

# Avoiding gas-phase calculations in theoretical $pK_a$ predictions

Rodrigo Casasnovas · David Fernández ·  
Joaquín Ortega-Castro · Juan Frau ·  
Josefa Donoso · Francisco Muñoz

Received: 13 December 2010 / Accepted: 16 April 2011 / Published online: 6 May 2011  
© Springer-Verlag 2011

**Abstract** CBS-QB3, two simplified and less computationally demanding versions of CBS-QB3, DFT-B3LYP, and HF quantum chemistry methods have been used in conjunction with the CPCM continuum solvent model to calculate the free energies of proton exchange reactions in water solution following an isodesmic reaction approach. According to our results, the precision of the predicted  $pK_a$  values when compared to experiment is equivalent to that of the thermodynamic cycles that combine gas-phase and solution-phase calculations. However, in the aqueous isodesmic reaction schema, the accuracy of the results is less sensitive to the presence of explicit water molecules and to the global charges of the involved species since the free energies of solvation are not required. In addition, this procedure makes easier the prediction of  $pK_a$  values for molecules that undergo large conformational changes in solvation process and makes possible the  $pK_a$  prediction of unstable species in gas-phase such as some zwitterionic tautomers. The successive  $pK_a$  values of few amino acids corresponding to the ionization of the  $\alpha$ -carboxylic acid and  $\alpha$ -amine groups, which is one of the problematic cases for thermodynamic cycles, were successfully calculated by employing the aqueous isodesmic reaction yielding mean absolute deviations of 0.22 and 0.19  $pK_a$  units for the first and second ionization processes, respectively.

**Keywords**  $pK_a$  Calculations · Isodesmic reaction · Thermodynamic cycles · CBS-QB3 · CPCM · Continuum solvent models

## 1 Introduction

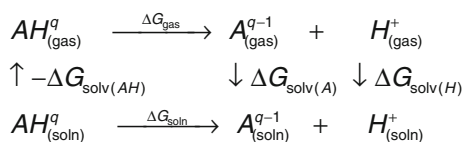
The continuous development of implicit solvent models (also known as continuum solvent models) during the last decades together with their simplicity and low computational costs in front of the more complex explicit solvent simulations, made possible the theoretical prediction of physicochemical properties of a broad variety of chemical species in different solution media [1–6]. One of the remaining subjects in continuum solvent calculations is the accurate calculation of acidity constants ( $K_a$ ) or  $pK_a$  in its negative logarithmic form [7].

Prior to introduce the ways in which  $pK_a$  values are obtained from quantum calculations and how solvent effects are included, we should consider that the state-of-the-art quantum chemistry calculations provide gas-phase energies with precision within 1 kcal mol<sup>-1</sup>. However, the continuum model approach would involve some sort of additional error. Since  $pK_a$  values are proportional to deprotonation free energies, accurate  $pK_a$  prediction is therefore a challenge considering that experimental errors are within 0.01–0.1  $pK_a$  units for many species.

The most common approach in  $pK_a$  calculations is the use of thermodynamic cycles that combine a gas-phase deprotonation reaction and desolvation/solvation of the involved reactants/products [7–18]. The simplest thermodynamic cycle for determination of absolute  $pK_a$  values (also known as “direct” method) is depicted in Scheme 1. In this cycle, the acid species ( $AH^g$ ) is desolvated, deprotonated in gas-phase and, in the last step, the conjugate

**Electronic supplementary material** The online version of this article (doi:10.1007/s00214-011-0945-5) contains supplementary material, which is available to authorized users.

R. Casasnovas · D. Fernández · J. Ortega-Castro ·  
J. Frau (✉) · J. Donoso · F. Muñoz  
Department of Chemistry, Universitat Illes Balears,  
07122 Palma de Mallorca, Spain  
e-mail: juan.frau@uib.es



**Scheme 1** Thermodynamic cycle 1 that considers dissociation of acid species in the conjugate base and proton. Total charge of the acid and the conjugate base are represented by  $q$  and  $q - 1$ , respectively

base ( $A^{q-1}$ ) and the isolated proton ( $H^+$ ) are solvated to complete the cycle.

According to this, the solution deprotonation free energy can be calculated as shown in (1). The factor  $RT \ln 24.46$  accounts for the change in gas-phase reference state from 1 atm to 1 M. Eventually,  $pK_a$  values can be obtained from (2):

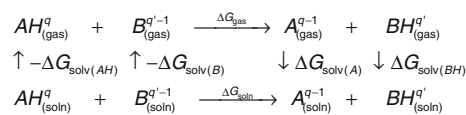
$$\begin{aligned}
 \Delta G_{\text{soln}} = & \{G_{\text{gas}}(A^{q-1}) + G_{\text{gas}}(H^+) - G_{\text{gas}}(AH^q) \\
 & + \Delta G_{\text{solv}}(A^{q-1}) + \Delta G_{\text{solv}}(H^+) - \Delta G_{\text{solv}}(AH^q) \\
 & + RT \ln 24.46\} \quad (1)
 \end{aligned}$$

$$pK_a = \frac{\Delta G_{\text{soln}}}{RT \ln 10} \quad (2)$$

This cycle yielded successful results for aqueous  $pK_a$  values of carboxylic acids [8–10], phenols [11], and pyridines [18] with mean absolute errors below 0.5  $pK_a$  units. On the other hand, much worse results were obtained for carbon acids [7, 13, 16]. According to this cycle, the experimental solvation energy for the isolated proton [ $\Delta G_{\text{solv}}(H^+)$ ] is required. This value is not easy to measure experimentally and may involve a significant error. Consequently, several values have been used in  $pK_a$  calculations [7].

When aqueous  $pK_a$  values are calculated, several alternatives to Scheme 1 are possible in which the proton is substituted in the right-hand side of the reaction for an hydronium cation (and a water molecule is added in the left-hand side to balance the cycle) [7, 8, 10, 15, 17–19]. Or, alternatively, the proton is replaced by a water molecule in the right-hand side and a hydroxide anion is included in the left-hand side [7, 15, 17, 19]. In these two alternative cycles, gas-phase and solvation free energies for water, hydronium and hydroxide can be either taken from experiment or calculated by quantum methods [20–24].

As shown in (1) and (2), the accuracy in  $pK_a$  calculations depends on the errors present in both gas-phase and solvation free energies. Several papers suggested to use high *ab initio*-level calculations for computing deprotonation free energies in gas-phase with errors near the so-called chemical accuracy ( $1 \text{ kcal mol}^{-1}$ ) in order to obtain accurate acidities [7–11, 19]. On the other hand, several benchmarking studies of continuum models showed



**Scheme 2** Thermodynamic cycle 2 that considers a proton exchange reaction between an acidic species and a reference acid molecule. Global charge of the acids and the conjugate bases are represented by  $q/q'$  and  $q - 1/q' - 1$ , respectively

that while solvation free energies for neutral species can be predicted with errors of  $1 \text{ kcal mol}^{-1}$ , average errors in the free energies of solvation corresponding to charged species usually reach  $5 \text{ kcal mol}^{-1}$  for the best continuum models [25, 26]. For that reason, the largest contribution to the total error frequently lies in the solvation free energies when  $pK_a$  values are calculated by using a thermodynamic cycle approach.

A different cycle that is built upon a proton exchange reaction (relative  $pK_a$  determination) has been successfully applied for some species [14, 16, 18] (Scheme 2). This approach is based on an isodesmic reaction in which the acid ( $AH^q$ ) donates a proton to the conjugate base ( $B^{q'-1}$ ) of a reference acid molecule ( $BH^{q'}$ ) in which  $pK_a$  is experimentally known.

According to this cycle,  $pK_a$  values can be calculated from (3) and (4). Note that in this case, the factor  $RT \ln 24.46$  is not present since the number of moles is the same in both sides of the reaction.

$$\begin{aligned}
 \Delta G_{\text{soln}} = & \{G_{\text{gas}}(A^{q-1}) + G_{\text{gas}}(BH^{q'}) - G_{\text{gas}}(AH^q) \\
 & - G_{\text{gas}}(B^{q'-1}) + \Delta G_{\text{solv}}(A^{q-1}) + \Delta G_{\text{solv}}(BH^{q'}) \\
 & - \Delta G_{\text{solv}}(AH^q) - \Delta G_{\text{solv}}(B^{q'-1})\} \quad (3)
 \end{aligned}$$

$$pK_a(AH^q) = \frac{\Delta G_{\text{soln}}}{RT \ln 10} + pK_a(BH^{q'}) \quad (4)$$

As can be inferred from (3), the advantage of this thermodynamic cycle lies in the expected cancellation of errors in the calculated  $\Delta G_{\text{soln}}$  due to the subtraction of the gas-phase and solvation free energies of the reactants and products. Such cancellation of errors is likely to be maximized between the gas-phase free energies for a given quantum chemistry method if the molecular and electronic structure of  $(AH^q)/(BH^{q'})$  and  $(A^{q-1})/(B^{q'-1})$  are similar. On the other hand, since the largest errors in the free energies of solvation lie in those of ionic species and principally depend on their global charge, the total error coming from the subtractions between  $\Delta G_{\text{solv}}(AH^q)/\Delta G_{\text{solv}}(BH^{q'})$  and  $\Delta G_{\text{solv}}(A^{q-1})/\Delta G_{\text{solv}}(B^{q'-1})$  is expected to be minimum when  $q$  and  $q'$  are equal.

