8.3) |
| Phthalimide | −9.6 | −9.7 (−8.5) | −9.0 (−8.5) |
| Sulfolane | −8.6 | −9.6 (−9.9) | −8.1 (−8.6) |
| D-glucose | −25.5 | −29.6 (−18.4) | −27.7 (−22.5) |
| D-xylose | −20.5 | −23.9 (−16.0) | −22.5 (−17.0) |
| Investigatory | | | |
| Hexamethyl phosphoramide | | −7.3 (−4.7) | −6.8 (−4.6) |
| Dimethyl methylphosphonate | | −9.6 (−8.2) | −8.7 (−8.4) |
| Methyl dimethylphosphinate | | −12.2 (−10.5) | −10.5 (−11.1) |
| Methyl phenyl sulfoxide | | −8.1 (−7.5) | −6.9 (−7.7) |
| Trifluoromethyl phenyl sulfoxide | −2.7 (−2.9) | −1.9 (−3.6) | |
| Methyl phenyl sulfone | | −9.4 (−8.6) | −8.6 (−9.1) |
| Trifluoromethyl phenyl sulfone | | −2.2 (−2.5) | −2.7 (−3.4) |
| Oxazole | | −3.9 (−4.0) | −3.9 (−4.0) |
| Thiazole | | −3.8 (−3.6) | −3.8 (−3.7) |
| Isothiazole | | −4.1 (−3.8) | −3.9 (−4.1) |

[a] Values determined using the geometries provided by the organizers and the original IEF-MST parametrization. The solvation free energies determined for the geometries optimized in the gas phase at the HF and B3LYP levels using readjusted surface tensions for fluorine and chlorine are given in parenthesis

**Table 2** Statistical analysis between predicted and experimental hydration free energies

| | IEF-MST | |
|---|---|---|
| | HF | B3LYP |
| Explanatory | | |
| mse | 1.4 (1.6) | 1.3 (0.6) |
| mue | 2.9 (2.1) | 2.7 (1.3) |
| rmsd | 3.6 (2.8) | 3.1 (1.6) |
| Obscure | | |
| mse | −0.1 (1.5) | 0.7 (1.5) |
| mue | 1.4 (1.9) | 1.4 (1.8) |
| rmsd | 2.2 (2.2) | 2.1 (2.2) |
| All | | |
| mse | 0.3 (1.5) | 0.9 (1.3) |
| mue | 1.8 (1.9) | 1.8 (1.7) |
| rmsd | 2.5 (2.4) | 2.3 (2.1) |

Statistical parameters derived using the geometries provided by the organizers and the original IEF-MST parametrization. The solvation free energies determined for the geometries optimized in the gas phase at the HF and B3LYP levels using readjusted surface tensions for fluorine and chlorine, and excluding glycerol, D-glucose and D-xylose, are given in parenthesis

*mse* mean-signed error, *mue* mean-unsigned error, *rmsd* root-mean square deviation (all values in kcal/mol)

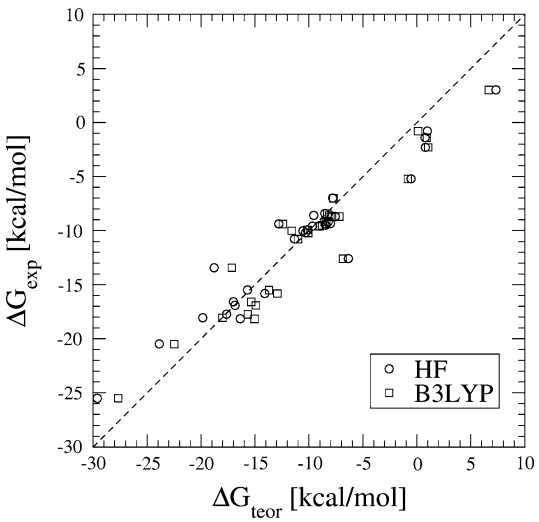


**Fig. 1** Representation of experimental versus theoretical hydration free energies predicted by using the original HF/6-31G(d) and B3LYP/6-31G(d) versions of the IEF-MST model and the molecular geometries provided by the organizers. The ideal regression equation is shown by the *dashed line*

These values compare with the error found for the COSMO-RS model [32] parametrized using the BP functional [33, 34] and the TZVP basis [35, 36], which amounts to 1.4 kcal/mol after having screened the conformational preferences of the compounds using the COSMOconf procedure and considered up to 10 conformers in the prediction of $\Delta G_{hyd}$. On the other hand, uncertainties of 1.9–3.4 kcal/mol are determined for the data computed using the SM8 [37], SM8AD [38] and SMD [39] models with the density functional M06-2X [40] and the 6-31G(d) basis, which were used to compute the hydration free energies for the single geometries provided by the organizers. Overall, these findings point out a similar global accuracy of the different QM-SCRF models, as well as the need to account for a proper description of the solvent influence on the conformational preferences of solutes.

