Theor Chem Acc (2007) 118:281–293
DOI 10.1007/s00214-006-0232-z

REGULAR ARTICLE

# Calculation of acidic dissociation constants in water: solvation free energy terms. Their accuracy and impact

**Nina Sadlej-Sosnowska**

Received: 22 December 2005 / Accepted: 21 November 2006 / Published online: 19 January 2007


**Abstract** Three polarizable continuum models, DPCM, CPCM, and IEFPCM, have been applied to calculate free energy differences for nine neutral compounds and their anions. On the basis of solvation free energies, the $pK_a$ values were obtained for the compounds in question by using three thermodynamic cycles: one, involving the combined experimental and calculated data, as well as two other cycles solely with calculated data. This paper deals with the influence of factors such as the SCRF model applied, choice of a particular thermodynamic cycle, atomic radii used to build a cavity in the solvent (water), optimization of geometry in water, inclusion of electron correlation, and the dimension of the basis set on the solvation free energies and on the calculated $pK_a$ values.

**Keywords** Solvation free energy · Acidity · $pK_a$ · Polarizable continuum model

## 1 Introduction

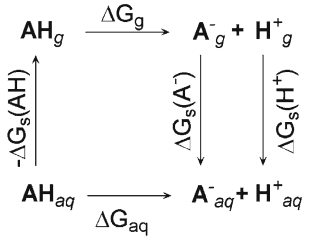
The tendency of a molecule to lose its hydrogen atom as an acidic proton is quantified as $pK_a$. During the last two decades there has been much interest in developing a methodology enabling theoretical prediction of $pK_a$ values, employing various quantum theoretical techniques.

As $pK_a$ equals $\Delta G/2.303\ RT$, where $\Delta G$ is a free energy change of the dissociation reaction either in a gas or solution, acidity of a compound is determined by the $\Delta G$ value. In the gas phase, calculation of $\Delta G_g$ is a fairly straightforward procedure, based on gas-phase geometry optimization and on vibrational analysis of the acid and its anion.

In order to compute the $pK_a$ values in an aqueous solution, in the framework of a continuum solvation model [1–29], the most frequently used procedure relies on an application of the appropriate thermodynamic cycles in which the computed gas-phase $\Delta G_g$ values are combined with the free energies of solvation for each of the species involved in the equilibrium, in order to obtain the value of total free energy change in a solution. The continuum solvation model belongs to the category of the so-called implicit models [30]. Here we shall present and discuss the results obtained only within the model.

Thermodynamic cycles can be constructed in several ways. The illustration below shows the simplest thermodynamic cycle which was at the same time the one most frequently applied with the purpose of $pK_a$ calculation in aqueous solution [4,5,10–22,26,28]:

Cycle A

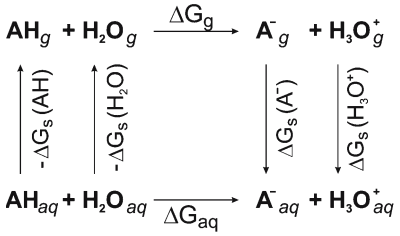



**Electronic supplementary material** The online version of this article (doi:10.1007/s00214-006-0232-z) contains supplementary material, which is available to authorized users.

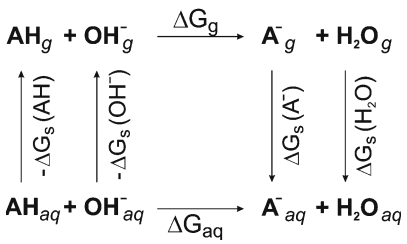
N. Sadlej-Sosnowska (✉)
National Institute of Public Health,
30/34 Chełmska Street, 00-725 Warsaw, Poland
e-mail: sadlej@il.waw.pl

The gas-phase dissociation reaction forms an upper leg of the cycle. The vertical legs consist of free energies of a neutral molecule, an anion, and proton solvation. Cycle A has been generally used for calculation of all the reported "experimental" free energies of solvation of ions, relative to experimental $\Delta G_s(H^+)$, with the $\Delta G_{gas}$ and $\Delta G_{aq}$ values also arrived at by way of experiment [31–35]. For the calculation of $pK_a$ in an aqueous solution two other cycles, named here B and C, have also been applied [3,5,6,8,9,23–25,29]. The two cycles are shown below.

Cycle B




Cycle C




As can be seen, the vertical legs are formed by the free energies of solvation of all species involved. The lower horizontal legs of the three cycles display three water-phase equilibria characteristic for the three cycles.

It should be noted that none of the cycles provides an exhaustive description of the equilibria taking place in aqueous solutions which are far more complicated and involve heterogeneous ion-water clusters of incompletely identified stoichiometry, and as such are far from being thoroughly recognized.

If one neglects the more complicated water-phase equilibria involving ion–water clusters other than $H_3O^+$ occurring in Cycle B, and consistent with yet another set of cycles [1,24,35], the $pK_a$ value of a given HA can be determined by using one of the three cycles, A, B, or C.

The $pK_a$ calculated on the basis of Cycle A, $pK_a^A$, is expressed as

$$2.303\,RT\mathrm{p}K_a^A = \Delta G_{aq} = G_g(A^-) + G_g(H^+) - G_g(HA) + \Delta G_s(H^+) + \Delta G_s(A^-) - \Delta G_s(HA) \quad (1)$$

whereas the corresponding expressions for $pK_a$'s calculated based on the two other cycles are as follows [23,24]:

$$2.303\,RT\mathrm{p}K_a^B = G_g(A^-) + G_g(H_3O^+) - G_g(HA) - G_g(H_2O) + \Delta G_s(A^-) + \Delta G_s(H_3O^+) - \Delta G_s(HA) - \Delta G_s(H_2O) - 2.303\,RT\log[H_2O]_{aq} \quad (2)$$

$$2.303\,RT\mathrm{p}K_a^C = G_g(A^-) + G_g(H_2O) - G_g(HA) - G_g(OH^-) + \Delta G_s(A^-) + \Delta G_s(H_2O) - \Delta G_s(HA) - \Delta G_s(OH^-) + 2.303\,RT\mathrm{p}K_a(H_2O) \quad (3)$$

Subtracting Eq. (1) from Eq. (2) one obtains

$$2.303\,RT(\mathrm{p}K_a^B - \mathrm{p}K_a^A) = G_g(H_3O^+) - G_g(H^+) - G_g(H_2O) + \Delta G_s(H_3O^+) - \Delta G_s(H_2O) - \Delta G_s(H^+) - 2.303\,RT\log[H_2O]_{aq} \quad (4)$$

The right-hand side of Eq. (4) constitutes the sum of free energy terms in the thermodynamic Cycle A for dissociation of the $H_3O^+$ acid and should equal zero provided all the free energy terms took their absolute, true values.

Similarly, on subtraction of Eq. (1) from Eq. (3) one obtains:

$$2.303\,RT(\mathrm{p}K_a^C - \mathrm{p}K_a^A) = G_g(H_2O) - G_g(H^+) - G_g(OH^-) + \Delta G_s(H_2O) - \Delta G_s(H^+) - \Delta G_s(OH^-) + 2.303\,\mathrm{RT}\,\mathrm{p}K_a(H_2O) \quad (5)$$

This time the right-hand side of Eq. (5) constitutes the sum of free energy terms in the thermodynamic Cycle A for dissociation of $H_2O$ and would also equal zero provided all the free energy terms took their absolute, true values.

What follows, under these conditions, $pK_a^A = pK_a^B = pK_a^C$. The equality also holds in the case when the experimental values taken for $\Delta G_g$ in Eqs. (2) and (3) are the same as those which were applied for the calculation of the "experimental" values of $\Delta G_s(H_3O^+)$ and $\Delta G_s(OH^-)$, and the latter two are inserted in Eqs. (2) and (3). In fact, it has been found that for six carboxylic acids, mean unsigned error (m.u.e.) was nearly the same regardless whether Cycle A [13] or Cycle B were used with a reliable (experimental) solvation free energy of $H_3O^+$ [23].