One may think that more accurate  $pK_a$  predictions could be obtained by considering some procedure in which the free energies in solution are calculated directly instead of using gas-phase and solvation free energies. In fact, few

authors predicted some  $pK_a$  values from free energies in solution calculated as

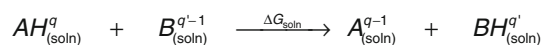
$$G_{soln} = (E_{soln} + G_{nes}) + \Delta G_{corr\_soln} \quad (5)$$

where  $E_{soln}$  is the potential energy of the solute structure optimized in the continuum model,  $G_{nes}$  includes the cavitation, dispersion, and repulsion solvent effects, and  $\Delta G_{corr\_soln}$  is a correction accounting for the nuclear motions at room temperature calculated within the harmonic rigid rotor approximation also in the presence of the dielectric continuum [15, 27, 28]. However, Coote and coworkers [29] pointed out that this procedure for obtaining free energies in solution is not correct since: (1) potential energies in solution include (in an unknown extent) a residual contribution of the differences between gas-phase and solution-phase free energies through parametrization of the continuum solvent model and/or the set of radii used to define the solute cavity in the dielectric continuum. And (2) the rotor rigid harmonic approximation is dubiously valid to describe the behavior of the nuclear motions in solution phase, specially the translations and rotations. Instead, free energies in solution should be calculated by (6) in order to be consistent with the design of continuum models.

$$G_{soln} = (E_{soln} + G_{nes}) + \Delta G_{corr\_gas} \quad (6)$$

Since during the parametrization of either the continuum solvation models or the set of radii Ben-Naim's definition for solvation free energy is usually assumed [25, 30], to be consistent with (6) one should use gas-phase optimized geometries. Otherwise, there is no guarantee that solution-phase optimized geometries correspond to minima in the gas-phase potential energy surface and gas-phase thermal corrections ( $\Delta G_{corr\_gas}$ ) could be meaningless in that case. Then,  $pK_a$  calculations seem to be limited by (6) to those cases where gas-phase geometries are representative of solution-phase structures.

The present study focuses on using an aqueous isodesmic reaction (Scheme 3) as an approach to avoid gas-phase calculations in  $pK_a$  predictions. In fact, this procedure can be interpreted as an alternative formulation of the thermodynamic cycle 2 in which gas-phase calculations are not present. Then, according to Scheme 3,  $pK_a$  values are computed by (4) as for thermodynamic cycle 2 but by using



**Scheme 3** Isodesmic reaction that considers a proton exchange reaction between an acidic species and a reference acid molecule. Global charge of the acids and the conjugate bases are represented by  $q/q'$  and  $q - 1/q' - 1$ , respectively

(7) instead of (3) to obtain the free energies of deprotonation in solution ( $\Delta G_{soln}$ )

$$\Delta G_{soln} = G_{soln}(A^{q-1}) + G_{soln}(BH^{q'}) - G_{soln}(AH^q) - G_{soln}(B^{q'-1}) \quad (7)$$

A previous study on acidity of pyridines revealed that very low errors (ca. 0.3–0.7 log units) for  $pK_a$  values can be obtained with different thermodynamic cycles [18]. On the contrary,  $pK_a$  values of carbon acids have been difficult to predict accurately with thermodynamic cycles and the lowest obtained errors range 1–4  $pK_a$  units depending on the global charges of the acid and its conjugate base [16].

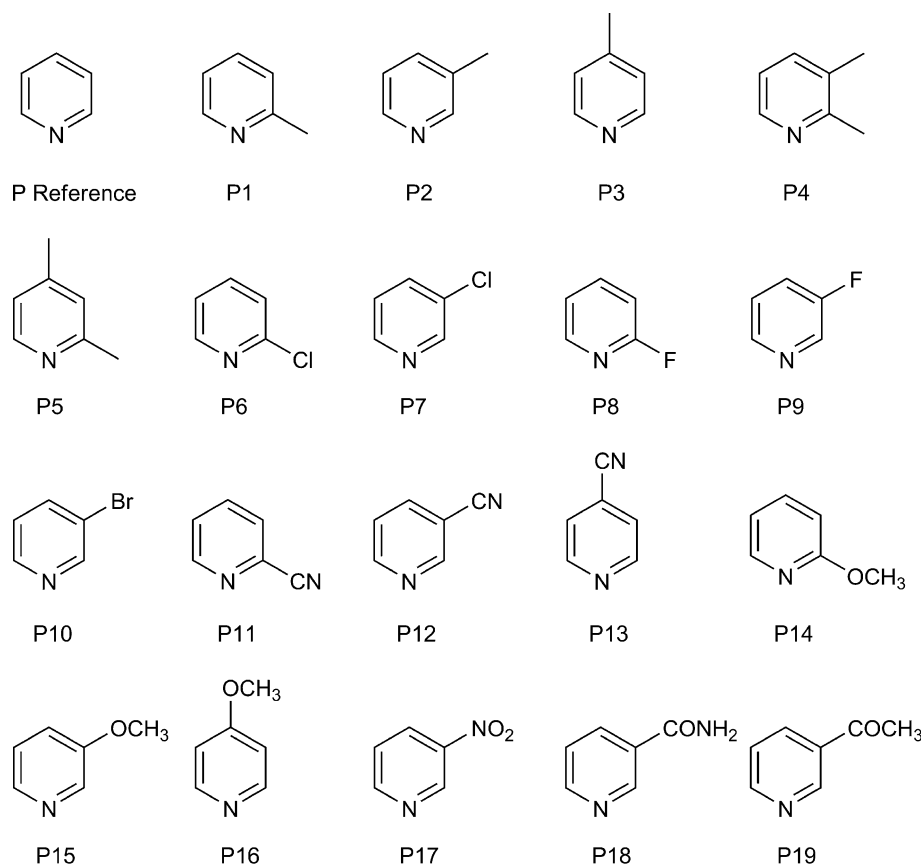
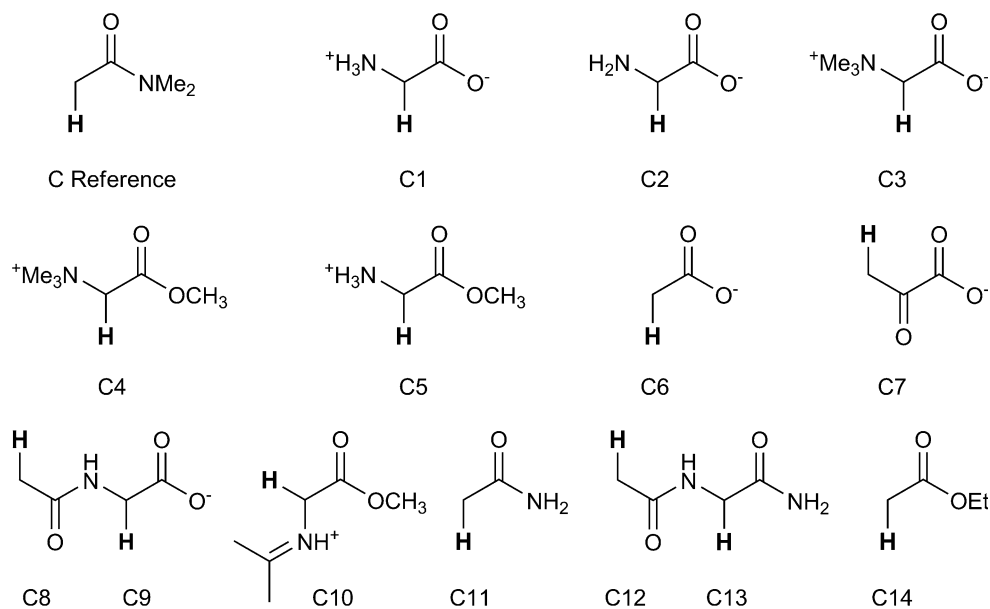
We will use the aqueous isodesmic reaction approach depicted in Scheme 3 to calculate the acidity of a set of substituted pyridines (Chart 1) and a set of carbon acids (Chart 2) so that we can compare the performance of this procedure in relation to thermodynamic cycles. In addition, the two conceptually different ways to obtain free energies in solution represented by (5) and (6) are considered for the  $pK_a$  calculations.