## Tautomerism

The study of tautomeric equilibrium in aqueous solutions is a challenging task due to the need to have accurate estimates of both the intrinsic (gas phase) free energy difference between tautomers as well as of their differential hydration preferences. For the purpose of this challenge we decided to explore the uncertainties in the prediction of the tautomeric preferences by resorting to a variety of methods involving single-point calculations at the MP2 level with extrapolation to complete basis set (using geometries optimized at the MP2/6-31+G(d) level), inclusion of higher-order electron correlation effects from the difference between CCSD and MP2 calculations performed with the 6-31+G(d), and addition of the hydration free energies determined using both HF/6-31G(d) and B3LYP/6-31G(d) optimized versions of the IEF-MST model (see "Methods").

The free energy differences in water between tautomers determined at different levels of theory are reported in Table 3, which includes the whole set of compounds presented by the organizers, though the discussion will be focused only on those compounds where experimental data are available. The only compound not included in Table 3 is xanthine (compound 17), which will be analyzed separately.

The lowest error in the prediction of the relative stabilities is obtained at the MP2/CBS+[CCSD-MP2/6-31+G(d)] level, which yields rmsd values of 3.2 and 3.6 kcal/mol when combined with the HF and B3LYP $\Delta G_{hyd}$ values, respectively (see Table 4). The error for computations carried out at the MP2/6-31 + G(d) level increases up to *ca.* 4.1 kcal/mol, and it is further increased by near 1 kcal/mol for the results determined at the MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ levels, which provide highly similar relative stabilities between tautomers.

The different behavior found for the prediction of the tautomeric preferences of pyrazolones and isoxazolones (compounds 10–16) is quite surprising. Thus, while the tautomerism of 6-membered aromatic heterocycles (compounds 1–6) and α-diketones (compounds 7 and 8) is predicted with good accuracy (i.e., the uncertainty ranges from

**Table 3** Tautomerization free energies (kcal/mol) determined by combining MP2 calculations in the gas phase and hydration free energies determined from IEF-MST computations

