JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

# Searching for Computational Strategies to Accurately Predict p$K_a$s of Large Phenolic Derivatives

Aida Mariana Rebollar-Zepeda, Tania Campos-Hernández, María Teresa Ramírez-Silva, Alberto Rojas-Hernández, and Annia Galano*

Departamento de Química, División de Ciencias Básicas e Ingeniería, Universidad Autónoma Metropolitana-Iztapalapa, Av San Rafael Atlixco No.186, Col.Vicentina C.P.09340, México D.F.

Supporting Information

**ABSTRACT:** Twenty-two reaction schemes have been tested, within the cluster-continuum model including up to seven explicit water molecules. They have been used in conjunction with nine different methods, within the density functional theory and with second-order Møller−Plesset. The quality of the p$K_a$ predictions was found to be strongly dependent on the chosen scheme, while only moderately influenced by the method of calculation. We recommend the **E1** reaction scheme [HA + $OH^-$ ($3H_2O$) ↔ $A^-$ ($H_2O$) + $3H_2O$], since it yields mean unsigned errors (MUE) lower than 1 unit of p$K_a$ for most of the tested functionals. The best p$K_a$ values obtained from this reaction scheme are those involving calculations with PBE0 (MUE = 0.77), TPSS (MUE = 0.82), BHandHLYP (MUE = 0.82), and B3LYP (MUE = 0.86) functionals. This scheme has the additional advantage, compared to the proton exchange method, which also gives very small values of MUE, of being experiment independent. It should be kept in mind, however, that these recommendations are valid within the cluster-continuum model, using the polarizable continuum model in conjunction with the united atom Hartree−Fock cavity and the strategy based on thermodynamic cycles. Changes in any of these aspects of the used methodology may lead to different outcomes.

## INTRODUCTION

There is a huge amount of substances that behave as Brønsted acids or bases in aqueous solution. Proton-transfer equilibriums are therefore very important for a large variety of chemical compounds and in particular for pharmaceuticals, which frequently are weak acids or bases.[1] Acid dissociation constants ($K_a$) not only characterize the acidity of these compounds but also influence their reactivity. They are commonly reported as p$K_a$s, and their values are related to numerous properties of drugs, such as solubility and rate of absorption.[2] They are also taken into account to decide dosage forms and regimes of drugs.[3] Therefore the accurate knowledge of p$K_a$s is highly important for practical purposes as well as for understanding the behavior of chemicals under different conditions.

There are several experimental techniques that have been successfully applied to accurately determine p$K_a$s. However sometimes this becomes a challenging task,[4] for example, for short-living intermediates and for very weak or very strong acids.[5] Thus a large amount of works have been devoted to obtain p$K_a$ values using theoretical methods.[2,5−26] Different calculation strategies and current trends have been recently and thoroughly reviewed by Ho and Coote.[27]

Considering all the information gathered so far it becomes evident that estimating accurate p$K_a$s using computational methodologies remains a very complicated problem. The difficulties are numerous. To start with, deprotonation processes do not conserve the number of charged species on both sides of the equilibrium. Therefore there is no cancellation of errors when computing energies of reaction. As a result, the accuracy of these relative energies is not as predictable as for processes where the number of charged species is conserved. A second problem is that even though high-level composite methods, such as Gaussian-n (Gn)[28] and complete basis set (CBS)[29] variations, are now available for producing accurate gas phase energies, the calculations in solution have not reached such level of accuracy yet. Moreover, Gn and CBS methods are still computationally unfeasible for moderately large systems, which is the case of most chemicals. An additional problem, particularly important for calculating p$K_a$s in aqueous solution, is that specific short-range solute−solvent interactions are not included when continuum models are used. On the other hand, using discrete models, including a large enough number of solvent molecules, within ab initio or density functional theory (DFT) frameworks, would be so computationally demanding that it is currently an unattainable option. If we take into account that all the above-mentioned problems represent sources of error, that do not necessarily cancel out, and that an error of 1.36 kcal/mol in the Gibbs free energy of the deprotonation reactions ($\Delta G$) represents an error of 1 p$K_a$ unit, then it becomes evident that producing accurate p$K_a$ values from theoretical calculations is a very challenging task. In fact it is currently accepted that mean absolute deviations smaller than 2 units of p$K_a$ are reasonably accurate.[27]