## 2 Computational details

All the calculations were performed with the Gaussian03E.01 package [31]. Different sets of  $pK_a$  values were calculated by using HF/6–31+G(d), DFT-B3LYP/6–311++G(d,p), CBS-QB3 [32–34], CBS-4B3 [37], and CBS-4B3\* quantum chemistry methods. The small-sized 6–31+G(d) basis set was used for HF calculations. However, the larger 6–311++G(d,p) basis set was required for DFT-B3LYP calculations in order to avoid low imaginary frequencies for several species.

Solvent effects were taken into account by using the CPCM continuum model [35, 36] and the UAHF radii [30] for HF and post-HF methods or the UAKS radii [31] for B3LYP calculations. The CPCM continuum model has been used since previous works reported considerably accurate  $pK_a$  predictions for several acid functionalities including substituted pyridines and carbon acids [7–11, 14, 16, 18]. Furthermore, using the same continuum model seemed necessary for rigorous comparisons between the  $pK_a$  predictions with the isodesmic reaction reported here and previous calculations with thermodynamic cycles [16, 18].

The original CBS-QB3 model involves geometry optimization and frequency calculations at B3LYP/6–311G(2d,d,p) level. Afterward, the energy is calculated on the optimized geometry in three steps. Second-order energies are computed at MP2/6–311+G(3d2f,2df,2p) and later corrected to the complete basis set limit (CBS extrapolation). The second and third steps correct second-order

**Chart 1** Set of substituted pyridines**Chart 2** Set of carbon acids

energies by MP4SDQ/6–31+G(d,f,p) and CCSD(T)/6–31+G<sup>†</sup> calculations. Zero-point energies and further thermal corrections are taken from B3LYP frequency calculations. At last, empirical and spin contamination corrections are added to the energy. CBS-4B3 method is a

simplification of the original CBS-QB3 in which CCSD(T), spin contamination, and empirical corrections to the energy are not included [37]. The resulting method is considerably less computational demanding due to the absence of coupled cluster calculations. This method has been named CBS-4B3

since it retains the original CBS extrapolation of all CBS methods, “4” because the highest level of calculation is MP4SDQ and “B3” due to the B3LYP geometries and frequencies. Eventually, CBS-QB3 and CBS-4B3 energies are computed by using (8) and (9).

$$G_{CBSQB3} = E_{SCF} + \Delta E_{MP2} + \Delta E_{CBS} + \Delta E_{MP4SDQ} + \Delta E_{CCSD(T)} + \Delta E_{emp} + \Delta E_{spin} + \Delta G_{corr} \quad (8)$$

$$G_{CBS4B3} = E_{SCF} + \Delta E_{MP2} + \Delta E_{CBS} + \Delta E_{MP4SDQ} + \Delta G_{corr} \quad (9)$$

The summation of the first three terms ( $E_{SCF} + \Delta E_{MP2} + \Delta E_{CBS}$ ) corresponds to the second-order energies, the  $\Delta G_{corr}$  term accounts for the thermal corrections to the potential energy calculated within the harmonic approximation at B3LYP level scaled by 0.99 with the unscaled zero point energy correction removed according to the implementation in G03 [31]. The later terms are obtained as

$$\Delta E_{MP4SDQ} = E_{MP4SDQ/6-31+G(d(f),p)} - E_{MP2/6-31+G(d(f),p)} \quad (10)$$

$$\Delta E_{CCSD(T)} = E_{CCSD(T)/6-31+G^\dagger} - E_{MP4SDQ/6-31+G^\dagger} \quad (11)$$

$$\Delta E_{emp} = -0.0057 \sum_{i=1}^{n_f} \left( \sum_{\mu=1}^{N_{virt+1}} C_{\mu ii} \right)^2 |S_{ii}^2| \quad (12)$$

$$\Delta E_{spin} = 0.00954 [\langle S^2 \rangle - S_z(S_z - 1)]. \quad (13)$$

Since CBS-QB3 and CBS-4B3 models use 6–311G(2d,d,p) basis set for geometry optimizations and the absence of diffuse functions may not describe properly hydrogen bonding interactions, all the geometries were reoptimized with B3LYP/6–311++G(d,p). Afterward, CBS-4B3 energies according to (9) were calculated on the resulting structures (named here as CBS-4B3\*). The scale factor for the thermal corrections was modified to 0.9806 according to the combination of 6–311++G(d,p) basis set and B3LYP functional [38].

Free energies in solution for each species as defined in (5) were obtained by performing both geometry optimizations and frequency calculations (thermal corrections) together with the CPCM model for all HF, B3LYP, and CBS methods. Moreover, all CBS energy calculations were performed in the presence of the CPCM continuum model.

On the other hand, free energies in solution as defined in (6) require the geometry optimizations and frequency calculations to be performed in the gas-phase. Subsequently, an energy calculation on the fixed gas-phase optimized geometry is performed in the presence of the CPCM dielectric continuum. In this case, only HF and B3LYP but not CBS methods were used to calculate the free energies in solution according to (6).

All the structures were optimized and characterized as true minima by the absence of imaginary frequencies except for the C9 acid species optimized at CPCM/HF/6–31+G(d) level, which presented a low ( $17.5 \text{ cm}^{-1}$ ) imaginary frequency corresponding to the torsion of a methyl group. The free energies in solution according to (5) were also calculated for all pyridines and carbon acids adding a single water molecule to study the effects of explicit short-range solvent interactions in the accuracy of  $pK_a$  predictions.

### 3 Results and discussion

Before showing the results, some comments should be done concerning the expected precision in the results provided by the isodesmic approach. According to Ho and Coote [7], computational absolute and relative  $pK_a$  predictions should present errors within 3.5/2  $pK_a$  units, respectively, when thermodynamic cycles are used.

According to (4) (see Introduction section), the errors in the calculated  $pK_a$  values lie in the calculated free energy of the proton exchange reaction in solution and in the experimental uncertainty of the acid reference  $pK_a$  value. The former free energy is obtained by introducing in (7) the free energies in solution calculated with the quantum chemistry method. In this work, we used different CBS models in which errors in gas-phase deprotonation energies were ranged within  $1.2 \text{ kcal mol}^{-1}$  [39, 40]. Bearing this in mind and considering addition of errors, the expected accuracy should be  $2.4 \text{ kcal mol}^{-1}$  (approximately 1.8  $pK_a$  units at 298.15 K). However, the lower ab initio level HF and DFT-B3LYP calculations may be expected to yield larger errors due to a worse description of electron correlation and exchange interactions.

In addition to the inherent errors of the quantum chemistry methods, the errors due to the introduction of solvent effects as a dielectric continuum should also be considered. However, it is somewhat difficult to define the magnitude of these errors within the aqueous isodesmic reaction approach because continuum solvent models (or the set of solvation radii) are constructed to yield free energies of solvation that are neither used in (5) nor in (6). For example, UAHF and UAKS spheres were parametrized to yield accurate potential energy differences between gas and solution phases computed with small basis sets at HF or PBE0 levels, respectively [30, 31]. As long as HF or DFT gas-phase energies are far from “exact”, it is not possible to decompose the total error of solvation energies in their gas-phase and solution-phase contributions.

In a first approximation, we can set two error limits for the computed aqueous free energies. First, if errors in solvation energies are only located in the solution energies

(due to the CPCM model), aqueous-phase errors will be added to the inherent errors of the quantum chemistry method. Considering that the average errors in solvation energies for the CPCM/UAHF and CPCM/UAKS solvent models reported by Takano and Houk [26] were approximated located within 1–5 kcal mol<sup>-1</sup> depending on the global charge of the solute, the error in each deprotonation energy can easily increase in the same extent depending on the charges of the acid and its conjugate base. Moreover, this error will be easily enlarged in case of poly-charged ions. On the other hand, if errors in solvation energies are only due to gas-phase energies (due to the quantum chemistry method), the global errors for each method will remain at its intrinsic level.

Despite it is not possible to predict whether the errors in the free energies in solution [either calculated by (5) or (6)] are larger than the errors in the free energies of solvation, it is reasonable to expect that errors in the solution free energies present a systematic dependence on the structural features of the solutes such as global charges, molecular, and electronic structures (in particular the electrostatic potential) due to the parametrization of the continuum model/cavities in the same extent than the free energies of solvation. Therefore, some cancellation of errors is expected for the results of the aqueous isodesmic reaction (Scheme 3) as occurs with the proton exchange thermodynamic cycle (Scheme 2) [7, 9, 12, 15, 16, 18].