| Pair of tautomers | MP2/6-31 + G(d) | | MP2/aug-cc-pVDZ | | MP2/aug-cc-pVTZ | | MP2/CBS + [CCSD-MP2/6-31 + G(d)] | | Exp. |
|---|---|---|---|---|---|---|---|---|---|
| | HF | B3LYP | HF | B3LYP | HF | B3LYP | HF | B3LYP | |
| 1A-1B | −7.0 | −6.1 | −4.3 | −3.4 | −4.1 | −3.3 | −4.6 | −3.7 | −4.8 |
| 2A-2B | −8.3 | −7.9 | −5.2 | −4.9 | −5.2 | −4.9 | −6.3 | −6.0 | −6.1 |
| 3A-3B | −8.9 | −7.9 | −7.0 | −6.0 | −6.9 | −5.9 | −7.7 | −6.7 | −7.2 |
| 4A-4B | −2.3 | −0.3 | 0.3 | 2.2 | 0.5 | 2.5 | 0.6 | 2.5 | −2.3 |
| 5A-5B | −8.0 | −6.9 | −5.3 | −4.2 | −5.2 | −4.1 | −5.6 | −4.5 | −4.8 |
| 5B-5C | 1.1 | 2.2 | 0.4 | 1.5 | 0.3 | 1.4 | 0.0 | 1.2 | 0.5 |
| 6A-6B | −11.1 | −10.3 | −9.0 | −8.2 | −8.9 | −8.1 | −9.8 | −9.0 | −9.2 |
| 6A-6Z | −8.7 | −4.6 | −7.7 | −3.6 | −7.4 | −3.3 | −2.4 | 1.7 | −2.4 |
| 7A-B | 12.0 | 11.9 | 7.8 | 7.7 | 6.8 | 6.7 | 5.5 | 5.4 | 7.0 |
| 8A-8B | 3.1 | 2.6 | −1.0 | −1.5 | −2.6 | −3.1 | −3.5 | −4.1 | −3.0 |
| 10B-10C | 2.4 | 3.2 | 4.9 | 5.6 | 5.2 | 5.9 | 2.2 | 2.9 | −2.9 |
| 10D_10C | 4.7 | 4.7 | 8.0 | 7.9 | 8.2 | 8.2 | 5.0 | 5.0 | −1.2 |
| 11D-11C | 5.1 | 5.2 | 8.2 | 8.3 | 8.6 | 8.8 | 5.5 | 5.6 | −0.5 |
| 12D-12C | 3.5 | 4.3 | 5.5 | 6.3 | 6.1 | 6.8 | 3.0 | 3.7 | −1.8 |
| 13D-13C | 4.2 | 4.2 | 7.4 | 7.4 | 8.2 | 8.1 | 5.2 | 5.2 | 0.1 |
| 14D-14C | 2.4 | 2.8 | 4.2 | 4.5 | 4.2 | 4.6 | 4.0 | 3.7 | 0.3 |
| 15A-15B | 2.9 | 2.0 | −0.9 | −1.7 | −2.3 | −3.1 | 0.9 | 0.1 | 0.9 |
| 15A-15C | 4.2 | 4.1 | 2.6 | 2.5 | 1.3 | 1.2 | 1.4 | 1.3 | −1.2 |
| 15B-15C | 1.4 | 2.1 | 3.5 | 4.2 | 3.6 | 4.3 | 0.5 | 1.2 | −2.2 |
| 16A-16C | 4.9 | 4.5 | 3.0 | 2.6 | 1.4 | 1.0 | 1.5 | 1.2 | 0.5 |
| 19A-19B (R=H) | 5.0 | 4.3 | 1.2 | 0.5 | −0.2 | −0.8 | −0.8 | −1.4 | |
| 20A-20B (Z=CH2) | 4.0 | 3.7 | −1.3 | −1.6 | −2.8 | −3.1 | −2.9 | −3.1 | |
| 20A-20B (Z=O) | 6.0 | 5.1 | 1.6 | 0.8 | 0.1 | −0.8 | −0.6 | −1.4 | |
| 20A-20B (Z=NCH3) | 10.6 | 8.4 | 6.4 | 4.3 | 5.1 | 2.9 | 4.3 | 2.2 | |
| 20A-20B (Z=S) | 4.0 | 3.5 | −1.4 | −1.9 | −3.0 | −3.4 | −3.4 | −3.9 | |
| 21A_21B | 17.9 | 17.0 | 13.9 | 13.0 | 12.5 | 11.6 | 11.4 | 10.5 | |
| 22A-22B | −14.9 | −14.1 | −12.8 | −12.0 | −12.6 | −11.8 | −12.0 | −11.2 | |
| 23A-23B | −16.0 | −15.3 | −14.0 | −13.2 | −13.9 | −13.2 | −13.1 | −12.4 | |
| 24A-24B | −18.8 | −19.6 | −14.7 | −15.6 | −14.8 | −15.6 | −14.1 | −14.9 | |
| 25A-25B | −20.0 | −20.8 | −16.1 | −16.9 | −16.3 | −17.0 | −15.7 | −16.4 | |
| 26A-26B | −20.3 | −20.1 | −18.2 | −18.0 | −18.3 | −18.1 | −17.5 | −17.3 | |
| 27A-27B (Z=NH) | −0.3 | −1.3 | −0.2 | −1.1 | −0.3 | −1.2 | −0.4 | −1.3 | |
| 27A_27C | 1.5 | 0.6 | 1.6 | 0.7 | 1.7 | 0.8 | 1.5 | 0.7 | |
| 27A_27D | 1.5 | 0.6 | 1.5 | 0.6 | 1.3 | 0.4 | 1.5 | 0.6 | |
| 27A-27B (Z=O) | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.0 | −0.1 | |
| 27A-27C | −0.1 | −0.2 | −0.3 | −0.4 | −0.2 | −0.3 | −0.3 | −0.4 | |
| 27A-27D | 8.1 | 8.2 | 8.0 | 8.2 | 8.4 | 8.5 | 7.9 | 8.1 | |
| 27A-27B (Z=S) | 0.7 | 0.7 | 0.9 | 1.0 | 0.9 | 0.9 | 0.8 | 0.8 | |
| 27A-27C | 4.8 | 5.8 | 4.4 | 5.4 | 4.8 | 5.8 | 4.3 | 5.3 | |
| 27A-27D | 9.5 | 10.6 | 9.6 | 10.7 | 9.6 | 10.7 | 9.0 | 10.1 | |
| 28A_28B | −1.1 | −0.5 | −1.4 | −0.8 | −1.3 | −0.7 | −0.9 | −0.3 | |
| 29A_29B | −3.5 | −2.6 | −3.8 | −2.9 | −3.6 | −2.7 | −3.2 | −2.3 | |
| 30A-30B | −3.9 | −4.0 | −3.8 | −4.0 | −3.7 | −3.9 | −3.4 | −3.6 | |
| 31A-31B | −8.9 | −8.7 | −9.2 | −9.0 | −8.9 | −8.7 | −8.0 | −7.8 | |
| 32A-32B | −8.0 | −8.6 | −9.9 | −10.5 | −11.4 | −12.0 | −9.4 | −10.0 | |
| 33A-33B | 11.4 | 11.1 | 10.1 | 9.8 | 8.9 | 8.6 | 2.4 | 2.1 | |