However, one can also find statements that application of one of the Cycles, B or C [6,24,25], generates better results than application of Cycle A. The differences

stem from taking the calculated values rather than the experimental ones for $\Delta G_g$ and/or $\Delta G_s(H_3O^+)$ and $\Delta G_s(OH^-)$.

This investigation mainly aimed to analyse the impact of the calculated free energies of solvation of acids, HA, and the corresponding anions $A^-$ on the calculated $pK_a$ accuracy (agreement with experimental data). To this end, experimental values were applied to the free energy differences in the gas phase and to free energy of proton solvation, while Cycle A has been used. However, the Reviewer recommended that the authors include theoretical calculations of all of the species involved in the acid–base reactions, which is possible only for Cycles B and C because $\Delta G_s(H^+)$, occurring in Cycle A, cannot be calculated directly and tends to be derived experimentally. Therefore, the results are also presented for Cycles B and C with theoretical $\Delta G_s(H_3O^+)$ and $\Delta G_s(OH^-)$. The $\Delta G_g$ values were also replaced by the ones calculated by using the multilevel complete basis set method (CBS-QB3) [36,37], since calculations of $\Delta G_g$, at this level, combined with solvation models, rendered very accurate $pK_a$ values [13,14].

For calculations with Cycle A, $\Delta G_s(H_2O)$ and $\Delta G_s(H^+)$ have been taken from subject literature. The first value equals to −6.32 kcal/mol [38]. However, the second value has evoked much controversy for a long time. Currently, the values proposed by Tissandier et al., [39] namely $-264.0 \pm 0.2$ and Tuttle et al. [40] $-263.7 \pm 0.2$ kcal/mol, determined on the basis of cluster ion solvation data, seem to be the most accurate [41]. The first value has been adopted here. It has recently been reproduced with accuracy of some tenths of kcal/mole [35,42].

In this context we ought to address the issue of standard states. All experimental and calculated gas-phase free-energies (Table 1S) are quoted using an ideal gas at 1 atm as a reference state. All "experimental" solvation free energies (used in Tables 1–3 and 2S–7S) were calculated for an ideal gas at a gas-phase concentration of 1 mol/L dissolving in liquid phase at the same concentration. In order to convert gas phase data from one convention to another (1 atm to 1 mole/L) we refer to ref. [30]. For the gas phase reaction A→ B + C (as in Cycle A),

$$\Delta G(1\ \text{mol/L}) = \Delta G(1\ \text{atm}) + RT\ \ln(24.5) \qquad (6)$$

In the case of reactions $A + B \rightarrow C + D$, as in Cycles B and C, the $\Delta G$ values for both conventions are equal.

The value for $\Delta G_s(H^+)$ reported by Tissandier et al. (−264.0 kcal/mol) concerns the standard state concentration of 1 atm in gas phase and 1 mol/L in aqueous phase. Converting to the standard state that uses a concentration of 1 mol/L in both phases requires that $RT$ ln 24.5 (1.9 kcal/mol) be subtracted from this value producing the value of −265.9 kcal/mol. The issue of standard state for the theoretical solvation free energies will be addressed in Results.

Among the solvation models, referred to as self-consistent reaction field (SCRF), the polarizable continuum model (PCM) is presumably the most frequently used one [43] and the one that has undergone a number of evolutionary changes (compare Gaussian 03 with Gaussian 98 programs for reference) [44,45]. The original PCM version [46,47] is now being named Dielectric PCM (DPCM) to distinguish it from the two successive reformulations of the model, CPCM [48] and integral equation formalism PCM (IEFPCM) [49,50]. In the G98 and G03 programs all three versions are implemented. In G98, DPCM is the default version of PCM, whereas in the latest G03 version IEFPCM has become the default PCM formalism. Detailed description, comparison, evaluation, and applications of the models and other continuum solvation ones can be found in ref. [30] and in the reviews [51–54].

All continuum models require definition of the shape and size of a cavity occupied by a solute molecule in the solvent. PCM models built the cavities in the form of envelopes of spheres centered on atoms and atomic groups. The Gaussian programs offer several possibilities of choosing a set of the atomic radii. In G98, the set Radii = UAHF is the default one, where UAHF stands for the United Atom Topological Model for Hartree–Fock [55]. In G03, the default setting has changed to Radii = UA0 (the United Atom Topological Model applied on atomic radii of the UFF force field) [56].

The papers experimenting with factors such as various solvation models, different calculation levels, and different basis sets tended to investigate the influence of the said factors on the final result, namely the value of $pK_a$. The results differed much in their agreement with the corresponding experimental data. For instance, the B3LYP method was reported to have given less accurate $pK_a$ results for substituted phenols [14] and substituted tetrazoles [17] than the HF did. This effect was mentioned by Barone and Cossi [48] and explained as resultant from underestimation of the solvent reaction field at MP2 and DFT level. However, in another study, the absolute $pK_a$ values were very close to experimental ones, but only when the solutes were treated at a correlated ab initio level [10]. The results of the calculation of the first two steps of deprotonation of histamine obtained with HF were much poorer than those obtained using B3LYP, for different basis sets [6]. Despite a number of papers dealing with the $pK_a$ calculations at many

**Table 1** $\Delta G_s$ calculated using DPCM model: experimental $\Delta G_s$

| Radii = UAHF $n = 9$ | Neutral molecules | | | | Anions | | | |
|---|---|---|---|---|---|---|---|---|
| | Gas-phase geometry IComp = 4[a] | | Geometry optimized in water, Icomp = 2[a] | | Gas-phase geometry IComp = 4[a] | | Geometry optimized in water, Icomp = 2[a] | |
| | m.u.e. | Mean error | m.u.e. | Mean error | m.u.e. | Mean error | m.u.e. | Mean error |
| HF/6-31+G* | 0.51 | −0.51 | 0.99 | −0.92 | 3.27 | 1.24 | 5.42 | 4.15 |
| HF/6-311++G** | 0.77 | −0.67 | 0.89 | −0.84 | 3.87 | 2.25 | 5.43 | 4.36 |
| B3LYP/6-31+G* | 0.93 | −0.57 | 1.69 | −1.52 | 3.78 | 3.78 | 7.21 | 6.41 |
| B3LYP/6-311++G** | 1.64 | −1.34 | 1.71 | −1.55 | 5.56 | 4.45 | 7.27 | 6.68 |
| HF/cc-pVDZ | 0.98 | 0.36 | 0.59 | −0.40 | 4.09 | 1.88 | 4.63 | 3.22 |
| HF/cc-pVTZ | 0.56 | 0.49 | 0.53 | −0.46 | 7.84 | 6.01 | 5.43 | 3.85 |
| B3LYP/cc-pVDZ | 1.54 | 0.55 | 1.69 | −1.04 | 5.57 | 4.69 | 6.34 | 5.13 |
| B3LYP/cc-pVTZ | 0.97 | 0.97 | 0.40 | −0.04 | 9.57 | 8.10 | 6.89 | 5.51 |

Gaussian 98

[a] Default settings in Gaussian 98. Results of the gas-phase geometry calculations using Icomp = 2 setting are displayed in Table 2S

**Table 2** $\Delta G_s$ calculated using IEFPCM model: experimental $\Delta G_s$

| Radii = UAHF | Neutral molecules, $n = 9$ | | | | Anions, $n = 8$ | |
|---|---|---|---|---|---|---|
| | Gas-phase geometry | | Geometry optimized in water | | Gas-phase geometry | Geometry optimized in water |
| | m.u.e. | Mean error | m.u.e. | Mean error | m.u.e. = mean error | m.u.e. = mean error |
| HF/6-31+G* | 0.56 | −0.56 | 0.72 | −0.72 | 5.16 | 3.79 |
| HF/6-311++G** | 0.45 | −0.45 | 0.62 | −0.62 | 5.30 | 3.91 |
| B3LYP/6-31+G* | 0.37 | −0.15 | 0.48 | −0.32 | 8.13 | 6.97 |
| B3LYP/6-311++G** | 0.25 | −0.06 | 0.42 | −0.30 | 8.34 | 6.80 |
| HF/cc-pVDZ | 0.28 | 0.17 | 0.33 | 0.04 | 5.17 | 3.84 |
| HF/cc-pVTZ | 0.20 | 0.01 | 0.27 | −0.13 | 5.04 | 3.69 |
| B3LYP/cc-pVDZ | 0.92 | 0.92 | 0.80 | 0.75 | 8.37 | 7.30 |
| B3LYP/cc-pVTZ | 0.51 | 0.51 | 0.38 | 0.37 | 7.95 | 6.47 |

Gaussian 03

theoretical levels, and the application of different basis sets, there is no general agreement as to which variations of the calculation methods may be relied on to produce better results than others.