At last, the experimental uncertainties in the pK<sub>a</sub> of the acid reference should be taken into account. These errors usually fall in the range of 0.1–0.01 pK<sub>a</sub> units for most of the typical organic and inorganic compounds; therefore, these are the less important contributions to the global error in the calculated pK<sub>a</sub> values. Nevertheless, for some very strong or very weak acid species (i.e., those species which pK<sub>a</sub> value falls out of the range 0–14), the experimental values could show errors of several pK<sub>a</sub> units [7] and become as significant as the errors in the calculated free energies of deprotonation.

We have shown earlier that the aqueous isodesmic reaction should not necessarily involve additional sources of error respect to thermodynamic cycles. Instead, both approaches seem to be equivalent. Therefore, an obvious question comes out: Which are the advantages of using the aqueous isodesmic reaction instead of thermodynamic cycles? (1) Some species adopt very different gas- and solution-phase geometries that conformational free energy change should be computed [7, 27, 28]. (2) Some acids can be gas-phase unstable, particularly in the case of polyprotic species where intramolecular proton transfers are possible [27, 28, 41]. (3) When “rare” acid–base functionalities are not included in the parametrization set of radii, there may be large errors when computing solvation free energies. (4) Previous studies illustrated that thermodynamic cycles

usually fail when calculating pK<sub>a</sub> values of polyprotic species, probably due to large errors in the solvation energies of poly-charged molecules (anions/cations). However, the application of the aqueous isodesmic reaction improved drastically the results in some cases [27, 28]. (5) Accurate results resulted strongly dependent on the global charge of the acid reference when using thermodynamic cycles [16]. Conversely, accurate results were obtained for a polyprotic species, in a different study, where the studied and reference acids/bases bear different global charges [28]. (6) For thermodynamic cycles, addition of explicit solvent molecules was critical to obtain accurate pK<sub>a</sub> values when strong, short-range solute–solvent interactions are not well described by the continuum model [18, 19, 42].

Results for pK<sub>a</sub> prediction of substituted pyridines and carbon acids calculated by using (5) and (6) are shown in Tables 1 and 2, respectively. It should be noted that only vibrational partition functions were taken into account for the thermal corrections calculated in solution ( $\Delta G_{\text{corr\_soln}}$ ) since ideal gas partition functions for rotational and translational motions do not describe the real physics in solution as noted by Coote and coworkers [29]. On the other hand, it may seem reasonable to consider that translations and rotations of the solutes do not change dramatically after the deprotonation reaction. Therefore, we assume that all the non-vibrational thermal contributions cancel out in (7). Moreover, we also assume that any contribution due to the structural reorganization of the solvent will be similar between the studied and reference species so they will also cancel out in (7).

Pyridine acid–base behavior is well known from both experiment [43–46] and computational studies [18, 46–48]. It should be noted that pyridine functionality has not only been chosen because of the high accuracy in the computational determinations of pK<sub>a</sub> (errors below 1 pK<sub>a</sub> unit) but also because good results have been reported with different thermodynamic cycles [18, 48]. Considering that the estimated errors vary in several pK<sub>a</sub> units for different thermodynamic cycles [7], equivalent results for different cycles can be only obtained if all gas-phase free energies of deprotonation and free energies of solvation are accurate. This means that low-error cancellation is present in the pK<sub>a</sub> calculation, so substituted pyridines will be a proper set of molecules to assess the performance of the aqueous isodesmic reaction approach.

On the other hand, the experimental determination of acidity constants for carbon acids is rather difficult because of their very weak acidity in water. In fact, experimental uncertainties of such pK<sub>a</sub> values are significantly larger (i.e. 0.5–1 pK<sub>a</sub> unit or larger) than those obtained for common acid species [49–51]. The reference carbon acid was chosen in the post-processing stage as the species that yields

**Table 1** Calculated  $pK_a$  values for substituted pyridines using the aqueous isodesmic reaction according to (5) or (6)

	Exptl.	HF (5)	HF (6)	B3LYP (5)	B3LYP (6)	CBS-4B3* (5)	CBS-4B3 (5)	CBS-QB3 (5)
PRef	5.23 [43]							
P1	5.97 [44]	6.81 (0.84)	6.84 (0.87)	7.01 (1.04)	6.73 (0.76)	7.03 (1.06)	6.64 (0.67)	6.60 (0.63)
P2	5.68 [44]	6.04 (0.36)	5.98 (0.30)	6.11 (0.43)	6.06 (0.38)	6.09 (0.41)	6.10 (0.42)	6.12 (0.44)
P3	6.02 [44]	6.36 (0.34)	6.26 (0.24)	6.38 (0.36)	6.47 (0.45)	6.53 (0.51)	6.43 (0.41)	6.35 (0.33)
P4	6.57 [43]	7.51 (0.94)	7.65 (1.08)	7.65 (1.08)	7.85 (1.28)	7.62 (1.05)	7.63 (1.06)	7.60 (1.03)
P5	6.99 [43]	7.88 (0.89)	8.05 (1.06)	8.53 (1.54)	8.45 (1.46)	8.39 (1.40)	7.92 (0.93)	7.79 (0.80)
P6	0.49 [43]	-1.15 (1.64)	-1.25 (1.74)	-0.65 (1.14)	-0.73 (1.22)	-0.26 (0.75)	-0.27 (0.76)	-0.25 (0.74)
P7	2.81 [43]	2.54 (0.27)	2.42 (0.39)	2.87 (0.06)	2.77 (0.04)	2.83 (0.02)	2.84 (0.03)	2.89 (0.08)
P8	-0.44 [43]	-1.80 (1.36)	-1.89 (1.45)	-1.60 (1.16)	-1.74 (1.30)	-1.52 (1.08)	-1.55 (1.11)	-1.58 (1.14)
P9	2.97 [44]	2.65 (0.32)	2.50 (0.47)	3.03 (0.06)	2.87 (0.10)	3.16 (0.19)	3.16 (0.19)	3.17 (0.20)
P10	2.84 [43]	2.60 (0.24)	2.50 (0.34)	2.87 (0.03)	2.78 (0.06)	2.77 (0.07)	2.79 (0.05)	2.85 (0.01)
P11	-0.26 [45]	-4.06 (3.80)	-4.12 (3.86)	-2.62 (2.36)	-2.72 (2.46)	-2.98 (2.72)	-2.98 (2.72)	-2.94 (2.68)
P12	1.45 [45]	-0.37 (1.82)	-0.41 (1.86)	-0.28 (1.73)	-0.30 (1.75)	-0.08 (1.53)	-0.07 (1.52)	-0.08 (1.53)
P13	1.90 [45]	-0.50 (2.40)	-0.50 (2.40)	0.23 (1.67)	0.22 (1.68)	0.15 (1.75)	0.16 (1.74)	0.17 (1.73)
P14	3.28 [44]	1.66 (1.62)	1.67 (1.61)	1.25 (2.03)	1.41 (1.87)	1.65 (1.63)	1.70 (1.58)	1.59 (1.69)
P15	4.88 [44]	5.27 (0.39)	5.11 (0.23)	6.00 (1.12)	5.88 (1.00)	5.84 (0.96)	5.83 (0.95)	5.93 (1.05)
P16	6.62 [44]	7.96 (1.34)	7.85 (1.23)	7.87 (1.25)	7.81 (1.19)	7.86 (1.24)	7.84 (1.22)	7.68 (1.06)
P17	0.81 [44]	-1.60 (2.41)	-1.74 (2.55)	-1.24 (2.05)	-1.35 (2.16)	-0.92 (1.73)	-0.87 (1.68)	-0.87 (1.68)
P18	3.40 [44]	2.79 (0.61)	2.57 (0.83)	3.04 (0.36)	2.93 (0.47)	3.10 (0.30)	3.08 (0.32)	3.09 (0.31)
P19	3.18 [44]	2.84 (0.34)	2.79 (0.39)	2.80 (0.38)	2.80 (0.38)	3.10 (0.08)	3.10 (0.08)	3.03 (0.15)
MAD		1.15	1.21	1.05	1.05	0.97	0.92	0.91
SD		0.96	0.97	0.73	0.73	0.72	0.71	0.71
ADmax		3.80	3.86	2.36	2.46	2.72	2.72	2.68