**Table 3** continued

| Pair of tautomers | MP2/6-31 + G(d) | | MP2/aug-cc-pVDZ | | MP2/aug-cc-pVTZ | | MP2/CBS + [CCSD-MP2/6-31 + G(d)] | | Exp. |
|---|---|---|---|---|---|---|---|---|---|
| | HF | B3LYP | HF | B3LYP | HF | B3LYP | HF | B3LYP | |
| 34A_34B | 6.5 | 6.2 | 4.7 | 4.4 | 3.4 | 3.1 | 0.4 | 0.1 | |
| 35C-35A (Z=NCH3) | 0.7 | 0.6 | 0.8 | 0.7 | 0.7 | 0.6 | 1.0 | 0.9 | |
| 35A-35B | 8.7 | 8.1 | 4.0 | 3.4 | 2.8 | 2.2 | 5.5 | 4.9 | |
| 35C-35A (Z=O) | 4.6 | 4.2 | 4.9 | 4.4 | 4.6 | 4.2 | 4.3 | 3.9 | |
| 35A-35B | 11.7 | 11.0 | 7.6 | 6.8 | 6.1 | 5.3 | 7.6 | 6.9 | |
| 35C-35A (Z=S) | 4.6 | 4.5 | 5.3 | 5.2 | 5.0 | 4.9 | 4.3 | 4.2 | |
| 35A-35B | 5.6 | 4.9 | 1.4 | 0.7 | −0.8 | −1.4 | 1.2 | 0.6 | |

See opening article of this issue for representation of the compounds

**Table 4** Statistical analysis between predicted and experimental tautomerization free energies determining by combining MP2 calculations in the gas phase and hydration free energies determined from IEF-MST computations

| | MP2/6-31 +G(d) | | MP2/aug-cc-pVDZ | | MP2/aug-cc-pVTZ | | MP2/CBS+[CCSD-MP2/6-31+G(d)] | |
|---|---|---|---|---|---|---|---|---|
| | HF | B3LYP | HF | B3LYP | HF | B3LYP | HF | B3LYP |
| Compounds 1–8 | | | | | | | | |
| mse | −0.6 | 0.5 | 0.1 | 1.2 | 0.0 | 1.0 | −0.2 | 0.9 |
| mue | 2.9 | 2.3 | 1.3 | 1.4 | 1.1 | 1.3 | 0.8 | 1.5 |
| rmsd | 3.8 | 2.9 | 2.1 | 1.9 | 2.0 | 1.9 | 1.2 | 2.3 |
| Compounds 10–16 | | | | | | | | |
| mse | 4.4 | 4.5 | 5.4 | 5.6 | 5.3 | 5.4 | 3.7 | 3.8 |
| mue | 4.4 | 4.5 | 5.8 | 6.1 | 5.9 | 6.2 | 3.7 | 4.0 |
| rmsd | 4.8 | 5.0 | 6.7 | 7.0 | 6.9 | 7.2 | 4.5 | 4.7 |
| Compounds 1–16 | | | | | | | | |
| mse | 1.9 | 2.5 | 2.8 | 3.4 | 2.6 | 3.2 | 1.8 | 2.4 |
| mue | 3.6 | 3.4 | 3.6 | 3.8 | 3.5 | 3.7 | 2.2 | 2.7 |
| rmsd | 4.2 | 4.0 | 4.8 | 5.0 | 5.0 | 5.2 | 3.2 | 3.6 |