Looking into the reason behind the discrepancies, we took here a few of AH/A-systems for which experimental values of $\Delta G_s(HA)$ and $\Delta G_s(A^-)$ are known, in order to compare accuracy of the two terms with the accuracy of $pK_a$. The solvation data for ions have been evolving until recently together with the evolution of the accepted value for $\Delta G_s(H^+)$. Finally, the values given by Kelly et al. [35] were used for nine neutral molecules, i.e. methanol, acetic acid, phenol, acetonitrile, methanethiol, toluene, thiophenol, acetone and $NH_3$, and their negatively ionized forms.

The present investigation is confined to the PCM set of solvation codes implemented in the recent Gaussian versions (G98 and G03), easily accessible to the computational chemists community.

## 2 Methodology

Using the thermodynamic Cycle A, the $pK_a$ values were calculated from Eqs. (1) and (6):

$$pK_a = \Delta G_{aq}/2.303\,RT = [\Delta G_g + 1.9 + \Delta G_s(A^-) - \Delta G_s(AH) - 265.9]/1.364$$

where $\Delta G_g = G_g(A^-) + G_g(H^+) - G_g(AH)$ is the sum of the first three terms on the right hand side of Eq. (1) and was taken from the NIST data [57] for the gas phase dissociation of given acid, HA. List of the NIST data is displayed in Table 1S.

The $pK_a$ values were also calculated from Eqs. (2) and (3) derived with the use of the thermodynamic Cycles B and C. To obtain free energy differences: $G_g(A^-) - G_g(AH)$, $G_g(H_2O) - G_g(H_3O^+)$, and $G_g(OH^-) - G_g(H_2O)$, the CBS-QB3 method was used. The values are also displayed in Table 1S.

**Table 3** $\Delta G_s$ calculated using IEFPCM model: experimental $\Delta G_s$

| Radii = UAO $n = 9$ | Neutral molecules | | | | Anions | |
|---|---|---|---|---|---|---|
| | Gas-phase geometry | | Geometry optimized in water | | Gas-phase geometry | Geometry optimized in water |
| | m.u.e. | Mean error | m.u.e. | Mean error | m.u.e. = mean error | m.u.e. = mean error |
| HF/6-31+G* | 2.37 | 2.19 | 2.27 | 2.00 | 11.67 | 10.90 |
| HF/6-311++G** | 2.50 | 2.28 | 2.28 | 2.00 | 11.96 | 11.08 |
| B3LYP/6-31+G* | 2.65 | 2.65 | 2.42 | 2.32 | 14.38 | 13.72 |
| B3LYP/6-311++G** | 2.77 | 2.70 | 2.54 | 2.39 | 14.43 | 13.58 |
| HF/cc-pVDZ | 3.05 | 2.90 | 2.87 | 2.66 | 11.87 | 11.16 |
| HF/cc-pVTZ | 2.89 | 2.81 | 2.70 | 2.56 | 11.37 | 10.58 |
| B3LYP/cc-pVDZ | 3.71 | 3.71 | 3.46 | 3.46 | 14.43 | 13.87 |
| B3LYP/cc-pVTZ | 3.34 | 3.34 | 3.07 | 3.07 | 13.73 | 13.03 |

Gaussian 03

For the calculation of solvation free energies: $\Delta G_s(A^-)$, $\Delta G_s(AH)$, $\Delta G_s(H_2O)$, $\Delta G_s(H_3O^+)$, and $\Delta G_s(OH^-)$, geometry of the species in gas phase was also determined at the HF and B3LYP levels, using Pople 6-31+G* and 6-311++G** basis sets, as well as Dunning cc-pVDZ and cc-pVTZ basis sets. Then the solvation free energies were calculated by applying DPCM, CPCM, and IEFPCM methods, using the same level and basis set which were used for geometry determination in the gas phase. The calculations were made both with and without geometry optimization in water. Namely, calculation of solvation energies employed DPCM/HF(6-31+G*)//HF(6-31+G*), DPCM/HF(6-311++G**)//HF(6-311++G**), DPCM/B3LYP(6-31+G*)//B3LYP(6-31+G*), DPCM / B3LYP (6-311 ++ G**)//B3LYP(6-311++G**), DPCM/HF(cc-pVDZ)//HF(cc-pVDZ), DPCM/HF(cc-pVTZ)//HF(cc-pVTZ), DPCM/B3LYP(cc-PVDZ)//B3LYP(cc-pVDZ), and DPCM/B3LYP(cc-pVTZ)//B3LYP(cc-pVTZ) procedures as well as using similar procedures differing according to the solvation model used, that is CPCM and IEFPCM, instead of DPCM.

In Gaussian 98 program [58] the PCM method of reaction field calculation is by default a synonym of the polarizable dielectric model (DPCM). The method performs normalization of polarization charges in order to arrive at the value of the charge reaching out the cavity [59] predicted by the Gauss' law (Keyword Icomp = $N$). For DPCM and single point calculations, the default is $N = 4$ (the effect of outlying charge is accounted for by means of an additional effective charge, distributed according to the solute electronic density); while for geometry optimization in DPCM the default is $N = 2$ (the calculated charge on each tessera is scaled by a constant factor). For the CPCM calculations (which are less affected by the outlying charge effects) the default value is always Icomp = 2.

During the study, Gaussian 98 program was used for the DPCM calculations, with the Icomp default values and without any special adjustments of atomic radii or other parameters which also take the default values. The parameters are: scaling factors of acidic hydrogens and the parameters used in the construction of the cavity (which is defined as a set of intersecting spheres centered on atoms)—that is atomic radii, the number of the tesserae on each sphere and the areas of the tesserae. For atomic radii, the default setting was Radii = UAHF [55].

Calculations applying the CPCM and IEFPCM methods have been carried out using Gaussian 03 program [60]. In the latter, the SCRF = PCM option has changed its meaning with respect to Gaussian 98 and this input implies performing a reaction field calculation using the integral equation formalism model (IEFPCM). The changes of formalism and implementation versus the previous PCM versions are described in ref. [61]. In this implementation, as well as when using an SCRF = CPCM option (in Gaussian 03) no normalization of polarization charge was employed and the default atomic radii changed to UA0. However, the recommended radii for the calculation of $\Delta G_s$ in the PCM model are those calculated using the UAHF radii [55]. This divergence prompted us to try both possibilities. (There are other optional sets of atomic radii, but they have not been explored). The SCFVAC keyword performing the gas phase calculation before the calculation in a solution was also applied.

## 3 Results and discussion

### 3.1 Solvation free energy terms

The experimental gas-phase free energy differences (for Cycle A) as well as calculated ones used in Cycles B and

C are displayed in Table 1S of the Supporting Information. Tables 2S–7S displaying results of calculation of individual $\Delta G_s$ values for all species under investigation are also available. Tables 2S–4S display the calculated values of $\Delta G_s$ for acids, whereas Tables 5S–7S show those obtained for the conjugated anions. The summary results showing m.u.e. and mean errors of the calculated solvation free energies, according to the calculation level and basis set, are given in Tables 1–3. In the first two tables the results obtained with the Radii = UAHF setting are shown (for DPCM and IEFPCM), while Table 3 presents the results for Radii = UA0 (IEFPCM). The Tables inserted in the body of the text do not display results obtained using the CPCM model, because they are very similar to those from IEFPCM. All results can be found in the Supporting Information. The data presented in the tables, all given in kcal/mol, reveal that errors in determination of $\Delta G_s$ are much greater for the anions than for the corresponding acids. What is more, the mean errors for anions are always positive and frequently equal mean unsigned errors; what follows, in case of all compounds the calculated $\Delta G_s$ values are positively biased versus the experimental ones. It has been previously reported that too small (negative) $\Delta G_s$ values for anions could be largely improved by the outlying charge compensation performed otherwise [59]. The smallest error in the calculated $\Delta G_s(A^-)$ has been found (for the gas-phase geometry) in the DPCM method. For the latter, the best results were obtained for gas-phase geometry and the HF/6-31+G* basis set. After the optimization in the DPCM method, the results with the Pople basis set were poorer than for gas-phase geometry. The result is not unexpected because the DPCM-UAHF model was parametrized by Tomasi and his coworkers [55] at the HF/6-31+G* level for anions and at the HF/6-31G* level for neutral molecules, using gas-phase geometry. They found, however, that geometry reoptimization in the presence of a solvent had a limited effect on the computed solvation free energies and that even large structural modifications lead to negligible changes in energy.