The absolute errors in  $pK_a$  units are given in brackets

**Table 2** Calculated  $pK_a$  values for carbon acids using the aqueous isodesmic reaction according to (5) or (6)

	Exptl.	HF (5)	HF (6)	B3LYP (5)	B3LYP (6)	CBS-4B3* (5)	CBS-4B3 (5)	CBS-QB3 (5)
CRef	29.4 ± 0.5 [52]							
C1	28.9 ± 0.5 [51]	24.7 (4.2)	-/	26.1 (2.8)	-/	26.4 (2.5)	26.4 (2.5)	26.7 (2.2)
C2	34 [52]	36.4 (2.4)	38.7 (4.7)	35.3 (1.3)	43.0 (9.0)	35.7 (1.7)	36.5 (2.5)	36.2 (2.2)
C3	27.3 ± 1.2 [51]	28.0 (0.7)	28.8 (1.5)	29.0 (1.7)	29.1 (1.8)	28.9 (1.6)	28.8 (1.5)	29.0 (1.7)
C4	18.0 ± 1.0 [51]	15.5 (2.5)	15.3 (2.7)	14.2 (3.8)	13.6 (4.4)	14.6 (3.4)	14.3 (3.7)	14.4 (3.6)
C5	21.0 ± 1.0 [50]	16.4 (4.6)	17.3 (3.7)	15.3 (5.7)	14.3 (6.7)	16.0 (5.0)	15.9 (5.1)	16.0 (5.0)
C6	33.5 [52]	33.0 (0.5)	34.3 (0.8)	33.3 (0.2)	34.9 (1.4)	33.3 (0.2)	33.8 (0.3)	34.1 (0.6)
C7	17 [53]	15.8 (1.2)	15.9 (1.1)	17.4 (0.4)	15.4 (1.6)	16.8 (0.2)	16.2 (0.8)	16.1 (0.9)
C8	30.3 [54]	30.6 (0.3)	29.1 (1.2)	29.5 (0.8)	28.4 (1.9)	29.4 (0.9)	31.6 (1.3)	30.7 (0.4)
C9	30.8 [54]	38.5 (7.7)	40.4 (9.6)	33.4 (2.6)	35.9 (5.1)	36.8 (6.0)	37.0 (6.2)	36.9 (6.1)
C10	14 ± 1 [55]	14.7 (0.7)	14.4 (0.4)	9.2 (4.8)	7.3 (6.7)	12.0 (2.0)	11.8 (2.2)	11.3 (2.7)
C11	28.4 ± 0.5 [52]	26.1 (2.3)	27.1 (1.3)	25.6 (2.8)	26.4 (2.0)	26.3 (2.1)	26.3 (2.1)	26.2 (2.2)
C12	29.1 [54]	28.0 (1.1)	29.2 (0.1)	29.0 (0.1)	27.7 (1.4)	27.7 (1.4)	28.0 (1.1)	27.6 (1.5)
C13	23.9 [54]	27.4 (3.5)	27.8 (3.9)	23.5 (0.4)	23.3 (0.6)	26.0 (2.1)	26.5 (2.6)	26.6 (2.7)
C14	25.6 ± 0.5 [49]	25.4 (0.2)	25.6 (0.0)	24.6 (1.0)	24.2 (1.4)	25.5 (0.1)	25.2 (0.4)	25.3 (0.3)
MAD		2.3	2.4	2.0	3.4	2.1	2.3	2.3
SD		2.1	2.6	1.8	2.7	1.7	1.7	1.7
ADmax		7.7	9.6	5.7	9.0	6.0	6.2	6.1

The absolute errors in  $pK_a$  units are given in brackets

the minimum errors respect to the experimental values. Moreover, the experimental uncertainty of this species  $pK_a$  is 0.5 log units [51], which is one of the lowest among the studied carbon acids. At this point, it should be taken into account that the expected accuracy of the aqueous isodesmic reaction is modified due to the uncertainty in the  $pK_a$  value of the reference carbon acid. Adding this value to the lowest expected errors (1.8  $pK_a$  units for CBS methods as stated previously), the resulting expected accuracy is 2.3 log units. Note that we have just considered errors of CBS gas-phase energies, but larger errors are expected due to the continuum solvent effects.

As can be seen in Tables 1 and 2, the  $pK_a$ s calculated by considering the free energies in solution defined either by (5) or (6) do not differ significantly for CPCM/HF/6–31+G(d) or for CPCM/B3LYP/6–311++G(d,p) calculations when rigid molecules are considered (i.e., pyridines) since in such cases gas-phase geometries are good models of their solution counterparts. However, slightly worse agreement is found between the  $pK_a$  values obtained by using (5) and (6) when carbon acids are considered due to conformational differences of gas-phase and solution geometries. Actually, the case of zwitterionic glycine species exemplifies a situation in which free energies in solution cannot be obtained according to the more rigorous (6) since that tautomeric state is not stable in the gas-phase. Therefore, in that cases correct  $pK_a$  values can be predicted by performing geometry optimizations and frequency calculations for the thermal corrections in combination with the dielectric continuum according to (5) despite of the fact that the calculated free energies in solution are not obtained in the most consistent way respect to the continuum model design.

The  $pK_a$  values calculated by using free energies in solution according to (5) show mean absolute deviations (MAD) below or close to 1 and 2  $pK_a$  units for the substituted pyridines and carbon acids, respectively, which is in agreement with the previously expected range of errors for the aqueous isodesmic reaction (1.8 and 2.3  $pK_a$  units). In the case of pyridines, the standard deviation values (SD) are also considerably low (ranging 0.71–0.96 log units), which indicates that most of the errors fall also within the expected limit of 1.8  $pK_a$  units. On the contrary, the standard deviations for carbon acids vary from 1.7 to 2.1 log units and accordingly, 4 to 6 species for each quantum method present errors in the calculated  $pK_a$ , which are higher than 2.3 log units. However, it should be considered that: (1) all the pyridine acids and conjugate bases bear the same global charge that the pyridinium/pyridine references, whereas the global charges of carbon acids and their conjugate bases differ from those of the reference carbon acid and (2) carbon acids (particularly the anionic species) were not considered in the parametrization

of the UAHF cavities [30], so larger errors are expected in the SCRF energies.

The magnitude of the standard deviation therefore gives an indication of how the mean absolute deviations will change when the reference species is changed by another molecule within the same set. Accordingly, MAD values change from 0.9 to 2.6  $pK_a$  units (for CBS methods) depending on the pyridine reference but changing the carbon acid reference causes much larger MAD variations, 2.1–6.8  $pK_a$  units in the worst case scenario (see detailed results in Supporting Information).

It is deduced from the analysis of errors of Tables 1 and 2 that mean absolute deviations of the  $pK_a$  values calculated by using (5) and CPCM/HF or CPCM/B3LYP methods show a comparable accuracy respect to the CPCM/CBS results, yielding MAD differences of 0.1–0.2  $pK_a$  units. In fact, CPCM/B3LYP calculations provide slightly better results than CBS models for carbon acids. These results seem to point out that the cancellation of errors in the aqueous isodesmic reaction is efficient up to the SCRF level (i.e., the energy level at which the continuum model affects the solute wavefunction/electron density) and the inclusion of second-order and further corrections to the SCRF energy does not improve significantly the precision of the predicted  $pK_a$  values. This means that the accuracy to which the continuum solvent model and/or the solvation radii are parametrized limits the accuracy of the aqueous isodesmic reaction. Actually, the accuracy of the results provided here is in agreement with the errors reported in previous works where thermodynamic cycles are used for  $pK_a$  calculations on similar sets of molecules [16, 18, 48]. This fact possibly points out that both procedures are equivalent since they are limited by the continuum solvent models in a similar extent.

A closer examination of the results show that for pyridines, the largest errors are found in those species with substituents capable of electron delocalization from or to the aromatic ring (i.e., –F, –CN, –OCH<sub>3</sub>, and –NO<sub>2</sub>). However, errors are reduced to less than 1  $pK_a$  unit for –F and –OCH<sub>3</sub> moieties when placed in position 3 or 4 (P9, P15, and P16). Interestingly, pyridines substituted at position 2 present larger errors than their 3 or 4 counterparts for all the considered substituents placed in different positions (–CH<sub>3</sub>, –Cl, –F, –CN, and –OCH<sub>3</sub>) (Table 1). This is in agreement with previous calculations where the solvation sphere of pyridine nitrogen was found to be embedded by the spheres of the neighbor atoms making difficult the correct evaluation of the electrostatic effects by the continuum solvent on the nitrogen portion of the cavity [18]. In fact, the results of Table 1 for pyridines substituted at position 2 reinforce the previous hypothesis as long as the largest errors are found for electronegative or hardly polarizable substituents, which may present abrupt changes



of the electric field in the local region of the pyridine nitrogen, especially in pyridinium cations.

On the other hand, the calculated  $pK_a$  values for pyridines with electronegative substituents are underestimated (i.e., pyridinium acidity is overestimated), which is probably due to the lack of stabilization of the protonated nitrogen by the pure continuum model in the acid species. For carbon acids, the trend indicates that acidity is typically overestimated for anionic species (C2 and C9) and underestimated for cationic species (C4, C5, and C10), while for the neutral ones (including zwitterions) both cases are found. These results indicate that the stability ionic species is not properly considered by the continuum model. Considering that the most important interaction between ionic groups and the solvent is the electrostatic one, the immediate conclusion is that errors are mainly due to errors in energies of such interactions, which is again pointing out that the errors are strongly dependent on the solute–continuum electrostatic interaction.