*mse* mean-signed error, *mue* mean-unsigned error, *rmsd* root-mean square deviation (all values in kcal/mol)

1.2 to 2.3 kcal/mol at the MP2/CBS+[CCSD-MP2/6-31+G(d)] level), predictions are worse (error of ~4.6 kcal/mol) when the prototropic tautomerism involves the five-membered rings of compounds 10–16. This distinct trend is clearly reflected in the representation of the experimental versus calculated relative stabilities (see Fig. 2). Thus, the values determined for compounds 1–8 lie close to the ideal regression line, as noted in the regression equations $\Delta G_{exp} = 0.97\Delta G_{HF}$ ($r = 0.97$; $p < 0.0001$) and $\Delta G_{exp} = 0.98\Delta G_{B3LYP}$ ($r = 0.91$; $p < 0.0001$). In contrast, the values determined for the tautomerism of pyrazolones and isoxazolones exhibit a clear deviation from the ideal regression line.

Analysis of the results suggests that a major contribution for the large error in compounds 10–16 stems from the gas phase contribution, since correlation effects seem to be very strong for these molecules, as noted in the fact that the estimated values obtained from MP2/aug-cc-pVDZ and MP2/aug-cc-pVTZ levels calculations are up to 3 kcal/mol worse than those obtained with our best estimate (MP2/CBS+[CCSD-MP2/6-31+G(d)]; see Table 4), while the difference was very moderate for compounds 1–8.

This finding suggests a particularly significant role of high-order electron correlation effects in the tautomeric equilibria of pyrazolones and isoxazolones. In fact, when one considers the gas phase free energy differences determined directly from CCSD/6-31+G(d) computations in conjunction with the IEF-MST $\Delta G_{hyd}$ values (see Tables 5, 6), the error for compounds 1-16 (*ca.* 3.5 kcal/mol) is similar to that obtained at the MP2/CBS+[CCSD-MP2/6-31+G(d)] level, but is also internally consistent for the two subsets of compounds (1–8 and 10–16). Then, while the strategy of combining the MP2/CBS energies with the higher-order electron correlation correction through the difference between CCSD and MP2 computations with the 6-31+G(d) basis yields a small deviation (between 1 and 2 kcal/mol) from the experimental data for compounds 1–8, CCSD/6-31+G(d) computations provide estimates of relative stabilities between tautomers for compounds 10–16 that are closer to the experimental data. Moreover, even though the uncertainty obtained for CCSD/6-31+G(d) computations in the case of compounds 10–16 is still large (around 3.6 kcal/mol; see Table 6), a significant

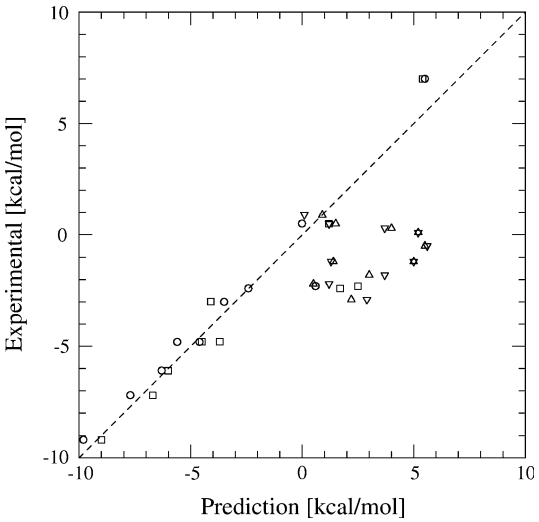


**Fig. 2** Representation of the experimental tautomerization free energies in water versus the theoretical values determined at the MP2/CBS + [CCSD-MP2/6-31*G(d)] level in conjunction with the hydration free energies predicted by using HF/6-31G(d) (*open circle*) and B3LYP/6-31G(d) (*square*) versions of the IEF-MST model. Compounds with five-membered rings are shown separately (HF: *triangle* B3LYP: *inverted triangle*). The ideal regression equation is shown by the *dashed line*

