J Comput Aided Mol Des (2010) 24:343–353
DOI 10.1007/s10822-010-9340-x

Prediction of tautomer ratios by embedded-cluster integral equation theory

Stefan M. Kast · Jochen Heil · Stefan Güssregen · K. Friedemann Schmidt

Received: 7 December 2009 / Accepted: 16 March 2010 / Published online: 30 March 2010

Abstract The "embedded cluster reference interaction site model" (EC-RISM) approach combines statistical-mechanical integral equation theory and quantum-chemical calculations for predicting thermodynamic data for chemical reactions in solution. The electronic structure of the solute is determined self-consistently with the structure of the solvent that is described by 3D RISM integral equation theory. The continuous solvent-site distribution is mapped onto a set of discrete background charges ("embedded cluster") that represent an additional contribution to the molecular Hamiltonian. The EC-RISM analysis of the SAMPL2 challenge set of tautomers proceeds in three stages. Firstly, the group of compounds for which quantitative experimental free energy data was provided was taken to determine appropriate levels of quantum-chemical theory for geometry optimization and free energy prediction. Secondly, the resulting workflow was applied to the full set, allowing for chemical interpretations of the results. Thirdly, disclosure of experimental data for parts of the compounds facilitated a detailed analysis of methodical issues and suggestions for future improvements of the model. Without specifically adjusting parameters, the EC-RISM model yields the smallest value of the root mean square error for the first set (0.6 kcal mol^{-1}) as well as for the full set of quantitative reaction data (2.0 kcal mol^{-1}) among the SAMPL2 participants.

Keywords Solvation model · Quantum chemistry · 3D RISM theory · Cluster embedding · Tautomer ratio

S. M. Kast (✉) · J. Heil
Theoretische Physikalische Chemie, Technische Universität Dortmund, Otto-Hahn-Str. 6, 44227 Dortmund, Germany
e-mail: kast@keychem.de

S. M. Kast · J. Heil
Eduard-Zintl-Institut für Anorganische und Physikalische Chemie, Technische Universität Darmstadt, Petersenstr. 20, 64287 Darmstadt, Germany

S. Güssregen · K. F. Schmidt
Sanofi-Aventis Deutschland GmbH, R&D CAS Drug Design FFM, Industriepark Hoechst, 65926 Frankfurt am Main, Germany

Introduction

In order to predict the standard Gibbs free energy change for a general reaction $A_{solv} \rightarrow B_{solv}$ involving two solvated molecules we combine quantum-chemical calculations with a solvation model based on three-dimensional "reference interaction site model" (3D RISM) integral equation theory [1]. Briefly, the method determines the electronic structure of the solute self-consistently with the structure of the solvent which is calculated by 3D RISM theory [2, 3]. The 3D RISM approach describes approximately the spatial distribution of solvent particles around a solute molecule using atomic models for both, in contrast to continuum solvation models. In this way, H bonding effects are accounted for explicitly. The solvent polarization affecting the solute's electronic wave function is modeled by mapping the continuous solvent site distribution onto a set of discrete background charges, i.e., by constructing an "embedding cluster" (EC) giving rise to the acronym "EC-RISM" for the procedure.

The tautomer set of the SAMPL2 challenge offers an excellent opportunity for calibrating and testing the EC-RISM approach. Following the separation into three compound classes provided by Taylor et al. [4] this paper is organized as follows: After describing the theoretical and

computational background, results for the group of compounds with quantitatively known experimental free energy data are reported in order to determine appropriate levels of quantum-chemical theory for geometry optimization and free energy prediction, and for testing the influence of nonbonded solute–solvent parameters. Consequently, the optimized workflow is applied to the full set, allowing for chemical interpretations of the results. Finally, the predictions are analyzed *ex post facto* with respect to methodical issues in light of the disclosure of experimental data for parts of the compounds.

Methods

Theory

The EC-RISM methodology benefits from the availability of closed-form expressions for the excess chemical potential μ^{ex} of the solute in a given conformation, which are functionals of solvent particle distribution functions, for a wide range of levels of theory [1, 5]. Following our earlier derivation [1], the Gibbs free energy for a reaction involving only a single conformation for A_{solv} and B_{solv} is written as

$$\Delta G = \Delta E_{\mathrm{solv}} + \Delta\mu^{\mathrm{ex}} \tag{1}$$

where ΔX denotes $X(B_{\mathrm{solv}}) - X(A_{\mathrm{solv}})$ and E_{solv} represents the per-particle (intensive) quantum-mechanical intramolecular energy in solution. μ^{ex} is the free energy change associated with transferring the solute (kept rigid and with solution-phase wave function) from the gas phase into solution at infinite dilution. If multiple conformations $C_{i,j}$ participate in the reaction, a discrete partition function approximation for species C_i can be invoked:

$$G(C_i) = -RT \ln \sum_j \exp[(-E_{\mathrm{solv}}(C_{i,j}) - \mu^{\mathrm{ex}}(C_{i,j}))/RT] \tag{2}$$

where R is the molar gas constant and T the absolute temperature. In this context, we neglect vibrational and rotational contributions to the partition function as well as pressure–volume work. If the calculations are performed at infinite dilution in a solvent environment at the standard pressure of 1 bar ΔG is equivalent to the standard Gibbs free energy ΔG^0 which, via

$$K = \exp(-\Delta G^0/RT), \tag{3}$$

is related to thermodynamic equilibrium constant.

The quantum-mechanical energy of a solute conformation composed of nuclei (n) and electrons (e) embedded in a cluster of point charges (q) is derived from the Hamiltonian

$$\hat{H}_{\mathrm{tot}} = \hat{H}_1 + \hat{H}_2 \tag{4}$$

where

$$\hat{H}_1 = \hat{H}_{ne} + \hat{H}_{ee} + \hat{H}_{nn} \tag{5}$$

represents pure solute properties and

$$\hat{H}_2 = \hat{H}_{nq} + \hat{H}_{eq} + \hat{H}_{qq} \tag{6}$$

describes the molecule–charge and charge–charge interactions. We are interested in

$$\begin{aligned} E_{\mathrm{solv}} &= \langle\psi_{\mathrm{tot}}|\hat{H}_1|\psi_{\mathrm{tot}}\rangle = E_{\mathrm{tot}} - E_2 \\ &= \langle\psi_{\mathrm{tot}}|\hat{H}_{\mathrm{tot}}|\psi_{\mathrm{tot}}\rangle - \langle\psi_{\mathrm{tot}}|\hat{H}_2|\psi_{\mathrm{tot}}\rangle \end{aligned} \tag{7}$$

with the wave function ψ_{tot} calculated for the embedded system, i.e., we have to subtract