The explicit treatment of water molecules together with the continuum solvent model has been demonstrated to improve the  $pK_a$  predictions with thermodynamic cycles by including the real quantum hydrogen bond interactions and by breaking the artificial boundary delimited by the dielectric continuum and therefore reducing the influence of the solute–continuum electrostatic interaction [16, 18, 19, 25, 42, 56–59]. Since the artificial interaction between the dielectric continuum and the solute also limits the performance of the aqueous isodesmic reaction, a similar behavior is expected when explicit water molecules are included. According to this, a water molecule was added to both pyridine and pyridinium molecules acting as hydrogen bond donor/acceptor, respectively. In the case of carbon acids, all the acidic carbons are placed in  $\alpha$ -position respect to a carbonyl group. So, a single water molecule was placed (in both acids and conjugate bases) donating a hydrogen bond to the carbonyl oxygen to describe the enolate resonance structures according to Ho and Coote's criterion [16].

The calculated  $pK_a$  values of the pyridine and carbon acid species in the presence of an explicit water molecule are shown in Tables 3 and 4, respectively. As shown, the presence of water molecules improves the mean absolute deviations by 0.20–0.42  $pK_a$  units for the pyridine species. For carbon acids, the MAD values are improved by 0.3  $pK_a$  units in the case of B3LYP calculations, whereas for HF and CBS methods the results are not improved or slightly worsened. As deduced by comparison of the  $pK_a$  values from Tables 1 and 3, the largest improvement takes place in those pyridines that present an electronegative substituent (P6–P17), in particular if such group is placed in position 2 (P6, P8, P11, and P14), which points out that the reason of the improvement is a better description of the

electrostatic continuum–solute interactions in agreement with previous  $pK_a$  calculations that used thermodynamic cycles [18]. However, it should be noted that for bare pyridines, the aqueous isodesmic reaction provides better results than the thermodynamic cycles [18].

An analogous comparison of the values reported in Tables 2 and 4 indicates that the presence of water molecules do not improve systematically the  $pK_a$  predictions for carbon acids. Actually, it can be seen that all the  $pK_a$  values are increased in a similar extent worsening or improving those values that either are or are not well predicted for the bare carbon acid species independently of their global charge. A closer inspection reveals that the free energy of deprotonation of each carbon acid increases approximately by 1–2 kcal mol<sup>-1</sup>, whereas the free energy of deprotonation of the acid reference increases by 4 kcal mol<sup>-1</sup> producing such shifting effect in all  $pK_a$  values.

In the case of the pyridine species, where the addition of explicit solvent has clear effects (Tables 1 and 3), comparison of the results provided by all HF, B3LYP, and CBS methods points out that to improve the  $pK_a$  predictions when explicit waters are present, electron correlation effects should be taken into account to treat hydrogen bonding interactions correctly. On the other hand, CBS-QB3 and CBS-4B3 calculations provide similar  $pK_a$  values than the CBS-4B3\* and B3LYP ones (which include diffuse functions for the geometry optimizations) for most of the microsolvated species (pyridines and carbon acids in Tables 3 and 4), which points out that the rigorous treatment of short-range hydrogen bonding interactions might not be the unique reason for the improvement in  $pK_a$  predictions.

In order to understand the relationship between the presence of explicit water molecules and the improvement in  $pK_a$  calculations, the electrostatic and non-electrostatic contributions of CBS-4B3\* free energies in solution were analyzed. The electrostatic effects of the continuum model at SCRf level reach 8.66/62.13 kcal mol<sup>-1</sup> (averages calculated from absolute values) for the bare pyridines/pyridinium cations, respectively, in the dielectric medium. However, what determines the continuum decisive contribution to  $pK_a$  values in the aqueous isodesmic reaction is the difference in deprotonation energies between the acid and the reference as in (7).

For the bare pyridines, the average differences in the electrostatic contributions between each acid/base couple  $\{\Delta G_{el} = G_{el}(A) - G_{el}(AH^+)\}$  is 53.5 kcal mol<sup>-1</sup>, while the average contribution respect to the reference species  $\{\text{i.e., } \Delta \Delta G_{el} = \Delta G_{el}(A/AH^+) - \Delta G_{el}(\text{ref}/\text{refH}^+)\}$  is 4.22 kcal mol<sup>-1</sup>  $\pm$  2.90 kcal mol<sup>-1</sup> (MAD  $\pm$  SD). The presence of a single explicit water molecule interacting with the pyridine nitrogen produces significant changes in the electrostatic contributions, reducing  $\Delta G_{el}$  by almost 20% to an

**Table 3** Calculated  $pK_a$  values for the single water microsolvated substituted pyridines using the aqueous isodesmic reaction according to (5)

	Exptl.	HF (5)	B3LYP (5)	CBS-4B3* (5)	CBS-4B3 (5)	CBS-QB3 (5)
PRef	5.23 [43]					
P1	5.97 [44]	6.55 (0.58)	6.94 (0.97)	6.66 (0.69)	6.97 (1.00)	6.90 (0.93)
P2	5.68 [44]	5.75 (0.07)	6.24 (0.56)	6.01 (0.33)	6.04 (0.36)	6.04 (0.36)
P3	6.02 [44]	6.16 (0.14)	6.14 (0.12)	5.93 (0.09)	5.79 (0.23)	5.71 (0.31)
P4	6.57 [43]	7.39 (0.82)	7.62 (1.05)	7.27 (0.70)	7.32 (0.75)	7.25 (0.68)
P5	6.99 [43]	7.41 (0.42)	7.75 (0.76)	7.39 (0.40)	8.15 (1.16)	8.00 (1.01)
P6	0.49 [43]	−0.95 (1.44)	0.58 (0.09)	0.56 (0.07)	0.55 (0.06)	0.62 (0.13)
P7	2.81 [43]	2.42 (0.39)	2.99 (0.18)	2.83 (0.02)	3.10 (0.29)	3.16 (0.35)
P8	−0.44 [43]	−1.71 (1.27)	−0.75 (0.31)	−0.91 (0.47)	−0.53 (0.09)	−0.50 (0.06)
P9	2.97 [44]	2.68 (0.29)	3.00 (0.03)	3.12 (0.15)	3.38 (0.41)	3.42 (0.45)
P10	2.84 [43]	2.65 (0.19)	3.03 (0.19)	2.70 (0.14)	3.09 (0.25)	3.17 (0.33)
P11	−0.26 [45]	−3.72 (3.46)	−1.31 (1.05)	−1.70 (1.44)	−1.65 (1.39)	−1.54 (1.28)
P12	1.45 [45]	−0.31 (1.76)	0.41 (1.04)	0.26 (1.19)	0.69 (0.76)	0.74 (0.71)
P13	1.90 [45]	−0.33 (2.23)	0.75 (1.15)	0.59 (1.31)	0.85 (1.05)	0.91 (0.99)
P14	3.28 [44]	2.46 (0.82)	3.58 (0.30)	3.11 (0.17)	3.17 (0.11)	3.08 (0.20)
P15	4.88 [44]	5.05 (0.17)	5.75 (0.87)	5.34 (0.46)	5.80 (0.92)	5.88 (1.00)
P16	6.62 [44]	7.60 (0.98)	7.65 (1.03)	7.42 (0.80)	7.67 (1.05)	7.51 (0.89)
P17	0.81 [44]	−1.41 (2.22)	−0.54 (1.35)	−0.39 (1.20)	0.07 (0.74)	0.12 (0.69)
P18	3.40 [44]	2.71 (0.69)	3.03 (0.37)	2.79 (0.61)	3.50 (0.10)	3.51 (0.11)
P19	3.18 [44]	3.07 (0.11)	2.64 (0.54)	3.05 (0.13)	3.37 (0.19)	3.35 (0.17)
MAD		0.95	0.63	0.55	0.57	0.56
SD		0.92	0.43	0.46	0.43	0.38
ADmax		3.46	1.35	1.44	1.39	1.28

The absolute errors in  $pK_a$  units are given in brackets

**Table 4** Calculated  $pK_a$  values for the single water microsolvated carbon acids using the aqueous isodesmic reaction according to (5)