All experimental free energies of solvation used for the evaluation of calculated data employ the standard state 1 mol/L both in gas and in liquid phase. The calculated values of $\Delta G_s$ employ the same standard state in both phases, e.g. 1 atm or 1 mol/L, because definition of the Hamiltonian of the solute in the PCM continuum models does not account for contributions due to thermal motions of molecules. This is equivalent to neglecting the difference between the molecular motion contributions calculated in vacuo and in solution [49].

It is worth following the individual factors influencing errors in the calculated values (in comparison with the experimental data). The factors have been specified as follows:

### *3.1.1 Method used to normalize the polarization charge*

The DPCM calculations were performed using G98 program, which applies different methods in order to normalize the polarization charge so as to obtain the value predicted by the Gauss' law. For single point calculations the default setting is Icomp = 4, whereas for geometry optimization the default setting is Icomp = 2 [44]. In order to see whether the results obtained in DPCM with the gas phase geometry are better than after optimization due to the geometry differences or due to the method of charge normalization, we repeated the DPCM single point calculations forcing Icomp = 2 (in Gaussian 98) and without any charge compensation (using Gaussian 03). In both cases the results were very similar one to the other and for acids they were also similar to the results arrived at when applying the default IComp = 4 setting. However, for anions and consequently for $pK_a$'s, they were much poorer than with the default Icomp = 4 setting for single point calculations (except at the HF/cc-pVTZ and B3LYP/cc-pVTZ levels). On the other hand, free energies of solvation calculated in DPCM with geometry optimization and using the default Icomp = 2 setting were very close to those obtained in DPCM single point with the the same Icomp = 2 setting imposed. This finding is parallel to that quoted in ref. [55], where it was stated that geometry optimization led to negligible changes in solvation energy, whereas the calculations with and without the optimization were performed in the same Icomp = 4 setting. Taking this into consideration, it can be inferred that the manner of the outlying charge compensation is of primary importance for the quality of the solvation free energy results, especially for anions. Data from the DPCM solvation free energies calculated using Icomp = 2 as compared to Icomp = 4 can be found in Tables 2S and 5S, while Tables 8S, 11S, and 14S display the $pK_a$ results for both settings.

The CPCM and IEFPCM calculations were performed using Gaussian 03 where the Icomp keyword is no longer used [45] because the polarization charge density is designed to take into account the outlying charge effects implicitly [61]. It was stated that in the IEFPCM method, the related errors are less pronounced than in the other PCM variant methods [49]. Yet the results obtained in CPCM and IEFPCM (using G03) for anions and $pK_a$'s are poorer than those from DPCM with Icomp = 4 (with the same UAHF radii). Similar findings have been recently reported in ref. [1]. This result might be interpreted as a proof that the errors

related to the outlying charge are still significant. Deviations of $\Delta G_s$ from the experimental data for anions are much greater than those for the neutral molecules and are always positive (mean errors equal mean unsigned errors).

### 3.1.2 Radii = UAHF versus Radii = UA0 (for the CPCM and IEFPCM methods)

The results for $\Delta G_s$ in the CPCM model are displayed in Tables 3S and 6S; those for $\Delta G_s$ in the IEFPCM in Tables 2, 3, 4S, and 7S. It can be seen that for acids, errors in the calculated values of $\Delta G_s$ are greater for the UA0 radii than for the UAHF ones by 2–3 kcal/mole (compare data in Tables 2, and 3 and in Tables 3S, 4S, 6S, and 7S). Deviations of $\Delta G_s$ from the experimental data for anions are much greater for the UA0 radii than for the UAHF ones. One exception is $NH_2^-$, for which the calculated solvation free energies with UAHF radii were extremely high [$\sim -150$ kcal/mol for CPCM and $-150/-260$ kcal/mol for IEFPCM for gas phase geometry, and up to $-780$ kcal/mol for geometry optimized in water (Tables 6S, 7S) whereas the experimental value is $-95$ kcal/mol]. Yet for the UA0 radii, the energies did not depart from the experimental values more than for other anions. That is why $NH_2^-$ was excluded from the calculations of mean errors of $\Delta G_s$ for anions (Tables 6S, 7S) and from the results of $pK_a$ calculations when using the UAHF radii (Tables 9S, 10S, 12S, 13S, 15S, and 16S). It is worth noticing that the solvation energies for $NH_2^-$ calculated with UAHF radii in DPCM (in Gaussian 98) do not differ from the experimental data more than they do in the case of other anions. The solvation free energies were also calculated for $H_3O^+$ and $OH^-$ occurring in Cycles B and C. In the case of $OH^-$, optimization resulted in oscillating $\Delta G_s$ values between $-90$ and $-300$ kcal/mol (UAHF radii, CPCM, and IEFPCM). For the same ion, irrationally low $\Delta G_s$ absolute values ($\Delta G_s = -85$ kcal/mol) were found in some other cases, namely in DPCM (single point) at HF/cc-pVTZ and B3LYP/cc-pVTZ levels, and in CPCM and IEFPCM models with the UA0 radii. For $H_3O^+$, UA0 radii also produced solvation free energies, the absolute values of which were too low by approximately 25 kcal/mol (the experimental values for $\Delta G_s(H_3O^+)$ and $\Delta G_s(OH^-)$ are $-110.3$ and $-104.7$ kcal/mol, respectively [35]). The calculated values of $\Delta G_s(H_3O^+)$ together with the $\Delta G_s(H_2O)$ values are displayed in Tables 11S–13S, concerned with the results obtained with Cycle B, whereas the calculated values of $\Delta G_s$ $(OH^-)$ together with the values of $\Delta G_s(H_2O)$ are displayed in Tables 14S–16S (Cycle C).

### 3.1.3 Optimization of geometry in $H_2O$

As it has been mentioned previously, after the geometry optimization in the DPCM method, the results in G98 achieved with the Pople basis were poorer than for the gas-phase geometry, especially in case of the anions. However, this effect may be attributed to different methods of charge normalization. On the other hand, in the CPCM and IEFPCM models, geometry optimization exerted a negligible influence on the results obtained for neutral molecules, although for the anions, optimization resulted in lower errors, at all levels/basis sets, and with both radii settings.

### 3.1.4 Electron correlation on the DFT (B3LYP) level

For the three SCRF models the results for anions were always poorer when using B3LYP than when HF was applied (mean error of $\Delta G_s$ calculated with B3LYP/6-31+G* and B3LYP/6-311++G** is higher by ca. 2–3 kcal/mol than that of $\Delta G_s$ calculated at the HF/6-31+G* and HF/6-311++G** level. Similarly, mean error calculated with B3LYP/cc-pVDZ or cc-pVTZ is 2–3 kcal/mol higher than with HF/cc-pVDZ or HF/cc-pVTZ). For neutral molecules and Radii = UAHF the trend could be seen for $\Delta G_s$ to be poorer at the B3LYP level than at the HF one, but the differences are fairly small (less than 1 kcal/mol). Similarly, it was stated in ref. [61] that the agreement of the calculated free energies with experimental ones was slightly less satisfactory in DFT calculations (PBE0 functional) and that the trend can be explained by the fact that the cavity size (based on the UAHF radii) has been optimized for Hartree–Fock. However, in the present work, the same trend is even more distinct for UA0 radii.