$$E_2 = \langle\psi_{\mathrm{tot}}|\hat{H}_{nq} + \hat{H}_{eq}|\psi_{\mathrm{tot}}\rangle + \langle\psi_{\mathrm{tot}}|\hat{H}_{qq}|\psi_{\mathrm{tot}}\rangle = E_q + E_{\mathrm{self}} \tag{8}$$

from the results of embedded cluster quantum-chemical calculations. The explicit evaluation of the last term in Eq. 8, the self energy of the background charges, is not necessary since it cancels upon subtracting E_2 from E_{tot}. Its computation should be avoided since it scales with the square of the number of charges which is typically of the order 10^6. The first term represents physically the solute–solvent electrostatic interaction energy. It can be written as

$$E_q = \int \mathrm{d}\mathbf{r}\rho_q(\mathbf{r})\varphi(\mathbf{r}) \tag{9}$$

with the quantum-mechanical electrostatic potential of the solute φ at spatial points $\mathbf{r}$ and the charge density ρ_q. The latter is related to the pair distribution function g_γ between solute and solvent site γ by

$$\rho_q(\mathbf{r}) = \sum_\gamma q_\gamma \rho_\gamma g_\gamma(\mathbf{r}) \tag{10}$$

where ρ_γ and q_γ are the bulk solvent site density and site charge, respectively. The value of a discrete background charge at $\mathbf{r}_i$ that contributes to $\hat{H}_2$ results from integration over the corresponding grid point volume ΔV approximately as

$$q(\mathbf{r}_i) = \rho_q(\mathbf{r}_i)\Delta V \tag{11}$$

which leads to

$$E_q \approx \sum_i \sum_\gamma q_\gamma \rho_\gamma g_\gamma(\mathbf{r}_i)\varphi(\mathbf{r}_i)\Delta V = \sum_i q(\mathbf{r}_i)\varphi(\mathbf{r}_i). \tag{12}$$

The determination of E_{solv} by the preceding formalism requires as input the solute–solvent pair distribution functions g that are obtained by 3D RISM calculations on a grid surrounding the solute molecule. The distribution functions are related to the total correlation functions

$h = g - 1$ which are connected to the direct correlation functions c by

$$\rho_\gamma h_\gamma(\mathbf{r}) = \sum_{\gamma'} c_{\gamma'} \times \chi_{\gamma\gamma'}(\mathbf{r}). \tag{13}$$

The solvent site–site susceptibility χ is computed only once for a given solvent model at specified thermodynamic conditions and is typically obtained from one-dimensional dielectrically corrected DRISM theory [6, 7], the star denotes convolution. A second relation, a so-called closure is needed for finding solutions to the system of integral equations. Here we use the a third order partial series expansion (PSE-3) of the hypernetted chain (HNC) closure which has been shown to yield quantitatively very good results whereas numerical difficulties are efficiently avoided [5]:

$$h_\gamma(\mathbf{r}) = \begin{cases} \sum_{i=0}^{3} (t_\gamma^*(\mathbf{r}))^i / i! - 1 & \Leftrightarrow \quad t_\gamma^* > 0 \\ \exp(t_\gamma^*(\mathbf{r})) - 1 & \Leftrightarrow \quad t_\gamma^* \leq 0. \end{cases} \tag{14}$$

The so-called renormalized indirect correlation function

$$t_\gamma^*(\mathbf{r}) = h_\gamma(\mathbf{r}) - c_\gamma(\mathbf{r}) - \beta u_\gamma(\mathbf{r}) \tag{15}$$

($\beta = (k_B T)^{-1}$, k_B is the Boltzmann constant) contains the solute–solvent site interaction u_γ, which consists of apolar as well as Coulombic contributions φ, as discussed above. The HNC approximation corresponds to unconditional application of the second line of Eq. 14. Starting from a vacuum wave function, the EC-RISM iteration cycle proceeds by calculating the solute–solvent interaction on the 3D grid, solving the integral Eqs. 13–15, computation of the charge density by Eq. 10, mapping of ρ_q onto discrete background charges according to Eq. 11 which in turn enter the solute Hamiltonian (4)–(6). These steps are repeated until self-consistency between electronic and solvent structure is achieved. After convergence, μ^{ex} in Eq. 1 is computed by evaluating the PSE-3 expression [5]

$$\mu^{ex} = \beta^{-1} \sum_\gamma \rho_\gamma \times \int d\mathbf{r} \left[\frac{1}{2} h_\gamma^2(\mathbf{r}) - c_\gamma(\mathbf{r}) - \frac{1}{2} h_\gamma(\mathbf{r}) c_\gamma(\mathbf{r}) - \Theta(h_\gamma(\mathbf{r}))(t_\gamma^*(\mathbf{r}))^4/4! \right] \tag{16}$$

where Θ is the Heaviside step function. E_{solv} finally follows from Eqs. 7–12.

Computational details

The quantum-chemical calculations were performed by Gaussian 03 [8], the EC-RISM code was developed by the Kast group at the Technische Universität Darmstadt. The solvent susceptibility χ was calculated using the DRISM/HNC [6, 7] method for a modified SPC/E water model [9] (Lennard-Jones (LJ) parameter σ of hydrogen set to 1.0 Å [10, 11]) at $T = 298.15$ K, a bulk density of $\rho = 0.03334$ Å^{-3}, and with the dielectric constant set to 78. The equations were solved with the modified direct inversion of iterative subspace (MDIIS) technique [12] on a logarithmically spaced grid [13, 14] ranging from 0.0059 to 164.02 Å using the convergence criterion that the maximum residual of the direct correlation functions should decay below 10^{-8}. Solute–solvent interactions were described by the sum of LJ (using standard Lorentz-Berthelot mixing rules) and Coulomb terms. Solute LJ parameters were taken from the general Amber force field (GAFF) [15], whereas the Coulomb part was approximated by atom-centered point charges derived by fitting to the electrostatic potential originating from nuclei and wave function in the Hartree–Fock approximation using the CHelpG scheme [16]. For periodic 3D RISM calculations, the electrostatic potential was treated by Ewald summation. The integral equations were solved on rectangular grids of adjustable dimension with a uniform grid spacing of 0.3 Å. The grid size was determined by adding 12.5 Å in each direction to the site position with the largest distance to the molecule's centroid during EC-RISM iterations, and adding 15 Å in the last step. The MDIIS method [12] was applied during iterations up to a convergence threshold of 10^{-6} in the direct correlation functions. The temperature was set to 298.15 K. 3D integrals occurring in free energy evaluations were approximated by discrete sums over the grid. For the EC-RISM iterations, we used the total change in $E_{tot} + \mu^{ex}$ as convergence criterion, which should fall below 0.01 kcal mol^{-1} between two subsequent cycle steps. The expensive calculation of E_q by Eq. 12 was performed in the last step only. We have checked in preliminary calculations that the target value $E_{solv} + \mu^{ex}$ is equally well converged in this way.