	Exptl.	HF (5)	B3LYP (5)	CBS-4B3* (5)	CBS-4B3 (5)	CBS-QB3 (5)
CRef	29.4 ± 0.5 [52]					
C1	28.9 ± 0.5 [51]	27.7 (1.2)	28.7 (0.2)	29.3 (0.4)	29.6 (0.7)	29.4 (0.5)
C2	34 [52]	38.6 (4.6)	36.9 (2.9)	37.9 (3.9)	38.8 (4.8)	38.6 (4.6)
C3	27.3 ± 1.2 [51]	30.0 (2.7)	31.4 (4.1)	31.2 (3.9)	31.7 (4.4)	31.5 (4.2)
C4	18.0 ± 1.0 [51]	16.2 (1.8)	15.7 (2.3)	16.2 (1.8)	16.0 (2.0)	15.7 (2.3)
C5	21.0 ± 1.0 [50]	19.9 (1.1)	19.7 (1.3)	20.4 (0.6)	20.3 (0.7)	20.4 (0.6)
C6	33.5 [52]	33.9 (0.4)	35.7 (2.2)	35.8 (2.3)	36.2 (2.7)	36.5 (3.0)
C7	17 [53]	17.0 (0.0)	19.3 (2.3)	19.4 (2.4)	20.4 (3.4)	19.6 (2.6)
C8	30.3 [54]	31.2 (0.9)	31.7 (1.4)	32.4 (2.1)	32.0 (1.7)	31.0 (0.7)
C9	30.8 [54]	39.6 (8.8)	35.6 (4.8)	38.8 (8.0)	39.6 (8.8)	39.7 (8.9)
C10	14 ± 1 [55]	17.2 (3.2)	13.7 (0.3)	13.9 (0.1)	13.9 (0.1)	13.1 (0.9)
C11	28.4 ± 0.5 [52]	27.3 (1.1)	27.7 (0.7)	28.0 (0.4)	28.3 (0.1)	27.8 (0.6)
C12	29.1 [54]	27.8 (1.3)	28.8 (0.3)	29.5 (0.4)	28.3 (0.8)	28.0 (1.1)
C13	23.9 [54]	28.4 (4.5)	24.5 (0.6)	28.4 (4.5)	27.0 (3.1)	27.2 (3.3)
C14	25.6 ± 0.5 [49]	26.6 (1.0)	26.3 (0.7)	26.7 (1.1)	26.6 (1.0)	26.0 (0.4)
MAD		2.3	1.7	2.3	2.5	2.4
SD		2.3	1.4	2.2	2.4	2.4
ADmax		8.8	4.8	8.0	8.8	8.9

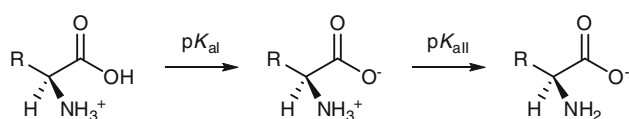
The absolute errors in  $pK_a$  units are given in brackets

average value of  $43.57 \text{ kcal mol}^{-1}$ . In addition,  $\Delta\Delta G_{\text{el}}$  falls to  $2.92 \pm 2.27 \text{ kcal mol}^{-1}$ .

Some fraction of such a decrease in both  $\Delta G_{\text{el}}$  and  $\Delta\Delta G_{\text{el}}$  is due to an equalization of the electrostatic solute–continuum interaction between the involved species in the proton exchange reaction (Scheme 3) by the presence of water molecules. To get some insight into that, the individual electrostatic interaction contributions [ $G_{\text{el}}(N_{\text{pyr}})$ ] of the cavity region around each atom can be recovered. Accordingly, the average difference between the electrostatic interactions of the bare pyridine and pyridinium nitrogen atoms [ $\Delta G_{\text{el}}(N_{\text{pyr}})$ ] is  $14.28 \pm 3.07 \text{ kcal mol}^{-1}$ , and the average difference of  $\Delta G_{\text{el}}(N_{\text{pyr}})$  with respect to the pyridine reference is  $\Delta\Delta G_{\text{el}}(N_{\text{pyr}}) = 2.87 \pm 1.66 \text{ kcal mol}^{-1}$ .

In contrast, when water molecules are included, the electrostatic interactions between the dielectric continuum and the nitrogen atoms are nearly cancelled [ $\Delta G_{\text{el}}(N_{\text{pyr}}) = 0.57 \text{ kcal mol}^{-1} \pm 0.28 \text{ kcal mol}^{-1}$  and  $\Delta\Delta G_{\text{el}}(N_{\text{pyr}}) = 0.31 \text{ kcal mol}^{-1} \pm 0.25 \text{ kcal mol}^{-1}$ ]. As indicated in a previous work, pyridine–water hydrogen bonding interactions polarize the water molecule, so that the continuum model calculates the electrostatic interaction with the water molecule in a much larger surface reducing large part of the errors of bare species [18].

In fact, the water–continuum electrostatic interactions contributions to the total  $\Delta G_{\text{el}}$  and  $\Delta\Delta G_{\text{el}}$  are  $12.21 \text{ kcal mol}^{-1} \pm 1.37 \text{ kcal mol}^{-1}$  and  $1.41 \text{ kcal mol}^{-1} \pm 1.03 \text{ kcal mol}^{-1}$ , respectively (comparable to that of bare nitrogen atoms). Nevertheless, it should be noted that the average surface of the solvation cavity over bare nitrogen is  $9.0 \text{ \AA}^2$ , while the average surface over water molecules is considerably larger ( $34.3 \text{ \AA}^2$ ).



**Scheme 4** Successive ionization reactions considered for the  $pK_a$  prediction of amino acids

Eventually, the non-electrostatic contributions to the free energies in solution can be also taken into account. In the PCM models (including CPCM), cavitation, dispersion, and repulsion energies are calculated classically and are dependent on the volume of the solute cavity [1, 5, 35]. Therefore, the magnitude of such energies will remain approximately constant if acids and their conjugate bases are geometrically similar and present comparable cavity sizes.

According to the UAHF and UAKS model cavities employed in this work, the solvation spheres are placed only over non-hydrogen atoms. However, the charge and the number of hydrogens linked to a heavy atom do modify (slightly) the final radius of the sphere [30]. Consequently, small variations of the volume cavities are expected due to the deprotonation reaction.

For the bare pyridines, the absolute differences in the non-electrostatic energies between bases and acids [ $\Delta G_{\text{nes}} = G_{\text{nes}}(\text{A}) - G_{\text{nes}}(\text{AH}^+)$ ] are  $0.08 \pm 0.06 \text{ kcal mol}^{-1}$  (MAD  $\pm$  SD), whereas the average (MAD  $\pm$  SD) difference respect to the pyridine reference is  $\Delta\Delta G_{\text{nes}} = 0.19 \pm 0.09 \text{ kcal mol}^{-1}$ . Larger differences may be expected when water molecules are present since they can interact as proton donor or acceptor depending on the protonation state of the pyridine molecules. However, the orientation of the hydrogen atoms of water molecules does not change the cavity volume. Eventually, the obtained  $\Delta G_{\text{nes}}$  and  $\Delta\Delta G_{\text{nes}}$  average differences for the water–pyridine clusters are (MAD  $\pm$  SD)  $0.46 \pm 0.15 \text{ kcal mol}^{-1}$  and  $0.28 \pm 0.14 \text{ kcal mol}^{-1}$ , respectively.

As a suggestion of a referee, we employed the aqueous isodesmic reaction schema to calculate the  $pK_a$  values of problematic species for thermodynamic cycles such as amino acid species, which deprotonation reactions in aqueous solution involve gas-phase unstable zwitterionic species. Hydrophobic side-chain amino acids (i.e., glycine, alanine, valine, leucine, and methionine) were considered in order to unambiguously study the carboxylic acid and amine ionizations (Scheme 4) and to avoid interferences from any other acid or base group. As depicted in

**Table 5** Calculated  $pK_a$  values for the first ionization reaction in aqueous solution of amino acid species

	Exptl.	HF (5)	B3LYP (5)	CBS-4B3 (5)	CBS-4B3* (5)	CBS-QB3 (5)
Glycine (Ref)	2.34 [43]					
Alanine	2.33 [43]	1.70 (0.63)	1.96 (0.37)	1.79 (0.54)	1.76 (0.57)	1.78 (0.55)
Valine	2.27 [43]	2.19 (0.08)	2.47 (0.20)	1.31 (0.96)	2.06 (0.21)	2.05 (0.22)
Leucine	2.32 [43]	2.37 (0.05)	2.26 (0.06)	2.22 (0.10)	1.94 (0.38)	1.95 (0.37)
Methionine	2.16 [43]	1.82 (0.34)	1.91 (0.25)	1.60 (0.56)	1.42 (0.74)	1.39 (0.77)
MAD		0.27	0.22	0.54	0.47	0.48
SD		0.27	0.13	0.35	0.23	0.24
ADmax		0.63	0.37	0.96	0.74	0.77

**Table 6** Calculated  $pK_a$  values for the second ionization reaction in aqueous solution of amino acid species

	Exptl.	HF (5)	B3LYP (5)	CBS-4B3* (5)	CBS-4B3 (5)	CBS-QB3 (5)
Glycine (Ref)	9.58 [43]					
Alanine	9.71 [43]	9.95 (0.24)	9.68 (0.03)	9.64 (0.07)	9.43 (0.28)	9.35 (0.36)
Valine	9.52 [43]	9.01 (0.51)	9.07 (0.45)	8.99 (0.53)	9.46 (0.06)	9.38 (0.14)
Leucine	9.58 [43]	9.91 (0.33)	9.98 (0.40)	9.61 (0.03)	10.15 (0.57)	10.05 (0.47)
Methionine	9.08 [43]	9.44 (0.36)	9.30 (0.22)	9.19 (0.11)	9.71 (0.63)	9.62 (0.54)
MAD		0.36	0.27	0.19	0.38	0.38
SD		0.11	0.19	0.23	0.26	0.18
ADmax		0.51	0.45	0.53	0.63	0.54

Scheme 4, no tautomeric equilibrium was considered for the intermediate protonation state since the majority of species present the zwitterionic state in aqueous solution.