### 3.1.5 Influence of basis set

In the DPCM model, the use of a more extended basis set (6-311++G** vs. 6-31+G*, cc-pVTZ vs. cc-pVDZ) frequently worsened the results, especially for the anions when the gas phase geometry was applied. For CPCM and IEFPCM, the influence of the completeness of basis set is hardly perceptible.

### 3.1.6 Influence of the SCRF model

A comparison of the results for the CPCM and IEFPCM models (Tables 3S and 6S vs. Tables 4S and 7S) reveals that the results for the two models are almost the same when the same radii setting is applied, in contrast to the far more significant differences found with the two different radii settings used. The results for DPCM depart

from the two other models especially in case of the anions, where the errors in $\Delta G_s$ are significantly lower than for CPCM and IEFPCM in six of the eight cases.

Let us now move on to our main goal, which is the investigation of the influence of the factors mentioned above on the ultimate $pK_a$ values determined by using the calculated values of $\Delta G_s$ for the acids and the anions.

### 3.2 Calculation of $pK_a$ with the experimental $\Delta G_s$ values for $H_2O$ and $H^+$, as well as experimental $\Delta G_g$, using Cycle A

The results of the $pK_a$ determination from Eq. (1) for the individual compounds are presented in the form of Tables 8S–10S in the Supporting Information. As for $\Delta G_s$, the summary results for $pK_a$ are given in Tables 4 and 5. Comparison of Tables 4 and 5 with Tables 1–3 reveals that the accuracy of the $pK_a$ results (i.e. their agreement with experimental data) is mainly governed by the quality of the $\Delta G_s$ results for the anions. Namely, the best result was obtained for DPCM (gas phase geometry, Radii = UAHF) at HF/6-31+G*, and only slightly poorer results were arrived at for HF/6-311++G**, B3LYP/6-31+G*, and HF/cc-pVDZ. Geometry optimization in water affected the results negatively in six cases out of eight. Yet, as it has been previously mentioned, the said effect may be attributed to the application of different default methods of polarization charge normalization. Only for HF/cc-pVTZ and B3LYP/cc-pVTZ, the results after optimization were better.

The influence of a basis set on the accuracy of $pK_a$'s was parallel to that recorded for $\Delta G_s$ values. In the DPCM model, application of a more extended basis set (6-311++G** vs. 6-31+G*, cc-pVTZ vs. cc-pVDZ) frequently made the results worse, when the gas phase geometry was applied. Nevertheless, for DPCM with geometry optimization, CPCM and IEFPCM, completeness of a basis set exerted a hardly noticeable influence.

For the CPCM and IEFPCM models, geometry optimization resulted in the improvement of the $pK_a$ values. The corresponding $pK_a$'s calculated using both settings of atomic radii are displayed in Tables 9S, 10S, and in Table 5. It can be seen that the results with UAHF radii were significantly better than those for the UA0 radii ($K_a$ better by ~3 orders of magnitude). However, for $NH_3$, the $pK_a$ values have not been calculated with Radii = UAHF because of $\Delta G_s$ values for $NH_2^-$ being exceedingly high. Yet application of the UA0 atomic radii resulted in obtaining $pK_a$ values for $NH_3$ which proved as good as these for other compounds. The error of $pK_a$ calculated using CPCM with gas-phase geometry and UA0 radii is about 4–5 units; for the water-optimized geometry it is about seven units. We can therefore state that the UA0 atomic radii and the cavities built up with the latter are useless for the determination of $pK_a$ because of the far too small (negative) values of $\Delta G_s$ for most of the anions. The results with the two radii sets differ to a far greater extent one from the other than they differ in the two SCRF models or when different levels/basis sets are used.

At gas phase geometry, UAHF radii, and Pople basis sets, the order of consistency of the results with experimental data is DPCM>CPCM≥IEFPCM. This order is valid only for the default IComp = 4 setting in G98 for the DPCM calculations. In G03 with no charge compensation, and in G98 with IComp = 2 setting, the result is CPCM ≈ IEFPCM ≥ DPCM. Thus the advantage of the single point DPCM method seems to lie in the use of the Icomp = 4 method, the default one in G98.

The influence of electron correlation at the investigated DFT (B3LYP) level is such that the B3LYP results are always poorer than the HF results obtained using the same basis set, with both UAHF and UA0 atomic radii.

All these inferences also hold true for Cycles B and C in case of application, for the two latter ones, of consistent experimental solvation free energies for $H_2O$, $H_3O^+$, and $OH^-$. In such a case the three cycles generate the same results, as it has been shown in the Introduction. The situation changes the moment when the theoretical values are used for all terms in the Cycles B and C due to the impact of the accuracy of calculated $\Delta G_s(H_3O^+)$ and $\Delta G_s(OH^-)$ terms.

### 3.3 Calculation of $pK_a$ with the calculated values of $\Delta G_s$ for $H_2O$, $H_3O^+$ and $OH^-$, as well as for $\Delta G_g$

#### *3.3.1 Cycle B*

The solvation free energies for $H_2O$ and $H_3O^+$ are displayed in the Excel sheets presenting the $pK_a$ calculations according to Cycle B (Tables 11S–13S).

In this case, errors for DPCM (gas phase geometry and geometry optimized in water), Table 6, and CPCM and IEFPCM (gas phase geometry, Radii = UAHF), Table 7, are far more substantial than those calculated using Cycle A. This may mainly be attributed to the calculated values of $\Delta G_s(H_3O^+)$ which are too low in relation to the experimental value by 2–5 kcal/mol (Table 11S). This fact has already been noticed by Pliego [23] (calculation at IEFPCM/HF/6-31G* level). Hence he deduced that the IEFPCM–UAHF solvation model was inconsistently parametrized. Now it appeared that the effect has also been found for other models and levels, e.g. mean value of $\Delta G_s(H_3O^+)$ for eight level/basis set combinations is −105.5. kcal/mol for DPCM (gas

**Table 4** $pK_a$ calculated using DPCM model: experimental $pK_a$—Cycle A

| Radii = UAHF, $n = 9$ | Gas-phase geometry Icomp = 4 (default) | | Geometry optimized in water, Icomp = 2 (default) | | Gas-phase geometry Icomp = 2 | |
|---|---|---|---|---|---|---|
| | m.u.e. | Mean error | m.u.e. | Mean error | m.u.e. | Mean error |
| HF/6-31+G* | 2.37 | 1.03 | 4.33 | 3.47 | 4.43 | 3.56 |
| HF/6-311++G** | 3.06 | 1.88 | 4.34 | 3.56 | 4.43 | 3.66 |
| B3LYP/6-31+G* | 2.94 | 2.94 | 6.07 | 5.56 | 6.15 | 5.64 |
| B3LYP/6-311++G** | 4.79 | 3.99 | 6.21 | 5.78 | 6.27 | 5.84 |
| HF/cc-pVDZ | 3.04 | 0.87 | 3.54 | 2.41 | 3.69 | 2.56 |
| HF/cc-pVTZ | 5.28 | 3.79 | 4.11 | 2.91 | 4.22 | 3.02 |
| B3LYP/cc-pVDZ | 3.68 | 2.78 | 5.20 | 4.27 | 5.33 | 4.41 |
| B3LYP/cc-pVTZ | 6.23 | 4.98 | 4.91 | 3.81 | 5.08 | 3.98 |