The general workflow for processing the SAMPL2 compounds can be summarized as follows:

- Starting with the structure files that were provided [4], all chemically meaningful conformations including OH rotamers (dihedral angles formed by rotatable hydroxyl groups and the remaining molecular scaffold) were sampled exhaustively. For OH rotamers, this was achieved by generating input structures for subsequent geometry optimization in 90° increments of the dihedral angle.
- Geometry optimization was performed *in vacuo* using B3LYP/6-311++G(d,p) density functional theory calculations as well as under the influence of the polarizable continuum solvation model (PCM in water with the UA0 united atom topological model) at the same level of theory. Resulting unique conformations were extracted for subsequent EC-RISM treatment.
- As shown earlier [1], very accurate results can be expected if electron correlation is accounted for by

MP2 (frozen-core) perturbation theory. This method was applied throughout for EC-RISM calculations.

- The methodology was calibrated by variation of conditions for case (5), compounds **10–16** (in the nomenclature and notation of [4]). We examined the basis sets 6-311++G(d,p) and aug-cc-pVDZ for both vacuum and solution structures, the importance of partition functions over rotameric states, as well as slight variations of certain LJ parameters provided by GAFF. The most successful approach that emerged from these investigations was then chosen for treating the remaining compounds.

Results and discussion for the pre-disclosure phase

Preliminaries

The following notation is used throughout: ΔG^0_{exp} is the experimental value for the standard Gibbs reaction free energy derived from thermodynamic equilibrium constants, as provided [4]. In case (5), compound **10**, an estimate for the experimental reference value for **10B** → **10D** has been reconstructed by a thermodynamic cycle. ΔG^0_{min} represents the computational prediction if only the global conformational/rotameric minimum is taken into account, ΔG^0_Z indicates the result of conformational averaging by a discrete partition function according to Eq. 2. $\Delta\Delta G^0_{\text{min}/Z}$ denotes the difference between calculated and experimental values. For PCM calculations, nonpolar contributions to the free energy are neglected. Additionally, statistical data are reported including root mean square errors (RMSE) between theory and experiment and R^2 coefficients of determination as well as a linear regression analysis (slope m and intercept b) for the independent/dependent variable pairs (ΔG^0_{exp}, $\Delta G^0_{\text{min}/Z}$). For the final model, we provide the quantum-mechanical (ΔE_{solv}) and solvational ($\Delta\mu^{\text{ex}}$) contributions separately for the global conformational minima as well as the vacuum energy difference (ΔE_{vac}). Despite the limited accuracy of the experimental equilibrium constant data, we specify two decimals in units of kcal mol^{-1} in order to compare subtle computational energy differences of that order of magnitude.

Model calibration for explanatory compounds, case (5)

Table 1 shows EC-RISM results for different basis sets applied to the vacuum structures, using the original GAFF parameters for modeling LJ interactions. It is evident that the agreement between prediction and experiment is unsatisfactory for these settings. Statistical data show in general no significant correlation except for those pairs for which hydroxyl or carbonyl groups are retained upon reaction (**10B/D**, **16A/C**). Single-point PCM calculations (MP2/6-311++G(d,p)) on some of these compounds (data not shown) confirm this trend. It is therefore very likely that structural issues related to vacuum artifacts are responsible for the deviations.

Consequently, we reoptimized the structures under PCM(water) conditions, which also led to the stabilization of certain OH rotamers that are not found *in vacuo*. The

Table 1 Results of MP2/EC-RISM/PSE-3 calculations on vacuum structures optimized at B3LYP/6-311++G(d,p) for different basis sets; ΔG, RMSE, b in kcal mol^{-1}

Theory level		MP2/6-311++G(d,p)				MP2/aug-cc-pVDZ			
Reaction	ΔG^0_{exp}	ΔG^0_{min}	$\Delta\Delta G^0_{\text{min}}$	ΔG^0_Z	$\Delta\Delta G^0_Z$	ΔG^0_{min}	$\Delta\Delta G^0_{\text{min}}$	ΔG^0_Z	$\Delta\Delta G^0_Z$
10B → 10C	−2.87	0.54	3.41	0.54	3.41	0.53	3.40	0.53	3.40
10D → 10C	−1.23	2.66	3.89	2.78	4.01	2.78	4.01	2.90	4.13
10B → 10D	−1.64	−2.12	−0.48	−2.24	−0.60	−2.25	−0.61	−2.37	−0.73
11D → 11C	−0.55	2.76	3.31	2.96	3.51	2.67	3.22	2.86	3.41
12D → 12C	−1.78	1.84	3.62	2.04	3.82	1.80	3.58	1.96	3.74
13D → 13C	0.14	3.66	3.52	3.95	3.81	–	–	–	–
14D → 14C	0.27	0.92	0.65	1.14	0.87	0.99	0.72	0.99	0.72
15B → 15C	−2.18	0.36	2.54	0.37	2.55	–	–	–	–
15A → 15C	−1.23	−0.76	0.47	−0.76	0.47	–	–	–	–
15A → 15B	0.86	−1.12	−1.98	−1.12	−1.98	–	–	–	–
16A → 16C	0.55	1.27	0.72	1.27	0.72	1.51	0.96	1.51	0.96
Statistics	RMSE	2.61		2.72					
	R^2	0.02		0.03					
	m	0.22		0.25					
	b	1.10		1.21					

Table 2 Results of PCM optimizations at B3LYP/6-311++G(d,p) and subsequent PCM single-point (s.p.) calculations at MP2/aug-cc-pVDZ; ΔG, RMSE, b in kcal mol^{-1}

Theory level		B3LYP/6-311++G(d,p)				MP2/aug-cc-pVDZ (s.p.)			
Reaction	ΔG^0_{exp}	ΔG^0_{min}	$\Delta\Delta G^0_{min}$	ΔG^0_Z	$\Delta\Delta G^0_Z$	ΔG^0_{min}	$\Delta\Delta G^0_{min}$	ΔG^0_Z	$\Delta\Delta G^0_Z$
10B → 10C	−2.87	−3.00	−0.13	−2.91	−0.04	2.88	5.75	2.99	5.86
10D → 10C	−1.23	−0.76	0.47	−0.61	0.62	5.12	6.35	5.34	6.57
10B → 10D	−1.64	−2.25	−0.61	−2.30	−0.66	−2.23	−0.59	−2.35	−0.71
11D → 11C	−0.55	−0.15	0.40	−0.01	0.54	5.61	6.16	5.82	6.37
12D → 12C	−1.78	−0.68	1.10	−0.66	1.12	4.85	6.63	4.87	6.65
13D → 13C	0.14	−0.06	−0.20	0.07	−0.07	5.86	5.72	6.04	5.90
14D → 14C	0.27	−1.10	−1.37	−1.10	−1.37	3.50	3.23	3.51	3.24
15B → 15C	−2.18	−3.18	−1.00	−3.17	−0.99	2.80	4.98	2.85	5.03
15A → 15C	−1.23	−1.19	0.04	−1.19	0.04	−0.14	1.09	−0.14	1.09
15A → 15B	0.86	1.99	1.13	1.98	1.12	−2.94	−3.80	−2.99	−3.85
16A → 16C	0.55	−2.12	−2.67	−2.12	−2.67	0.54	−0.01	0.54	−0.01
Statistics	RMSE	1.10		1.11		4.66		4.76	
	R^2	0.45		0.44		0.03		0.03	
	m	0.82		0.82		−0.42		−0.43	
	b	−0.41		−0.38		1.97		2.03	