Albeit a wide variety of strategies that have been developed to overcome such difficulties, in this work we will focus on those involving thermodynamic cycles. This approach allows a large flexibility in designing reaction schemes that maximize systematic error cancelations.[27] Some of them can be designed in such a way

**Received:** March 17, 2011
**Published:** June 29, 2011

ACS Publications © 2011 American Chemical Society

that the number of charged species is conserved on both sides of the deprotonation equilibrium. Thermodynamic cycles also allow obtaining Gibbs free energies in solution $\Delta G_s$ from Gibbs free energies in gas phase ($\Delta G_g$) and Gibbs free energies of solvation ($\Delta G_{solv}$). This seems to be an efficient strategy since continuum solvation models are parametrized to produce accurate $\Delta G_{solv}$, but the levels of theory at which they are usually implemented are not sufficient to accurately reproduce $\Delta G_s$.[27] Therefore it is possible to take advantage of using higher levels of theory to calculate $\Delta G_g$ and therefore to improve the accuracy of the results. However, as mentioned before, bulk polarization effects are not necessarily enough to accurately reproduce $\Delta G_{solv}$ because of the short-range solute−solvent interactions. One way to overcome this issue, without making the calculations computationally prohibited, is to include a few solvent molecules in close proximity to the solute, in addition to using a dielectric continuum model. This hybrid explicit/implicit solvation model is known as the discrete-continuum,[30] the supermolecule-reaction field,[31] and the cluster-continuum model.[32]

Taking into account all the possible levels of theory, solvation methods, reaction schemes and number of explicit solvent molecules that can be used for calculating p$K_a$s with thermodynamic cycles, the amount of variations becomes almost infinite. Therefore in this work we will focus on levels of theory that are computationally feasible for relatively large molecules, mainly within the DFT framework. They are used in combination with the polarizable continuum model (PCM), which is currently one of the most widely used for calculating p$K_a$s in aqueous solution. Twenty-two reaction schemes have been used, some based on those previously reported by other authors and some proposed for the first time. In addition the cluster-continuum model has been used, including up to seven explicit water molecules. To test the resulting strategies we have chosen four nonsteroidal anti-inflammatory drugs, which are phenolic derivatives: acetaminophen, profadol, tapentadol, and ketobemidone. Then the study has been extended to a larger series of phenols.

Phenolic compounds are the focus of our study since they are ubiquitous and versatile substances. They have been identified to have multiple biological activities, including antioxidant[33] and cardioprotective[34] effects as well as anti-inflammatory,[35] antimicrobial, and antiviral[36] activities. They are also used to prevent and treat cancer[37] and neurodegenerative diseases[38] and to prevent skin damage[39] and osteoporosis.[40] Phenols are even proposed to have inhibitory effects against obesity.[41] All these biological activities take place within the human body, i.e., at different pHs depending on the specific site of action. Since their effects vary depending of their p$K_a$, which rules the fractions of their acid/base forms under specific conditions, it is vital to know this data.[33f] In addition, from a computational point of view, phenols represent an extra challenge for estimating acid constants, due to the influence of solvation in the proper description of the phenoxide anions' geometry.[10] Even though highly accurate strategies have been proposed for calculating the pKas of phenols, they involve CBS-QB3 calculations,[10] which are computationally unfeasible for most phenolic compounds with biological activity.

Based on the systematic analysis of the extensive data obtained, a few strategies are recommended for accurate calculations of aqueous p$K_a$s of relatively large-sized molecules with phenolic deprotonation sites, at reasonable computational cost.

## COMPUTATIONAL DETAILS

Full geometry optimizations and frequency calculations were performed with the package of programs Gaussian 03.[42] Different levels of theory have been used: BLYP, B3LYP, BHandHLYP, PBE, PBE0, PW91, BMK, TPSS, and M05-2x DFT methods and the MP2 wave function method for some of the smallest systems. All of them in conjunction with the 6-311++G(d,p) basis set. Local minima were identified by the absence of imaginary frequencies. The stationary points were first modeled in gas phase (vacuum), and solvent effects were included a posteriori by single point calculations using a PCM, specifically the integral equation formalism (IEF-PCM).[43] In PCM calculations, the choice of the solute cavity is important because the computed energies and properties strongly depend on the cavity size. In the present study the cavity has been built using the united atom model for Hartree−Fock (UAHF) method,[44] at HF/6-31+g(d), which is the recommended approach for predicting free energies of solvation according to the Gaussian 03 User's Reference.[45]