Gaussian 98

**Table 5** $pK_a$ calculated using IEFPCM model: experimental $pK_a$—Cycle A

| | Radii = UAHF, $n = 8$ | | | | Radii = UAO, $n = 9$ | |
|---|---|---|---|---|---|---|
| | Gas-phase geometry | | Geometry optimized in water | | Gas-phase geometry | Geometry optimized in water |
| | m.u.e. | Mean error | m.u.e. | Mean error | m.u.e. = mean error | m.u.e. = mean error |
| HF/6-31+G* | 3.93 | 3.93 | 3.05 | 3.05 | 6.79 | 6.28 |
| HF/6-311++G** | 3.98 | 3.98 | 3.19 | 3.19 | 6.94 | 6.41 |
| B3LYP/6-31+G* | 5.76 | 5.76 | 5.14 | 5.14 | 8.44 | 8.11 |
| B3LYP/6-311++G** | 5.88 | 5.88 | 5.04 | 5.04 | 8.44 | 7.96 |
| HF/cc-pVDZ | 3.49 | 3.45 | 2.88 | 2.67 | 6.41 | 5.98 |
| HF/cc-pVTZ | 3.46 | 3.46 | 2.80 | 2.69 | 6.11 | 5.63 |
| B3LYP/cc-pVDZ | 5.20 | 5.20 | 4.65 | 4.65 | 7.70 | 7.38 |
| B3LYP/cc-pVTZ | 5.19 | 5.19 | 4.31 | 4.31 | 7.46 | 7.05 |

Gaussian 03

phase geometry) as compared to the experimental value of −110.3 kcal/mol. However, for CPCM/UAHF and IEFPCM/UAHF with geometry optimization in water, values for $\Delta G_s(H_3O^+)$ were closer to the experimental value (av. −111.0 and −110.8 kcal/mol) and at the same time the results for the calculated $pK_a$'s were more accurate and comparable with those obtained using the same SCRF models and Cycle A (slightly poorer at HF level but better at B3LYP).

As concerns Radii = UA0, the calculations of $\Delta G_s(H_3O^+)$ were underestimated by 25 kcal/mol, which resulted in the $pK_a$'s error of approximately 25 $pK_a$ units. The same results for $\Delta G_s(H_3O^+)$ by using the UA0 radii were reported recently [25].

### *3.3.2 Cycle C*

Radii = UAHF. The solvation free energies for $OH^-$ and $H_2O$ are displayed in the Excel sheets presenting the $pK_a$ calculations according to Cycle C (Tables 14S–16S). The summary $pK_a$ results are displayed in Tables 8 and 9.

In this calculation variant, the best results were obtained for CPCM and IEFPCM at gas phase geometry and for Pople's basis sets, due to a partial compensation of errors in $\Delta G_s(A^-)$, $\Delta G_s(OH^-)$, and $\Delta G_s(H_2O)$. For example, taking the best result: m.u.e. and mean error 1.29, at B3LYP/6-311++G** level (Table 9), the sum $\Delta G_s(H_2O) - \Delta G_s(OH^-)$ equals 94.0 kcal/mol, which differs from the experimental values of 98.4 kcal/mol by 4.4 kcal/mol. This figure partially compensates for the errors resulting from the too low absolute values of $\Delta G_s(A^-)$ [Eq. (3)]. What is more, the calculated value of $G_g(H_2O) - G_g(OH^-)$, occurring in Eq. 3, is more negative than the experimental value. They equal −391.7 and −390 kcal/mol, respectively. (The results obtained for HF / 6-311 ++G**, B3LYP/6-31+G*, and B3LYP/6-311++G** are even better than those obtained by using Cycle A, DPCM, Icomp = 4, at these three levels). For Dunning basis sets the results of all methods are poor, especially in DPCM, gas-phase geometry, combined with cc-pVTZ basis set, where $\Delta G_s$ values of $OH^-$ were exceedingly low (−89 kcal/mol for HF and −86 kcal/mol for B3LYP).

**Table 6** $pK_a$ calculated using DPCM model: experimental $pK_a$—Cycle B

| Radii = UAHF, $n = 9$ | Gas-phase geometry Icomp = 4 (default) | Geometry optimized in water, Icomp = 2 (default) | | Gas-phase geometry Icomp = 2 | |
|---|---|---|---|---|---|
| | m.u.e. = mean error | m.u.e. | Mean error | m.u.e. | Mean error |
| HF/6-31+G* | 5.44 | 6.14 | 6.14 | 7.38 | 7.38 |
| HF/6-311++G** | 7.16 | 6.99 | 6.99 | 8.26 | 8.26 |
| B3LYP/6-31+G* | 6.72 | 7.15 | 7.15 | 8.72 | 8.72 |
| B3LYP/6-311++G** | 8.18 | 7.44 | 7.44 | 9.01 | 9.01 |
| HF/cc-pVDZ | 4.93 | 4.71 | 4.35 | 5.72 | 5.61 |
| HF/cc-pVTZ | 8.96 | 5.75 | 5.45 | 6.67 | 6.60 |
| B3LYP/cc-pVDZ | 6.16 | 5.44 | 5.01 | 6.67 | 6.56 |
| B3LYP/cc-pVTZ | 9.60 | 5.64 | 5.18 | 6.85 | 6.69 |

Gaussian 98

**Table 7** $pK_a$ calculated using IEFPCM model: experimental $pK_a$—Cycle B

| | Radii = UAHF, $n = 8$ | | | Radii = UAO, $n = 9$ | |
|---|---|---|---|---|---|
| | Gas-phase geometry | Geometry optimized in water | | Gas-phase geometry | Geometry optimized in water |
| | m.u.e. = mean error | m.u.e. | Mean error | m.u.e. = mean error | m.u.e. = mean error |
| HF/6-31+G* | 7.27 | 3.66 | 3.66 | 25.34 | 24.06 |
| HF/6-311++G** | 7.83 | 3.95 | 3.95 | 25.51 | 24.22 |
| B3LYP/6-31+G* | 8.72 | 4.04 | 4.04 | 26.67 | 25.44 |
| B3LYP/6-311++G** | 8.88 | 4.05 | 4.05 | 26.69 | 25.31 |
| HF/cc-pVDZ | 6.62 | 3.15 | 2.81 | 23.98 | 22.81 |
| HF/cc-pVTZ | 6.94 | 3.46 | 3.32 | 24.13 | 22.98 |
| B3LYP/cc-pVDZ | 7.50 | 3.20 | 3.09 | 24.68 | 23.48 |
| B3LYP/cc-pVTZ | 7.73 | 3.23 | 3.23 | 24.99 | 23.79 |

Gaussian 03

**Table 8** $pK_a$ calculated using DPCM model: experimental $pK_a$—Cycle C

| Radii = UAHF, $n = 9$ | Gas-phase geometry Icomp = 4 (default) | | Geometry optimized in water, Icomp = 2 (default) | Gas-phase geometry Icomp = 2 |
|---|---|---|---|---|
| | m.u.e. | Mean error | m.u.e. = mean error | m.u.e. = mean error |
| HF/6-31+G* | 2.86 | 1.74 | 7.01 | 6.80 |
| HF/6-311++G** | 3.01 | 1.51 | 5.83 | 5.63 |
| B3LYP/6-31+G* | 2.47 | 0.99 | 7.50 | 7.17 |
| B3LYP/6-311++G** | 3.06 | 1.64 | 7.16 | 6.86 |
| HF/cc-pVDZ | 4.43 | 4.43 | 8.40 | 8.46 |
| HF/cc-pVTZ | 7.87 | −8.89 | 8.62 | 8.57 |
| B3LYP/cc-pVDZ | 5.28 | 5.28 | 9.89 | 9.90 |
| B3LYP/cc-pVTZ | 9.10 | −9.99 | 9.31 | 9.27 |

Gaussian 98

**Table 9** $pK_a$ calculated using IEFPCM model: experimental $pK_a$—Cycle C

| | Radii = UAHF, $n = 8$ Gas-phase geometry | Radii = UAO, $n = 9$ Gas-phase geometry | | Geometry optimized in water | |
|---|---|---|---|---|---|
| | m.u.e. = mean error | m.u.e. | Mean error | m.u.e. | Mean error |
| HF/6-31+G* | 2.42 | 8.07 | −8.07 | 8.77 | −8.77 |
| HF/6-311++G** | 1.96 | 8.05 | −8.05 | 8.74 | −8.74 |
| B3LYP/6-31+G* | 1.44 | 8.20 | −8.20 | 8.72 | −8.72 |
| B3LYP/6-311++G** | 1.29 | 8.19 | −8.19 | 8.84 | −8.84 |
| HF/cc-pVDZ | 6.64 | 6.78 | −6.20 | 6.98 | −6.75 |
| HF/cc-pVTZ | 5.77 | 6.55 | −6.51 | 7.13 | −7.13 |
| B3LYP/cc-pVDZ | 7.70 | 6.46 | −5.41 | 6.58 | −5.79 |
| B3LYP/cc-pVTZ | 6.99 | 6.41 | −5.64 | 6.58 | −6.16 |

Gaussian 03

However, for CPCM and IEFPCM the calculation could not be performed after geometry optimization in water because for $OH^-$ the optimization resulted in $\Delta G_s(OH^-)$ values amounting to −300 kcal/mol.