PCM results are summarized in Table 2. If the free energy is evaluated at the energy minimum with the same level of theory used during optimization, the results are fairly satisfactory. However, the deviation of the intercept from zero indicates a systematic error component. More worrisome is the fact that, for a given structure, a generally more reliable level of theory (MP2/aug-cc-pVDZ) leads to drastically worse agreement between theory and experiment, leaving little room for systematic improvements. Here, as before, we find that conformational averaging over rotameric states shows no clear tendency towards better predictions.

Using the PCM/B3LYP/6-311++G(d,p)-optimized structures during EC-RISM calculations leads, however, to a considerable improvement as compared to the vacuum case reported above. Table 3 shows the results for MP2/6-311++G(d,p) taking into account the original GAFF LJ parameters (GAFF-P_0) as well as a slightly increased LJ parameter σ (+0.1 Å) of the hydroxyl-O (GAFF-P_1). Clearly, using solution phase-optimized structures is essential for obtaining reliable results. Even with the original GAFF parameter set (GAFF-P_0) reasonable agreement is achieved. The slightly modified GAFF-P_1 set yields considerably better values, as indicated by all statistical descriptors, most notably by a slope near one, a high degree of linear correlation, and small RMSE value of around 0.5 kcal mol^{-1}. The optimal LJ parameter for hydroxyl-O has been found by systematically scanning variations of σ(carbonyl-O) and σ(hydroxyl-O) in the vicinity of GAFF-P_0. For GAFF-P_1 we find that the maximum deviation is less than 1 kcal mol^{-1} and, again, no advantage of conformational averaging can be identified. Furthermore, it should be mentioned that it is critically important to find the globally optimal OH rotamer. If, for instance, the energetically *highest* rotamer is selected as being representative for the free energy, GAFF-P_1 yields an RMSE of 2.01 kcal mol^{-1} and a slope of 2.13, although a value of 0.81 for R^2 indicates significant correlation.

Finally, GAFF-P_0 and GAFF-P_1 were applied in EC-RISM calculations with the larger, correlation-consistent basis set aug-cc-pVDZ. The results are summarized in Table 4 and illustrated in Fig. 1. Judging from the statistical analysis and the maximum deviation, both sets perform quite well and could in principle be used for production calculations. On one hand, this justifies the attempt to calibrate adjustable parameters with a smaller basis. On the other hand, the fact that merely switching to a (typically better) basis set leads to better agreement without parameter adjustments emphasizes the soundness of the EC-RISM approach.

Only with the original, unmodified GAFF-P_0 we are, however, able to predict the correct sign of the reaction free energy for all cases. Together with the corresponding slope near one and the small value for the intercept, we conclude that MP2/aug-cc-pVDZ/EC-RISM/PSE-3 with GAFF-P_0 on globally optimal structures found by B3LYP/6-311++G(d,p)/PCM (i.e., without conformational averaging) is the best and most consistent procedure for treating the remaining cases. All data and interpretations reported in what follows are based on this methodology which contains no specifically adjusted parameters. As a (crude) measure of standard error, we determine the maximum deviation of values obtained from the three best strategies,

Table 3 Results of MP2/6-311++G(d,p)/EC-RISM/PSE-3 calculations on solution structures (B3LYP/6-311++G(d,p)/PCM) for GAFF-P_0 and GAFF-P_1; ΔG, RMSE, b in kcal mol^{-1}

Theory level		MP2/6-311++G(d,p)/GAFF-P_0				MP2/6-311++G(d,p)/GAFF-P_1			
Reaction	ΔG^0_{exp}	ΔG^0_{min}	$\Delta\Delta G^0_{min}$	ΔG^0_Z	$\Delta\Delta G^0_Z$	ΔG^0_{min}	$\Delta\Delta G^0_{min}$	ΔG^0_Z	$\Delta\Delta G^0_Z$
10B → 10C	−2.87	−2.61	0.26	−2.60	0.27	−3.78	−0.91	−3.78	−0.91
10D → 10C	−1.23	−0.45	0.78	−0.36	0.87	−1.67	−0.44	−1.53	−0.30
10B → 10D	−1.64	−2.15	−0.51	−2.24	−0.60	−2.12	−0.48	−2.25	−0.61
11D → 11C	−0.55	−0.19	0.36	−0.02	0.53	−1.36	−0.81	−1.14	−0.59
12D → 12C	−1.78	−0.38	1.40	−0.23	1.55	−1.49	0.29	−1.29	0.49
13D → 13C	0.14	1.53	1.39	1.82	1.68	0.38	0.24	0.72	0.58
14D → 14C	0.27	0.15	−0.12	0.33	0.06	−0.70	−0.97	−0.56	−0.83
15B → 15C	−2.18	−1.04	1.14	−1.04	1.14	−2.14	0.04	−2.14	0.04
15A → 15C	−1.23	−1.37	−0.14	−1.37	−0.14	−1.37	−0.14	−1.37	−0.14
15A → 15B	0.86	−0.32	−1.18	−0.32	−1.18	0.77	−0.09	0.77	−0.09
16A → 16C	0.55	0.40	−0.15	0.40	−0.15	0.40	−0.15	0.40	−0.15
Statistics	RMSE	0.84		0.93		0.53		0.52	
	R^2	0.58		0.55		0.89		0.88	
	m	0.73		0.76		1.04		1.07	
	b	0.06		0.16		−0.27		−0.17	

Table 4 Results of MP2/aug-cc-pVDZ/EC-RISM/PSE-3 calculations on solution structures (B3LYP/6-311++G(d,p)/PCM) for GAFF-P_0 and GAFF-P_1; ΔG, RMSE, b in kcal mol^{-1}