**Table 5** Tautomerization free energies and error (kcal/mol) predicted at the CCSD/6-31+G(d) level combined with hydration free energies determined from IEF-MST computations

| Pair of tautomers | CCSD/6-31+G(d) | | Exp. |
|---|---|---|---|
| | HF | B3LYP | |
| 1A-1B | −7.6 | −6.7 | −4.8 |
| 2A-2B | −9.4 | −9.0 | −6.1 |
| 3A-3B | −9.8 | −8.8 | −7.2 |
| 4A-4B | −2.3 | −0.4 | −2.3 |
| 5A-5B | −8.5 | −7.4 | −4.8 |
| 5B-5C | 0.9 | 2.0 | 0.5 |
| 6A-6B | −12.0 | −12.2 | −9.2 |
| 6A-6Z | −3.9 | 0.2 | −2.4 |
| 7A-B | 11.3 | 12.2 | 7.0 |
| 8A-8B | 3.1 | 2.6 | −3.0 |
| 10B-10C | −0.7 | 0.0 | −2.9 |
| 10D_10C | 1.4 | 1.3 | −1.2 |
| 11D-11C | 1.7 | 1.8 | −0.5 |
| 12D-12C | 0.1 | 0.9 | −1.8 |
| 13D-13C | 0.7 | 0.7 | 0.1 |
| 14D-14C | 0.7 | 1.0 | 0.3 |
| 15A-15B | 6.9 | 6.1 | 0.9 |
| 15A-15C | 5.0 | 4.9 | −1.2 |
| 15B-15C | −1.9 | −1.2 | −2.2 |
| 16A-16C | 5.9 | 5.6 | 0.5 |

See opening article of this issue for representation of the compounds

**Table 6** Statistical analysis between predicted and experimental tautomerization free energies determined from CCSD/6-31+G(d) computations combined with hydration free energies determined from IEF-MST computations

| | CCSD/6-31+G(d) | |
|---|---|---|
| | HF | B3 LYP |
| Compounds 1–8 | | |
| mse | −0.6 | 0.5 |
| mue | 2.8 | 2.9 |
| rmsd | 3.4 | 3.6 |
| Compounds 10–16 | | |
| mse | 2.8 | 2.9 |
| mue | 2.8 | 2.9 |
| rmsd | 3.7 | 3.6 |
| Compounds 1–16 | | |
| mse | 1.1 | 1.7 |
| mue | 2.8 | 2.9 |
| rmsd | 3.5 | 3.4 |

*mse* mean-signed error, *mue* mean-unsigned error, *rmsd* root-mean square deviation (all values in kcal/mol)

contribution arises from the tautomeric equilibria 15A-15B, 15A-15C and 16A-16C, which involve the prototropic tautomerism between NH/OH groups and an aromatic carbon, where the error between predicted and experimental value amounts to 5–6 kcal/mol. Exclusion of these cases then reduces the error to around 2 kcal/mol at the combined CCSD/6-31+G(d)+IEF-MST level.

The error found here for our best composite strategy (MP2/CBS+[CCS-MP2/6-31+G(d)]+IEF-MST) compares with the deviations found for the values determined with other QM-SCRF methods. COSMO-RS has been used with the intrinsic relative stabilities determined at (1) the BP/TZVP level or (2) from MP2 computations with the QZVPP basis [41, 42] using BP-TZVP geometries. The deviation from the experimental data given in Table 3 amounts to 3.3 and 3.0 kcal/mol, respectively. The BP-TZVP + COSMO-RS values deviate significantly for the subset of compounds 1–8 (3.9 kcal/mol), and there is a notable reduction when the MP2/QZVPP values are considered, leading to an error (1.6 kcal/mol) slightly larger than that obtained for our best composite approach (1.2 kcal/mol). Regarding the SMx models, geometries were reoptimized at the M06-2X level with the MG3S basis [43], which was also used to derive gas phase free energies. Combination with SM8, SM8AD and SMD models yields errors in the tautomerization free energies in aqueous solution of 3.6, 2.6 and 3.1 kcal/mol, which are close to the errors reported here for our best composite method. Again, prediction of the relative stabilities for compounds 1–8 (errors ranging from 1.7 to 2.5 kcal/mol) is more accurate than for compounds 10–16 (errors comprised between 3.3 and 4.6 kcal/mol).