Radii = UA0. In this case the results of CPCM and IEFPCM make no sense whatsoever, the $\Delta G_s$ absolute values for $OH^-$ are about 20 kcal/mol lower, than the experimental ones (Table 9 and Tables 15S, 16S). Here the $pK_a$ values are not as high as for Cycle B using UA0 Radii (Table 7) because the strongly biased values of $\Delta G_s(OH^-)$ and $\Delta G_s(A^-)$ are subtracted [Eq. (3)].

## 4 Conclusions

The results of $pK_a$ calculations are given for the thermodynamic Cycle A, with the combined experimental and calculated free energy terms, and for the Cycles B and C, with only calculated ones. In all cases the $pK_a$'s are positive, meaning that there is a marked tendency in the calculated values to be higher than the experimental ones. This effect is attributable mainly to the exceedingly low (negative) $\Delta G_s$ as calculated for anions. It turned out that the Radii = UA0 setting, which is a default one in the CPCM and IEFPCM calculation in Gaussian 03, produced far poorer results than the Radii = UAHF setting did (which, however, is recommended for the calculation of $\Delta G_s$ together with the keyword SCFVAC performing the gas phase calculation prior to the calculation in a solution [45]). Although the UAHF method for assigning atomic radii was optimized using DPCM, the UAHF radii set also works fairly well when it is applied together with the more recent versions of PCM. Yet, using the CPCM and IEFPCM methods with the UAHF radii produced irrationally large (negative) $\Delta G_s$ for the small anions, i.e. $NH_2^-$ and $OH^-$, which are supposed to generate strong electric field near their cavities.

When using Cycle A, the application of DPCM method implemented in Gaussian 98 produced the best results for anion solvation free energies (at the HF level) and consequently for $pK_a$'s. It was true given two conditions were met. The first one involved calculation of free energies of solvation for gas-phase geometry performed using Icomp = 4 default setting (The effect of the outlying charge is accounted for by means of an additional effective charge, distributed according to the solute electronic density [44]). With the forced Icomp = 2 setting (The calculated charge on each tessera is scaled by a constant factor), the results were much poorer. The second condition relates to the level/basis set used. To be specific, the best results have been obtained at HF/6-31+G*, B3LYP/6-31+G*, HF/6-311++G**, and HF/cc-pVDZ levels.

The conclusion that follows from the above-described results is that the manner in which the solute polarization and the outlying charges are treated is more important than the treatment of the solvent (as a dielectric or as a polarizable conductor). In other words, the impact of the boundary conditions (such as solute cavity/bulk of the solvent) surpasses that of the separate solute (geometry optimization) or solvent properties. It seems to us that the very noticeable influence of the atomic radii employed for the construction of cavity (UAHF vs. UA0 ones) may also be explicable by the impact of the cavity shape on the boundary phenomena.

Regardless of the applied model, the results deteriorated when more polarization and/or diffuse functions were added to the basis set used. These additional functions allow for displacement of electron distributions away from the nuclear positions, e.g. larger spatial extension of the wave function.

It appeared that for the four basis sets tried (two of them being Pople and two Dunning ones), the results at the B3LYP level were always poorer than the HF results arrived at using the same basis set.

As concerns Cycle B, the results are better than those obtained using Cycle A only for CPCM and IEFPCM models (Radii = UAHF, at B3LYP level, after geometry optimization). This is due to slightly more negative $\Delta G_s(H_3O^+)$ values (calculated at B3LYP) than the experimental value equaling −110.3 kcal/mol (Tables 12S, 13S). This entails lowering of the $pK_a^B$ values [Eq. (2)]. Application of Cycle C also resulted, in some cases (UAHF radii, Pople basis sets, gas phase geometry), in better $pK_a$'s than those from Cycle A. This improvement is due to the differences in the calculated (used here in Cycle C) and experimental (used in Cycle A) values of $G_g(H_2O) - G_g(OH^-)$, as well as $\Delta G_s(H_2O) - \Delta G_s(OH^-)$. Other results generated in Cycles B and C were poorer than those obtained by using Cycle A.

We thus concluded that of the three simple cycles used here, application of Cycle A is the most practical and recommendable for two reasons. The first reason is the uncertainty of the calculated $\Delta G_s$ values for small ions $OH^-$ and $H_3O^+$, occurring in cycles B and C. For example, by using UA0 radii, one obtains for $OH^-$ and $H_3O^+$ the $\Delta G_s$ absolute values about 25 kcal/mol too low. Such effect has also been reported for other radii sets [25]. What follows, application of the two cycles entailed very large errors in $pK_a$ values. Quality of the best results is attributable to compensation of errors in some terms. The second reason is that all three cycles generate the same $pK_a$ values when experimental values for $G_s(H_3O^+)$, $\Delta G_s(OH^-)$, $\Delta G_s(H_2O)$, $G_g(H_2O) - G_g(H_3O^+)$, and $G_g(OH^-) -$

$G_g(H_2O)$ are used in Eqs. (1), (2), and (3) (as it has been derived in the Introduction).

However, there is another possibility of $pK_a$ calculation, although it has not been investigated here. It consists in the application of cycles similar to those considered here in which one or more explicit water molecules are added to small ions [1,24,35,62,63] and some larger ones [1,34,35,62,64]. (In ref. [1] only to those bearing large partial negative atomic charges). This alternative is a very attractive one and in some cases it indeed produced better results than the three cycles taken into consideration presently. Nevertheless, so far it has been explored in a very limited scope.

We are of the opinion that future investigation will explore this option further. It would be reasonable to assume that the clusters had the structure similar to the one really existing in water solution, and this is the main difficulty because the genuine structure of solvated ions and that of water itself is far from being fully recognized. In practice, the most frequently encountered number of water molecules is one to four [1,24,35,62–64].

Results obtained with the simple Cycle A, preferred in this investigation, depend on the accepted experimental free energy of proton solvation. The recent value of Tissandier et al. [39] corrected to 1M–1M standard state (−265.9 kcal/mol), considered timely as the most accurate one and accepted here, have been recently reproduced to within 0.1 [42] and 0.2 kcal/mol [35]. Yet it may be superseded by better measurements in the future. In this case, the $pK_a$'s resulting from Cycle A can be simply shifted by adding a constant correction value. In order to see whether our conclusions are robust towards a change in the $\Delta G_s(H^+)$ value, we repeated the calculations with the theoretical value of Zhan and Dixon quoted as −262.4 kcal/mol [65]. Despite the fact that both the results of calculations of $\Delta G_s$ for anions (as compared to the experimental ones) and of the $pK_a$'s worsened, all of the conclusions were still valid.

**Acknowledgments** We should like to express our gratitude to Professor Christopher Cramer and to Casey Kelly for kindly providing us with their yet unpublished data on solvation energies of a number of compounds (now published in ref. [35]). We are also grateful to the Interdisciplinary Centre for Mathematical and Computational Modeling of Warsaw University for allowing the authors to use their computational resources. This work was supported in part by a grant from the State Scientific Research Committee.