Theory level		MP2/aug-cc-pVDZ/GAFF-P_0				MP2/aug-cc-pVDZ/GAFF-P_1			
Reaction	ΔG^0_{exp}	ΔG^0_{min}	$\Delta\Delta G^0_{min}$	ΔG^0_Z	$\Delta\Delta G^0_Z$	ΔG^0_{min}	$\Delta\Delta G^0_{min}$	ΔG^0_Z	$\Delta\Delta G^0_Z$
10B → 10C	−2.87	−2.84	0.03	−2.83	0.04	−3.93	−1.06	−3.91	−1.04
10D → 10C	−1.23	−0.55	0.68	−0.45	0.78	−1.69	−0.46	−1.54	−0.31
10B → 10D	−1.64	−2.29	−0.65	−2.38	−0.74	−2.24	−0.60	−2.38	−0.74
11D → 11C	−0.55	−0.39	0.16	−0.23	0.32	−1.49	−0.94	−1.27	−0.72
12D → 12C	−1.78	−0.79	0.99	−0.60	1.18	−1.82	−0.04	−1.58	0.20
13D → 13C	0.14	0.81	0.67	1.09	0.95	−0.27	−0.41	0.07	−0.07
14D → 14C	0.27	0.16	−0.11	0.32	0.05	−0.63	−0.90	−0.51	−0.78
15B → 15C	−2.18	−1.88	0.30	−1.88	0.30	−2.90	−0.72	−2.90	−0.72
15A → 15C	−1.23	−1.87	−0.64	−1.87	−0.64	−1.87	−0.64	−1.87	−0.64
15A → 15B	0.86	0.02	−0.84	0.01	−0.85	1.03	0.17	1.03	0.17
16A → 16C	0.55	0.56	0.01	0.56	0.01	0.57	0.02	0.56	0.01
Statistics	RMSE	0.57		0.66		0.64		0.59	
	R^2	0.77		0.73		0.94		0.94	
	m	0.89		0.91		1.17		1.19	
	b	−0.05		0.05		−0.36		−0.25	

aug-cc-pVDZ/GAFF-P_0, aug-cc-pVDZ/GAFF-P_1, and 6-311++G(d,p)/GAFF-P_1 calculations. This error represents certainly a lower limit at best; it can be used for characterizing the relative sensitivity towards variation within the EC-RISM approach, not the absolute accuracy of the method. All data are summarized in Table 5 (which additionally includes total free energy predictions for conformational averages, separate contributions ΔE_{solv} and $\Delta\mu^{ex}$ for the global minima, as well as the global minima energy difference of the gas phase optimization).

Remarks on obscure compounds, cases (1)–(3)

Case (1): The lactam form **B** is favored for all compounds **1**–**4** by an overcompensation of the electronic part by the solvation contribution; the relative stabilization increases

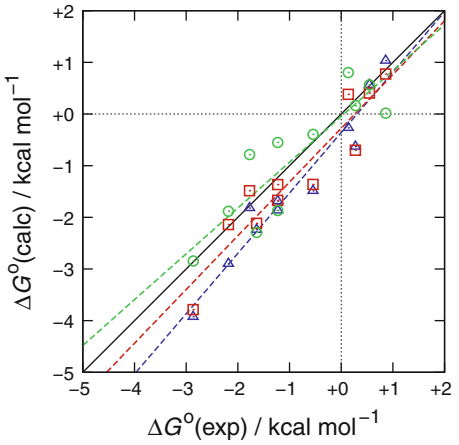

Fig. 1 Calculated vs. experimental standard reaction free energies for explanatory case (5), compounds **10**–**16** for the optimal models (global minima); the *solid black line* indicates perfect correlation. The structures of all compounds were optimized by B3LYP/6-311++G(d,p)/PCM and post-processed by MP2/aug-cc-pVDZ/EC-RISM/PSE-3 (GAFF-P_1) (*blue triangles*), MP2/6-311++G(d,p)/EC-RISM/PSE-3 (GAFF-P_1) (*red squares*), MP2/aug-cc-pVDZ/EC_RISM/PSE-3 (GAFF-P_0) (*green circles*). Regression lines are shown in the color of the data points. The latter level of theory has been used for production calculations, error estimates were determined from the maximum deviation between the models

in the direction **1B** < **2B** < **3B**. The effect of varying positions of the cyclic amide function on the free energy is large, corresponding to a range of ca. 6.6 kcal mol^{-1}. **4B** is relatively least favored, dominated by E_{solv}; it is also the only compound for which a pure gas phase calculation would lead to a preference of the iminol form **4A**. Apparently, the aromatic system of the fused rings is more strongly perturbed in the case of **4B** in comparison with the others.

Case (*2*): The stability of **5B** and **5C** is very similar. **5C** is slightly favored whereas **5A** is strongly disfavored—in this case the vacuum result shows the opposite trend, as is also seen for E_{solv}. Solvation effects are therefore decisive for this system. We note a dramatic E_{solv}/μ^{ex} compensation for **5B** → **5C** which is explained by a large change of the dipole moment (4.9 D → 15.0 D) leading to strong solvent sensitivity of K_m. The solvent-induced stability of **5B** over **5A** is more difficult to explain. The calculations reveal that the two OH rotamers of **5A** are almost equally populated, having dipole moments of 6.0 D and 2.5 D, respectively. On average, the dipole moment can be estimated to be around 4.2 D which is less than the 4.9 D found for **5B**.

Case (*3*): Here the stability decreases in the order **6B** > **6Z** > **6A**, **6A** does practically not exist. **6B** (dipole moment 7.9 D) is favored by both E_{solv} and μ^{ex} whereas the preference of **6Z** (15.6 D) over **6A** (7.6 D) is clearly a solvation effect explaining the incorrect prediction by the gas phase result. **6Z** therefore exists due to charge separation. On the other hand, **6B** and **6Z** are structurally quite similar. The bond lengths are for C(O)–N 1.31 Å (**6A**), 1.38 Å (**6B**), and 1.37 Å (**6Z**), i.e., shortest for **6A**, whereas the C–O bonds are 1.34 Å (**6A**), 1.24 Å (**6B**), and 1.25 Å (**6Z**), i.e., longest for **6A**. The C–O group of **6Z** therefore preserves the double-bond character.

Remarks on explanatory compounds, cases (4)–(6)

Case (*4*): The calculated tendency agrees semiquantitatively with the experimental estimate for **8** (the enol **B** is favored [4]), and more quantitatively for **7** (the diketone **A** is strongly preferred). **8A** is disfavored by conformational enforcement of parallel dipoles as illustrated by both E_{vac} and E_{solv} which overcompensates μ^{ex} (dipole moments for **8B**: 5.9 D, **8A**: 8.9 D). In contrast, we find for **7A** *trans*-oriented carbonyl groups in the dominant conformer. This leads to a vanishing dipole moment and almost perfect agreement of the vacuum energy difference with the reaction free energy in solution. A parallel alignment of carbonyl and hydroxyl groups is observed for both enol forms.