The first and second deprotonation reactions of glycine were used in combination with their experimental  $pK_a$  values (according to the Scheme 3) as reference reactions for the prediction of  $pK_{aI}$  and  $pK_{aII}$  values, respectively, of the other considered amino acid species. It has to be noted that no explicit water molecules were used in any of these calculations. As shown in Tables 5 and 6, the calculated  $pK_a$  values for both ionization processes are in excellent agreement with the experimental values for all species independently of the quantum chemistry method. Considering the first deprotonation reaction (corresponding to the carboxylic acid functionalities), the best results are provided by B3LYP and HF methods with mean absolute deviation (MAD) values below 0.3  $pK_a$  units, while the deviations (MAD) corresponding to CBS methods are slightly larger (i.e., below 0.54  $pK_a$  units). Concerning the second ionization process (corresponding to the amine groups), CBS-4B3\* and B3LYP yield the best prediction respect to the experimental values despite all the methods provide similar results with MAD values ranging 0.19–0.38  $pK_a$  units. These results show the potential applicability of the aqueous isodesmic reaction for those cases which gas-phase calculations are problematic for the thermodynamic cycles.

**Acknowledgments** This work was funded by the Spanish Government in the framework of Project CTQ2008-02207/BQU. One of us (R. C.) wishes to acknowledge a fellowship from the Spanish MEC within the FPU program. The authors are grateful to “Centro de Cálculo de Supercomputación de Galicia” (CESGA) and to “Centre de Supercomputació de Catalunya” (CESCA) for access to their computational facilities.

## References

- Tomasi J, Persico M (1994) Chem Rev 94:2027–2094
- Cramer CJ, Truhlar DG (1999) Chem Rev 99:2161–2200
- Orozco M, Luque FJ (2000) Chem Rev 100:4187–4225
- Tomasi J (2004) Theor Chem Acc 112:184–203
- Tomasi J (2005) Chem Rev 105:2999–3093
- Cramer CJ, Truhlar DG (2008) Acc Chem Res 41:760–768
- Ho J, Coote ML (2010) Theor Chem Acc 125:3–21
- Liptak MD, Shields GC (2001) J Am Chem Soc 123:7314–7319
- Toth AM, Liptak MD, Phillips DL, Shields GC (2001) J Chem Phys 114:4595–4606
- Liptak MD, Shields GC (2001) Int J Quant Chem 85:727–741
- Liptak MD, Gross KC, Seybold PG, Feldgus S, Shields GC (2002) J Am Chem Soc 124:6421–6427
- Król M, Wrona M, Page CS, Bates PA (2006) J Chem Theory Comput 2:1520–1529
- Charif IE, Mekelleche SM, Villemin D, Mora-Díez N (2007) J Mol Struct Theochem 818:1–6
- Gao D, Svoronos P, Wong PK, Maddalena D, Hwang J, Walker H (2005) J Phys Chem A 109:10776–10785
- Brown TN, Mora-Díez N (2006) J Phys Chem B 110:9270–9279
- Ho J, Coote ML (2009) J Chem Theory Comput 5:295–306
- Sadlej-Sosnowska N (2007) Theor Chem Acc 118:281–293
- Casasnovas R, Frau J, Ortega-Castro J, Salvà A, Donoso J, Muñoz F (2009) J Mol Struct Theochem 912:5–12
- Pliego JR Jr, Riveros JM (2002) J Phys Chem A 106:7434–7439
- Zhan C, Dixon DA (2001) J Phys Chem A 105:11534–11540
- Tissandier MD, Cowen KA, Feng WY, Gundlach E, Cohen MH, Earhart AD, Coe JV (1998) J Phys Chem A 102:7787–7794
- Kelly CP, Cramer CJ, Truhlar DG (2006) J Phys Chem B 110:16066–16081
- Palascak MW, Shields GC (2004) J Phys Chem A 108:3692–3694
- Pliego JR Jr, Riveros JM (2000) J Phys Chem B 104:5155–5160
- Kelly CP, Cramer CJ, Truhlar DG (2005) J Chem Theory Comput 1:1133–1152
- Takano Y, Houk KN (2005) J Chem Theory Comput 1:70–77
- Govender KK, Cukrowski I (2009) J Phys Chem A 113:3639–3647
- Govender KK, Cukrowski I (2010) J Phys Chem A 114:1868–1878
- Ho J, Klamt A, Coote ML (2010) J Phys Chem A 114:13442–13444
- Barone V, Cossi M, Tomasi J (1997) J Chem Phys 107:3210–3221
- 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, 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, Yazayev 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, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2004) Gaussian 03. Gaussian, Inc, Wallingford
32. Ochterski JW, Petersson GA, Montgomery JA Jr (1996) *J Chem Phys* 104:2598–2619
33. Montgomery JA Jr, Frisch JM, Ochterski JW, Petersson GA (1999) *J Chem Phys* 104:2822–2827
34. Montgomery JA Jr, Frisch JM, Ochterski JW, Petersson GA (2000) *J Chem Phys* 104:6532–6542
35. Barone V, Cossi M (1998) *J Phys Chem A* 102:1995–2001
36. Cossi M, Rega N, Scalmani G, Barone V (2003) *J Comput Chem* 24:669–681
37. Casanovas R, Frau J, Ortega-Castro J, Salvà A, Donoso J, Muñoz F (2010) *Int J Quant Chem* 110:323–330
38. Scott AP, Radom L (1996) *J Phys Chem* 100:16502–16513
39. Pokon EK, Liptak MD, Feldgus S, Shields GC (2001) *J Phys Chem A* 105:10483–10487
40. Pickard FC, Griffith DR, Ferrara SJ, Liptak MD, Kirschner KN, Shields GC (2006) *Int J Quant Chem* 106:3122–3128
41. Sanz P, Mó O, Yáñez M, Elguero J (2007) *J Phys Chem A* 111:3585–3591
42. Kelly CP, Cramer CJ, Truhlar DG (2006) *J Phys Chem A* 110:2493–2499
43. Lide DR (ed) (2007) *CRC handbook of chemistry and physics*, 87th edn. CRC Press Inc., Boca Raton
44. Williams R, pKa compilation. [http://research.chem.psu.edu/brpgroup/pKa\\_compilation.pdf](http://research.chem.psu.edu/brpgroup/pKa_compilation.pdf). Accessed 28 July 010
45. Albert A (1963) Dipolar cycloaddition chemistry. In: Katritzky AR (ed) *Physical methods in heterocyclic chemistry*, vol 1. Academic Press, NY
46. Angulo G, Carmona C, Pappalardo RR, Muñoz MA, Guardado P, Sánchez-Marcos E, Balón M (1997) *J Org Chem* 62:5104–5109
47. Chen IJ, MacKerell AD Jr (2000) *Theor Chem Acc* 103:483–494
48. Caballero NA, Melendez FJ, Muñoz-Caro C, Niño A (2006) *Biophys Chem* 124:155–160
49. Amyes TL, Richard JP (1996) *J Am Chem Soc* 118:3129–3141
50. Rios A, Richard JP (1997) *J Am Chem Soc* 119:8375–8376
51. Rios A, Amyes TL, Richard JP (2000) *J Am Chem Soc* 122:9373–9385
52. Rios A, Richard JP, Amyes TL (2002) *J Am Chem Soc* 124:8251–8259
53. Chiang Y, Kresge J, Pruszynski P (1992) *J Am Chem Soc* 114:3103–3107
54. Richard JP, Williams G, O'Donoghue AC, Amyes TL (2002) *J Am Chem Soc* 124:2957–2968
55. Rios A, Crueiras J, Richard JP (2001) *J Am Chem Soc* 123:7949–7950
56. Sánchez-Marcos E, Terryn B, Rivail JL (1985) *J Phys Chem* 89:4695–4700
57. Claverie P, Daudey JP, Langlet J, Pullman B, Plazzola D, Huron MJ (1978) *J Phys Chem* 82:405–418
58. Vyacheslav S, Bryantsev S, Diallo MS, Goddard WA III (2008) *J Phys Chem B* 112:9709–9719
59. Pliego JR Jr, Riveros JM (2001) *J Phys Chem A* 105:7241–7247