## References

1. Kelly CP, Cramer CJ, Truhlar DG (2006) J Phys Chem A 110:2493
2. Chipman DM (2002) J Phys Chem A 106:7413
3. Schüürmann G, Cossi M, Barone V, Tomasi J (1998) J Phys Chem A 102:6706
4. Gao D, Svoronos P, Wong PK, Maddalena D, Hwang J, Walker H (2005) J Phys Chem A 109:10776
5. Liptak MD, Shields GC (2001) Int J Quantum Chem 85:727
6. De Abreu HA, De Almeida WB, Duarte HA. (2004) Chem Phys Lett 383:47
7. Klamt A, Eckert F, Diedenhofen M, Beck ME (2003) J Phys Chem A 107:9380
8. da Silva CO, da Silva EC, Nascimento MAC (1999) J Phys Chem A 103:11194
9. da Silva CO, da Silva EC, Nascimento MAC (2000) J Phys Chem A 104:2402
10. Topol IA, Tawa GJ, Burt SK, Rashin AA (1997) J Phys Chem A 101:10075
11. Lee I, Kim CK, Han IS, Lee HW, Kim WK, Kim JB (1999) J Phys Chem B 103:7302
12. Lim C, Bashford D, Karplus M (1991) J Phys Chem 95: 5610
13. Liptak MD, Shields GC (2001) J Am Chem Soc 123:7314
14. Liptak MD, Gross KC, Seybold PG, Feldgus S, Shields GC (2002) J Am Chem Soc 124:6421
15. Jang YH, Sowers LC, Çağin T, Goddard WA III (2001) J Phys Chem A 105:274
16. Klicìc JJ, Friesner RA, Liu S-Y, Guida WC (2002) J Phys Chem A 106:1327
17. Murłowska K, Sadlej-Sosnowska N (2005) J Phys Chem A 109:5590
18. Satchell JF, Smith BJ (2002) Phys Chem Chem Phys 4:4314
19. Jang YH, Goddard WA III, Noyes KT, Sowers LC, Hwang S, Chung DS (2003) J Phys Chem B 107:344
20. Chen IJ, MacKerell AD Jr (2000) Theor Chem Acc 103:483
21. Tuñón I, Silla E, Tomasi J (1992) J Phys Chem 96:9043
22. Kallies B, Mitzner R (1997) J Phys Chem B 101:2957
23. Pliego JR Jr (2003) Chem Phys Lett 367:145
24. Pliego JR Jr, Riveros JM (2002) J Phys Chem A 106:7434
25. Takano Y, Houk KN (2005) J Chem Theory Comput 1:70
26. Cao Z, Lin M, Zhang Q, Mo Y (2004) J Phys Chem A 108:4277
27. Popovìc DM, Quenneville J, Stuchebrukhov AA (2005) J Phys Chem B 109:3616
28. Jensen JH, Li H, Robertson AD, Molina PA (2005) J Phys Chem A 109:6634
29. Namazian M, Heidary H (2003) J Mol Struct (Theochem) 620:257
30. Cramer CJ (2003) Essentials of computational chemistry. Wiley, England
31. Pearson RG (1986) J Am Chem Soc 108:6109
32. Pliego JR Jr, Riveros JM (2002) Phys Chem Chem Phys 4:1622
33. Pliego JR Jr, Riveros JM (2000) Chem Phys Lett 332:597
34. Kelly CP, Cramer CJ, Truhlar DG (2005) J Chem Theory Comput 1:1133
35. Kelly CP, Cramer CJ, Truhlar DG (2006) J Phys Chem B 110:16066
36. Montgomery JA, Frisch MJ, Ochterski JW, Petersson GA (1999) J Chem Phys 110:2822
37. Ochterski JW, Petersson GA, Montgomery JA (1996) J Chem Phys 104:2598
38. Ben-Naim A, Marcus Y (1984) J Chem Phys 81:2016
39. Tissandier MD, Cowen KA, Feng WY, Gundlach E, Cohen MH, Earhart AD, Coe JV, Tuttle TR (1998) J Phys Chem A 102:7787
40. Tuttle TR, Malaxos S, Coe JV (2002) J Phys Chem A 106:925
41. Camaioni DM, Schwerdtfeger CA (2005) J Phys Chem A 109:10795
42. Bartels DM, Takahashi K, Cline JA, Marin TW, Jonah CD (2005) J Phys Chem A 109:1299

43. Foresman JB, Frisch Æ (1996) Exploring chemistry with electronic structure methods. Gaussian Inc., Pittsburgh
44. Frisch Æ, Frisch MJ (1999) Gaussian 98 user's reference, 2nd Edn. Gaussian Inc., Pittsburgh
45. Frisch Æ, Frisch MJ, Trucks GW (2003) Gaussian 03 user's reference. Gaussian Inc., Carnegie
46. Miertuš S, Scrocco E, Tomasi J (1981) J Chem Phys 55:117
47. Cossi M, Barone V, Cammi R, Tomasi J (1996) Chem Phys Lett 255:327
48. Barone V, Cossi M (1998) J Phys Chem 102:1995
49. Mennucci B, Cancès E, Tomasi J (1997) J Phys Chem B 101:10506
50. Cancès E, Mennucci B, Tomasi J (1997) J Chem Phys 107:3032
51. Tomasi J, Mennucci B, Cammi R (2005) Chem Rev 105:2999
52. Cramer CJ, Truhlar DG (1999) Chem Rev 99:2161
53. Orozco M, Luque FJ (2000) Chem Rev 100:4187
54. Luque FJ, Curutchet C, Muñoz-Muriedas J, Bidon-Chanal A, Soteras I, Morreale A, Gelpi JL, Orozco M (2003) Phys Chem Chem Phys 5:3827
55. Barone V, Cossi M, Tomasi J (1997) J Chem Phys 107:3210
56. Rappé AK, Casewit CJ, Colwell KS, Goddard WA III, Skiff WM (1992) J Am Chem Soc 114:10024
57. NIST Chemistry Webbook. NIST Standard Reference Database Number 69, June 2005 Release. (http://webbook.nist.gov)
58. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Zakrzewski VG, Montgomery JA, Stratmann RE, Burant JC, Dapprich S, Millam JM, Daniels AD, Kudin KN, Strain MC, Farkas O, Tomasi J, Barone V, Cossi M, Cammi R, Mennucci B, Pomelli C, Adamo C, Clifford S, Ochterski J, Petersson GA, Ayala PY, Cui Q, Morokuma K, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Cioslowski J, Ortiz JV, Baboul AG, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Gomperts R, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Gonzales C, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Andres JL, Head-Gordon M, Replogle ES, Pople JA (1999) Gaussian 98, revision A.11.4. Gaussian, Pittsburgh
59. Klamt A, Jonas V (1996) J Chem Phys 105:9972
60. Frisch M J, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA, Vreven T, Kudin KN, Burant JC, Millam JM, Iyengar SS, Tomasi J, Barone V, Mennucci B, Cossi M, Scalmani G, Rega N, Petersson GA, Nakatsuji H, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, NakajimaT, Honda Y, Kitao O, Nakai H, Klene M, Li X, Knox JE, Hratchian HP, Cross JB, Adamo C, Jaramillo J, Gomperts R, Stratmann RE, Yazyev O, Austin AJ, Cammi R, Pomelli C, Ochterski JW, Ayala PY, Morokuma K, Voth GA, Salvador P, Dannenberg JJ, Zakrzewski VG, Dapprich S, Daniels AD, Strain MC, Farkas O, Malick DK, Rabuck AD, Raghavachari K, Foresman JB, Ortiz JV, Cui Q, Baboul AG, Clifford S, Cioslowski J, Stefanov BB, Liu G, Liashenko A, Piskorz P, Komaromi I, Martin RL, Fox DJ, Keith T, Al-Laham MA, Peng CY, Nanayakkara A, Challacombe M, Gill PMW, Johnson B, Chen W, Wong MW, Gonzalez C, Pople JA (2003) Gaussian 03, revision B.04 Gaussian, Pittsburgh
61. Cossi M, Scalmani G, Rega N, Barone V (2002) J Chem Phys 117:43
62. Pliego JR Jr, Riveros JM (2001) J Phys Chem A 105:7241
63. Asthagiri D, Pratt LR, Asbaugh HS (2003) J Chem Phys 119:2702
64. Adam KR (2002) J Phys Chem A 106:11963
65. Zhan C-G, Dixon DA (2001) J Phys Chem A 105:11534