Case (*5*): The keto form **10C** is most stable, dominated by solvation which overcompensates the energetic term. The latter should favor the aromatic enols **10B** and **10D**. For the compounds **11**, **12**, and **13** we find that **11C** and **12C** are slightly favored (μ^{ex} overcompensates E_{solv}), but, in contrast, **13D** is preferred. In this case the differential solvation term which favors the keto form is not sufficiently large for overcompensating E_{solv}. This is probably related to the presence of the hydrophobically hydrated, bigger phenyl group which increases the solvent cavity in comparison to **11** and **12**. The polar contribution to the solvation free energy which discriminates between keto and enol forms becomes therefore less important. **14D** is slightly favored and has a dipole moment of 4.5 D (**14C**: 6.7 D) which explains why it is more preferred in apolar solvents. The free energy prediction for **15A** → **15B** has a comparatively large absolute error which becomes smaller for a smaller basis set after reparametrization of the hydroxyl-O LJ parameter (see Table 3). The hydroxyl group in **15B** is apparently critically positioned and sensitive to partial solvent shielding by the phenyl ring. The reaction **16A** → **16C** is very well described, accompanied by a small error and appears to be barely influenced by the choice of the basis set (Tables 3, 4). It is characterized by almost perfect E_{solv}/μ^{ex} compensation.

Case (*6*): The calculations correctly identify **17F** and **17B** as the only relevant tautomers, the free energies of the next nearest variants indicate negligible contribution to the mixture (**17C**: +8.5 kcal mol^{-1}, **17G**: +9,1 kcal mol^{-1}, **17A**: +12.6 kcal mol^{-1}). However, the **17F**/**17B** preference is less clear than expected from the experimental data

Table 5 Results of MP2/aug-cc-pVDZ/EC-RISM/PSE-3 (GAFF-P_0) calculations on solution structures (B3LYP/6-311++G(d,p)/PCM); vacuum results ΔE_{vac} (B3LYP/6-311++G(d,p)), ΔG, error, ΔE, $\Delta\mu^{ex}$ in kcal mol^{-1}

Reaction	ΔG^0_{min}	ΔG^0_Z	Error	ΔE_{solv}	$\Delta\mu^{ex}$	ΔE_{vac}
1A → 1B	−7.73	−7.57	1.15	5.93	−13.65	−0.88
2A → 2B	−9.66	−9.29	0.82	9.00	−18.67	−4.51
3A → 3B	−11.17	−11.12	0.73	6.57	−17.74	−6.22
4A → 4B	−4.57	−4.43	1.10	17.56	−22.13	4.65
5A → 5B	−6.16	−5.83	1.00	−0.08	−6.09	−2.01
5B → 5C	−0.51	−0.51	0.16	24.75	−25.26	9.64
6A → 6B	−11.15	−11.12	0.58	−8.12	−3.02	−13.65
6A → 6Z	−6.72	−6.69	1.04	32.60	−39.32	6.58
7A → 7B	5.11	4.71	1.78	5.02	0.09	5.02
8A → 8B	−1.01	−1.38	1.63	−6.94	5.93	−4.60
10B → 10C	−2.84	−2.83	1.08	15.04	−17.88	−0.86
10D → 10C	−0.55	−0.45	1.13	13.95	−14.50	2.85
11D → 11C	−0.39	−0.23	1.09	14.18	−14.57	4.11
12D → 12C	−0.79	−0.60	1.03	10.22	−11.01	2.82
13D → 13C	0.81	1.09	1.07	12.14	−11.33	3.96
14D → 14C	0.16	0.32	0.87	7.79	−7.63	−0.12
15A → 15B	0.02	0.01	1.02	1.23	−1.22	6.37
15A → 15C	−1.87	−1.87	0.50	13.03	−14.90	4.88
15B → 15C	−1.88	−1.88	1.02	11.80	−13.68	−1.49
16A → 16C	0.56	0.56	0.16	8.97	−8.40	2.37
17F → 17B[a]	0.07	0.07	0.10	19.65	−19.58	9.12
19A_h → 19B_h	−1.08	−1.25	1.25	6.08	−7.16	1.83
20A_ch2 → 20B_ch2	−4.50	−4.60	1.65	4.42	−8.92	0.47
20A_o → 20B_o	−1.92	−1.94	1.19	7.26	−9.18	3.65
20A_s → 20B_s	−2.66	−2.72	1.89	4.39	−7.05	2.12
20A_nch3 → 20B_nch3	3.36	3.31	1.14	8.63	−5.27	7.56
21A → 21B	13.92	13.86	0.96	17.60	−3.68	15.96
22A → 22B	−16.77	−16.57	1.20	−14.27	−2.51	−11.89
23A → 23B	−17.59	−17.33	0.84	−4.23	−13.36	−13.86
24A → 24B	−14.41	−14.35	1.02	−19.55	5.15	−18.04
25A → 25B	−14.24	−14.19	1.10	−23.24	8.99	−19.64
26A → 26B	−17.72	−17.68	1.03	−24.05	6.33	−21.78
27A_o → 27B_o	0.38	0.33	0.15	−0.45	0.84	−0.45
27A_o → 27C_o	0.49	0.50	0.48	0.66	−0.17	1.54
27A_o → 27D_o	5.11	5.12	0.42	4.06	1.05	4.37
27A_s → 27B_s	0.59	0.59	0.16	0.99	−0.40	0.63
27A_s → 27C_s	−0.27	−0.27	0.20	−0.25	−0.02	0.56
27A_s → 27D_s	3.02	3.02	0.66	4.93	−1.91	4.31
27A_nh → 27B_nh	−0.68	−0.67	0.06	−1.74	1.06	−1.22
27A_nh → 27C_nh	0.65	0.67	0.17	2.21	−1.56	2.34
27A_nh → 27D_nh	1.38	1.39	0.16	4.35	−2.98	2.64
28A → 28B	1.67	1.67	0.04	0.28	1.39	0.75
29A → 29B	−2.15	−2.15	0.25	10.56	−12.71	5.05
30A → 30B	−3.31	−3.31	0.07	−3.94	0.63	−3.39
31A → 31B	−7.78	−7.78	0.36	−0.61	−7.16	−1.20
32A → 32B	−10.67	−10.67	0.40	−1.07	−9.61	−10.51
33A → 33B	11.07	11.07	0.04	16.88	−5.81	9.43
34A → 34B	5.02	5.02	0.71	10.98	−5.96	2.95

Table 5 continued

Reaction	ΔG^0_{min}	ΔG^0_Z	Error	ΔE_{solv}	$\Delta\mu^{ex}$	ΔE_{vac}
35A_o → 35B_o	9.91	9.85	1.01	5.62	4.29	10.78
35C_o → 35A_o	6.29	6.29	0.51	−0.45	6.74	3.04
35A_s → 35B_s	1.82	1.47	0.99	0.28	1.55	7.46
35C_s → 35A_s	6.56	6.56	0.68	1.27	5.29	4.08
35A_nch3 → 35B_nch3	8.46	8.12	0.98	−1.57	10.03	10.64
35C_nch3 → 35A_nch3	3.19	3.19	0.24	−2.25	5.44	0.57

[a] In solution only **17B** and **17F** contribute to the mixture, the energies of all others are at least 8.5 kcal mol^{-1} higher

[4]. We find almost equal population of **17F** and **17B** with a very slight advantage for the former. This statement is accompanied by a very small error estimate, i.e., a weak dependence on the choice of the computational EC-RISM model. Zero-point energy corrections derived from the harmonic frequencies under vacuum conditions also show negligible impact. The equilibrium between **17F** and **17B** is characterized by substantial E_{solv}/μ^{ex} compensation. E_{solv} favors **17F** (as cited in the reference [4]) while μ^{ex} favors **17B** due to a strong dipole moment difference (**17B**: 13.3 D, **17F**: 7.6 D).