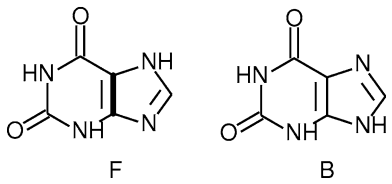


**Fig. 3** Main tautomeric forms of xanthine in aqueous solution

Overall, the present discussion illustrates the difficulty to predict with chemical accuracy the different classes of prototropic tautomerism in bioorganic compounds, as it is necessary to obtained a balanced, accurate description of the intrinsic relative stabilities in the gas phase as well as of the differential hydration preferences (for recent examples see refs. [44–46]).

To end this section, we discuss the tautomerism of xanthine, which has a large number of possible tautomeric species arising from the keto-enol tautomerism in the six-membered ring and the prototropic tautomerism between the nitrogen atoms in the five-membered ring. According to our calculations, only two species (F and B in Fig. 3) are predicted to be populated more than 1% in aqueous solution. Thus, the relative population of tautomer F is predicted to amount to 79.4%, and the population estimated for tautomer B is 20.6% when HF $\Delta G_{hyd}$ values are used. Instead, those contributions become 91.4% and 8.6% for tautomers F and B when the B3LYP $\Delta G_{hyd}$ values are considered. In the two cases, therefore, tautomer F is found to be the main species, though a significant amount of tautomer B is also predicted. These findings agree well with the experimental information provided by the SAMPL2 organizers, as the estimated fractions of tautomers F and B amount to 99.7% and *ca.* 0.3%, while excluding the presence of any other tautomeric form in aqueous solution.

## Conclusion

Prediction of tautomerism in bioorganic compounds is a challenging issue with clear implications in areas such as drug design [47], where elucidation of the population of tautomers, each with its own physico-chemical properties, is crucial to rationalize the interaction of a bioorganic compound with a macromolecular target. In this framework, the organization of blind test such as SAMPL2 is extremely helpful to genuinely calibrate the reliability of theoretical methods to calculate tautomer ratios, which demands a delicate balance between the intrinsic relative stabilities of tautomers and the differential stabilization arising from interaction with the solvent environment.

This contribution has examined the suitability of the IEF-MST model, a parametrized version of the IEF-PCM formalism, which was refined at both HF and B3LYP levels using the 6-31G(d) basis set by fitting to atomistic simulations (for cavity definition) and experimental measures (solvatochromic effects and solvation free energies). Comparison with the experimental hydration free energies collected for a subset of 21 polyfunctional compounds in the SAMPL2 contest yields a rmsd of *ca.* 2.3 kcal/mol, which compare with the uncertainties obtained in the preceding SAMPL1 contest [31]. Improvement of the predictive capabilities of the method could come from refinement of the parameters optimized in the fitting to experimental data, and from the inclusion of different conformations in the case of flexible compounds by developing efficient strategies for the exploration of the conformational preferences of compounds in solution.

With regard to tautomerism, the lowest error in the prediction of the relative stabilities between tautomers is obtained at the MP2/CBS+[CCSD-MP2/6-31+G(d)] level, which yields a rmsd of *ca.* 3.4 kcal/mol. Nevertheless, the results also stress the difficulty to keep a balanced, accurate description of the intrinsic relative stabilities of tautomers and of their hydration preferences able to describe with general validity the prototropic tautomerism in chemically diverse compounds. Overall, present computations, in conjunction with those performed by other colleagues in this issue, will provide valuable clues to advance in our definition of the best computational strategies to predict the tautomeric preferences in different solvent environments.

**Acknowledgments** We acknowledge the financial support from the Ministerio de Innovación y Ciencia (MICINN; SAF2008-05595) and Generalitat de Catalunya (2009-SGR00298), and the computational facilities provided by the Centre de Supercomputació de Cataluyna (CESCA) and the Barcelona Supercomputing Center (BSC) are acknowledged.

## References

1. Ben-Naim A (1987) Solvation Thermodynamics. Plenum Press, New York
2. Cramer CJ, Truhlar DG (1999) Chem Rev 99:2161
3. Orozco M, Luque FJ (2000) Chem Rev 100:4187
4. Luque FJ, Curutchet C, Muñoz-Muriedas J, Bidon-Chanal A, Soteras I, Morreale A, Gelpí JL, Orozco M (2003) Phys Chem Chem Phys 18:3827
5. Tomasi J, Mennucci B, Cammi R (2005) Chem Rev 105:2999
6. Rinaldi D, Ruiz-López MF, Martins Costa MTC, Rivail JL (1986) Chem Phys Lett 128:177
7. Olivares del Valle FJ, Aguilar MA (1993) J Mol Struct (THEOCHEM) 25(9): 99
8. Amovilli C, Mennucci B (1997) J Phys Chem B 101:1051
9. Borström M, Ninham BW (2004) J Phys Chem B 108:12593
10. Curutchet C, Orozco M, Luque FJ, Mennucci B, Tomasi J (2006) J Comput Chem 27:1769
11. Curuchet C, Cramer CJ, Truhlar DG, Ruiz-López MF, Rinaldi D, Orozco M, Luque FJ (2003) J Comput Chem 24:284