Remarks on investigatory compounds, cases (7)–(12)

Case (*7*): The enol **19B** is preferred as a result of solvation effects, **20B** is also solvent-dominated and favored in the cases **ch2**, **o**, and **s** whereas, in contrast, **20A_nch3** is found to be dominant. The latter case is the consequence of the electronic energy which cannot be overcompensated by the solvation term. More pronounced, the tautomers of **21** are clearly determined by electronic effects. Enolization is practically absent despite the solvation term since the structure would become antiaromatic, as can be deduced from the site charge of ca. +1 on the carbonyl C.

Case (*8*): For compounds **22**–**26** all **B** (lactam/imide) isomers are strongly favored, typically as a consequence of E_{solv} and also in agreement with the gas phase energies. μ^{ex} shows a less clear trend. Particularly for **23** (similarly but less pronounced also for **22**) the solvation term has the same direction as the electronic part. Here, the dipole moment is primarily due to the presence of only one carbonyl group. Note that case (8) contrast sharply with case (7), except only for **20** and, more clearly, **21**.

Case (*9*): Solvent effects are small throughout for these compounds. E_{solv} dominates the tendency as suggested [4], except for **27A_o → 27B_o**. In general, the reaction free energies for **27A → 27B** and **27A → 27C** are small, implying almost equal population. For **27A → 27D** the **A** form is favored since **D** is electronically disadvantaged by enolization of cyclic ester, thioester, or amide groups, in good agreement with computed gas phase data.

Case (*10*): The **B** forms are successively more favorable in the order **28** > **29** > **30** > **31**. In absolute terms, **28A** is slightly preferred. This result corresponds with the acidity of N in five-membered rings which increases with increasing number of heteroatoms corresponding to increasing electron deficiency. Note that we find no correlation with separate E_{solv}/μ^{ex} terms or the vacuum energies, only the total free energy in solution yields a reliable ranking.

Case (*11*): For compounds **33** and **34** the **A** forms are homoaromatic whereas the **B** forms are to some degree antiaromatic, more pronounced so for **33**. **A** is strongly favored, dominated by E_{solv} and supported by the gas phase result. In contrast, **32B** is heavily preferred due to electronic conjugation which is not present in **32A**, as deduced by E_{solv} and E_{vac}. The solvation effect has the same direction here (dipole moment for **32B**: 8.0 D, for **32A**: 4.2 D).

Case (*12*): Throughout we find for the reactions **35C → 35A** that **35C** is always favored whereas, for the reactions **35A → 35B**, **35A** is always preferred. This means that the enolized cyclic ester/thioester/amide forms **35B** are unfavorable as found similarly in case (9) above. **35C** is indeed identified to be the dominant species, and, again as suggested [4], the gas phase calculations show a similar trend which explains the low degree of solvent sensitivity for this equilibrium.

Results and discussion for the post-disclosure phase

Obscure compounds, cases (1)–(3)

In light of the promising results from the application of the optimal model to the explanatory compounds, the disclosure of free energy data for the obscure compounds revealed a significantly worse predictive performance, as shown in Table 6 and Fig. 2. Evaluating compounds **1**–**6** separately, the RMSE of 2.9 kcal mol^{-1} is beyond the target of "chemical accuracy". Remarkably, however, the linear correlation between model and experiment is quite

Table 6 Results of MP2/aug-cc-pVDZ/EC-RISM/PSE-3 calculations for standard reaction free energies on globally optimal solution structures (B3LYP/6-311++G(d,p)/PCM) for different LJ models; RMSE, b in kcal mol^{-1}

Compounds	LJ model	RMSE	R^2	m	b
1–6	GAFF–P_0	2.91	0.89	1.10	−2.20
10–16[a]	GAFF-P_0	0.57	0.77	0.89	−0.05
1–16	GAFF-P_0	1.93	0.86	1.18	−0.63
1–16	GAFF-P_0 (**7–16**)/ GAFF-P_2 (**1–6**)	1.04	0.91	0.86	0.00

[a] See also Table 4

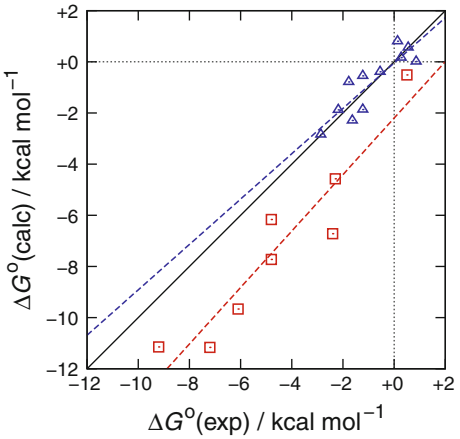

Fig. 2 Calculated (global minima) vs. experimental standard reaction free energies for obscure cases (1)–(3), compounds **1–6** (*red squares*) and explanatory case (5), compounds **10–16** (*blue triangles*); the *solid black line* indicates perfect correlation. The structures of all compounds were optimized by B3LYP/6-311++G(d,p)/PCM and post-processed by MP2/aug-cc-pVDZ/EC-RISM/PSE-3 (GAFF-P_0). Regression lines are shown in the color of the data points

good while the RMSE is almost completely contained in the intercept of the regression line. This means that the predictions systematically underestimate the experimental values, leading to almost parallel regression curves for compounds **1–6** vs. **10–16**.

Full set of compounds, cases (1)–(5)

A possible explanation for the apparent discrepancy between explanatory and obscure compounds becomes evident if a regression analysis is performed for the full set comprising both groups. Corresponding data is reported in Table 6 and in Fig. 3. Using the original LJ parameter set GAFF-P_0 yields an RMSE of 1.9 kcal mol^{-1} and a slope and intercept which reflect the influence of the problematic obscure group. However, if the LJ parameter σ(carbonyl-O) is modified selectively for the obscure compounds **1–6** by adding 0.25 Å (parameter set GAFF-P_2), the overall performance of the model is substantially improved. In this

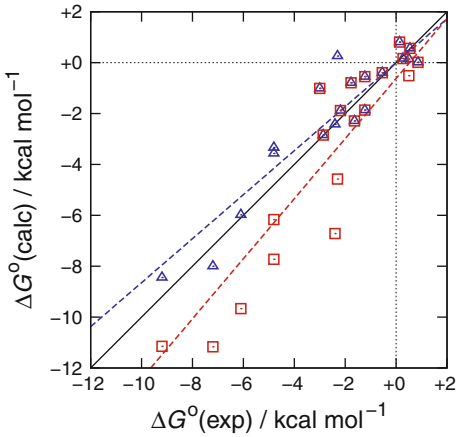

Fig. 3 Calculated (global minima) vs. experimental standard reaction free energies for obscure and explanatory cases (1)–(5), compounds **1–16**; the *solid black line* indicates perfect correlation. The structures of all compounds were optimized by B3LYP/6-311++G(d,p)/PCM and post-processed by MP2/aug-cc-pVDZ/EC-RISM (GAFF-P_0) (*red squares*) and by MP2/aug-cc-pVDZ/EC-RISM (GAFF-P_0 for **7–16**, GAFF-P_2 for **1–6**) (*blue triangles*). Regression lines are shown in the color of the data points

way, a total RMSE of 1.0 kcal mol^{-1} is obtained while the regression analysis shows a clear linear correlation with negligible intercept.

The approach to increase the carbonyl-O diameter is motivated by our recent development towards predictive pK_a models (manuscript in preparation). It turned out that the LJ-σ phenolate oxygen sites needs to be larger than the standard sp^2-O type (carbonyl and carboxylate) in GAFF which was assigned by default. This correlates with the structural difference between cases (1)–(3) which comprises essentially 6-membered rings and case (5) that consists of five-membered rings within which tautomeric proton shifts occur.

On the other hand, another striking difference between the groups is the magnitude of the reaction free energies. Those of the explanatory compounds are clustered around 0 whereas the range for the others is much larger, down to ca. −9 kcal mol^{-1}. It is expected that the experimental uncertainty is larger in this region if the tautomer ratio is determined by direct observation due to the negligible concentration of one the species. A critical re-examination of the experimental data would therefore be important and valuable before computational models are further developed and optimized.

Comparison with results of other groups

Not counting reaction **10B** → **10D** that has been reconstructed from the available thermodynamic data, our overall performance for compounds **1–16** is 2.0 kcal mol^{-1}

in terms of the RMSE metric without parameter adjustments. This is the smallest value in the field of all competitors who reported data for the full set of compounds. The second and third-ranking groups obtain 2.5 and 2.9 kcal mol^{-1}, respectively. In the case of compounds **1–8** for which no numerical data was known in advance our RMSE is 2.7 kcal mol^{-1}; in this group the best and worst-ranking competitors measured by the RMSE alone reach 1.1 and 4.0 kcal mol^{-1}, respectively. In light of the small sample size and the limitations of the RMSE metric for method comparison, it is clear that an objective and unbiased estimate of the overall performance requires much broader statistical analyses in future work.

Remarkably, the best-ranking competitors' models for **1–8** perform poorly for **10–16**. Again in terms of the RMSE, here the best-ranking model is ours with 0.6 kcal mol^{-1}, the second-ranking yields 2.7 kcal mol^{-1}, and the model which ranks first for **1–8** yields 4.2 kcal mol^{-1} for **10–16**. This is a clear indication that either the experimental uncertainty differs strongly between **1–8** and **10–16**, or that competing groups have similar problems in balancing prediction errors for five-membered and six-membered rings. Apparently only one of the types can be treated adequately without further adjustments. This issue warrants deeper investigation.

Conclusion

The challenge to predict tautomer ratios in aqueous solution resulted in the definition of an optimized workflow for routine application of EC-RISM calculations. We found that it is a useful strategy to optimize the geometry at B3LYP/6-311++G(d,p)/PCM followed by post-processing at MP2/aug-cc-pVDZ/EC-RISM/PSE-3. It is absolutely essential to locate the global minimum for OH rotamers whereas, for the cases studied here, conformational averaging turned out to be unnecessary. Application of Lennard-Jones parameters provided by GAFF yields an RMSE of 2.0 kcal mol^{-1}, which represents the smallest value among the models tested during the SAMPL2 challenge.

The analysis revealed that all models evaluated in this contest suffer from an imbalance of the predictive power depending on compound classes with distinct structural elements (here five-membered vs. six-membered rings). Although a closer examination of the experimental data is recommended in order to rule out model optimization with respect to wrong target data, it has been shown that existing deficiencies can be coped with by fine-tuning of carefully chosen Lennard-Jones parameters. In this way, an overall RMSE of 1 kcal mol^{-1} corresponding with "chemical accuracy" can be achieved which approaches our value of 0.6 kcal mol^{-1} already obtained for the five-membered systems without parameter adjustments. However, the assembly of an appropriately balanced, larger and chemically more diverse tautomer training set with well-defined experimental error margins is a prerequisite for a sensible parameter optimization and sensitivity analysis.

Acknowledgments We thank the Deutsche Forschungsgemeinschaft, the Fonds der Chemischen Industrie, and the Adolf-Messer-Stiftung for financial support.

References

1. Kloss T, Heil J, Kast SM (2008) J Phys Chem B 112:4337–4343
2. Beglov D, Roux B (1997) J Phys Chem 101:7821–7826
3. Kovalenko A, Hirata F (1998) Chem Phys Lett 290:237–244
4. Taylor PJ (2009) Tautomeric teasers: SAMPL (to be replaced by overview paper)
5. Kast SM, Kloss T (2008) J Chem Phys 129:236101
6. Perkyns J, Pettitt BM (1992) Chem Phys Lett 190:626–630
7. Perkyns J, Pettitt BM (1992) J Chem Phys 97:7656–7666
8. Frisch MJ et al (2004) Gaussian 03, Rev D.02/E.01. Gaussian, Inc., Wallingford
9. Berendsen HJC, Grigera JR, Straatsma TP (1987) J Phys Chem 91:6269
10. Maw S, Sato H, Ten-no S, Hirata F (1997) Chem Phys Lett 276:20–25
11. Sato H, Hirata F (1999) J Chem Phys 111:8545–8555
12. Kovalenko A, Ten-no S, Hirata F (1999) J Comput Chem 20:928–936
13. Talman JD (1978) J Comput Phys 29:35–48
14. Rossky PJ, Friedman HL (1980) J Chem Phys 72:5694–5700
15. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) J Comput Chem 25:1157–1174
16. Chirlian LE, Francl MM (1987) J Comput Chem 8:894–905