12. Klamt A, Mennucci B, Tomasi J, Barone V, Curutchet C, Orozco M, Luque FJ (2009) Acc Chem Res 42:489
13. Guthrie JP (2009) J Phys Chem B 113:4501
14. Curutchet C, Orozco M, Luque FJ (2001) J Comput Chem 22:1180
15. Soteras I, Curutchet C, Bidon-Chanal A, Orozco M, Luque FJ (2005) THEOCHEM 727:29
16. Cancès E, Mennucci B, Tomasi J (1997) J Chem Phys 107:3032
17. Mennucci B, Cancès B, Tomasi J (1997) J Phys Chem B 101:10506
18. Pierotti RA (1976) Chem Rev 76:717
19. Claverie P, Daudey JP, Langlet J, Pullman B, Piazzola D, Huron MJ (1978) J Phys Chem 82:405
20. Luque FJ, Bachs M, Orozco M (1994) J Comput Chem 15:111
21. Bachs M, Luque FJ, Orozco M (1994) J Comput Chem 15:446
22. Orozco M, Bachs M, Luque FJ (1995) J Comput Chem 16:563
23. Luque FJ, Zhang Y, Alemán C, Bachs M, Gao J, Orozco M (1996) J Phys Chem 100:4269
24. Luque FJ, Alemán C, Bachs M, Orozco M (1996) J Comput Chem 17:806
25. 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, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill, P MW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2003) Gaussian03, Rev. B.04. Gaussian Inc., Pittsburgh
26. Sun YP, Soon BH, Jae SL (2002) J Mol Struct (THEOCHEM) 586: 81
27. McQuerrie D (1973) Statistical Mechanics. Harper & Brown, New York
28. Luque FJ, López JM, López de la Paz M, Vicent C, Orozco M (1998) J Phys Chem A102:6690
29. Hernández B, Curutchet C, Colominas C, Orozco M, Luque FJ (2002) Mol Simul 28:153
30. Butler K, Luque FJ, Barril X (2009) J Comput Chem 30:601
31. Soteras I, Forti F, Orozco M, Luque FJ (2009) J Phys Chem B 113:9330
32. Klamt A, Jonas V, Burger T, Lohrenz JCW (1998) J Phys Chem 102:5074
33. Vosko SH, Wilk L, Nusair M (1980) Can J Phys 58:1200
34. Becke AD (1988) Phys Rev A 38:3098
35. Schaefer A, Huber C, Ahlrichs I (1994) J Chem Phys 100:5829
36. Eichkorn K, Weigend F, Treutler O, Ahlrichs R (1997) Theor Chem Acc 97:119
37. Marenich AV, Olson RM, Kelly CP, Cramer CJ, Truhlar DG (2009) J Chem Theory Comput. 3: 2011
38. Marenich AV, Cramer CJ, Truhlar DG (2009) J Chem Theory Comput (to be submitted)
39. Marenich AV, Cramer CJ, Truhlar DG (2009) J Phys Chem B 113:6378
40. Zhao Y, Truhlar DG (2008) Theor Chem Acc 120:215
41. Weigend F, Furche F, Ahlrichs R (2003) J Chem Phys 119:12753
42. Hättig C (2005) Phys Chem Chem Phys 7:59
43. Lynch BJ, Zhao Y, Truhlar DG (2003) J Phys Chem A 107:1384
44. Farrera JA, Canal I, Hidalgo-Fernandez P, Perez-Garcia ML, Huertas O, Luque FJ (2008) Chem Eur J 14:2277
45. Blas JR, Luque FJ, Orozco M (2004) J Am Chem Soc 126:154
46. Huertas O, Blas JR, Soteras I, Orozco M, Luque FJ (2006) J Phys Chem A 110:510
47. Martin YC (2009) J Comput Aided Mol Design 23:693