Relative Gibbs free energies in solution $\Delta G_s$ have been computed using the Hess law and thermodynamic cycles, as the sum of the corresponding gas-phase free energy $\Delta G_s$ and the free energy of salvation ($\Delta G_{solv}$). They include standard thermal corrections at 298.15 K. In all the cases the used reference state is 1 M. The aqueous Gibbs free energies for the deprotonation reactions are in turn used to compute the acid equilibrium constant ($K_a$), according to

$$K_a = e^{-\Delta G_s/RT} \quad (1)$$

Then, the p$K_a$ values are obtained using its definition:

$$pK_a = -\log(K_a) \quad (2)$$

which can also be calculated directly from $\Delta G_s$, from expressions that depend on the particular scheme of reaction used to model the deprotonation process.

**Reaction Schemes.** The reaction schemes tested in this work are reported in Table 1, together with the expressions corresponding to direct calculations of p$K_a$ from $\Delta G_s$. The latter have been calculated using the strategy based on thermodynamic cycles described above. For example for the reaction scheme **A**, the cycle is

$$\begin{array}{ccccc} HA_{(g)} & \xrightarrow{\Delta G_g} & A^-_{(g)} & + & H^+_{(g)} \\ \uparrow -\Delta G_{solv}(HA) & & \downarrow \Delta G_{solv}(A^-) & & \downarrow \Delta G_{solv}(H^+) \\ HA_{(s)} & \xrightarrow{\Delta G_s} & A^-_{(s)} & + & H^+_{(s)} \end{array}$$

and the Gibbs free energy of reaction in solution $\Delta G_s$ is obtained as the sum of the Gibbs free energy of reaction in vacuum ($\Delta G_g$) and the difference in solvation free energies ($\Delta\Delta G_{solv}$):

$$\Delta G_s = \Delta G_g + \Delta\Delta G_{solv} \quad (3)$$

where $\Delta\Delta G_{solv}$ and $\Delta G_g$ are calculated as

$$\Delta\Delta G_{solv} = \Delta G_{solv}(A^-) + \Delta G_{solv}(H^+) - \Delta G_{solv}(HA) \quad (4)$$

$$\Delta G_g = G_g(A^-) + G_g(H^+) - \Delta G_g(HA) \quad (5)$$

with $\Delta G_{solv}$ representing the free energies of solvation of each species.

Scheme **A** is by far the most frequently used, probably because of its simplicity. However it involves the proton, and it is known that computational methods poorly reproduce the solvation energies of this particular species. Therefore the $\Delta G_g(H^+)$ and

Table 1. Reaction Schemes and the Corresponding Expressions to Directly Calculate p$K_a$ Values From $\Delta G_s$

| scheme | equilibrium | p$K_a$ = |
|---|---|---|
| **A** | $HA \leftrightarrow H^+ + A^-$ | $\Delta G_s/RT \ln(10)$ |
| **B** | $HA + H_2O \leftrightarrow H_3O^+ + A^-$ | $\Delta G_s/RT \ln(10) - \log[H_2O]$ |
| **C1** | $HA + 2H_2O \leftrightarrow H_3O^+ + A^-(H_2O)$ | $\Delta G_s/RT \ln(10) - 2\log[H_2O]$ |
| **C2** | $HA + 3H_2O \leftrightarrow H_3O^+ + A^-(2H_2O)$ | $\Delta G_s/RT \ln(10) - 3\log[H_2O]$ |
| **C3** | $HA + 4H_2O \leftrightarrow H_3O^+ + A^-(3H_2O)$ | $\Delta G_s/RT \ln(10) - 4\log[H_2O]$ |
| **D** | $HA + ref^- \leftrightarrow A^- + HRef$ | $\Delta G_s/RT \ln(10) + pK_a(HRef)$ |
| **E1** | $HA + OH^-(3H_2O) \leftrightarrow A^-(H_2O) + 3H_2O$ | $(\Delta G_s)/RT \ln(10) + 14 + 3\log[H_2O]$ |
| **E2** | $HA + OH^-(3H_2O) \leftrightarrow A^-(2H_2O) + 2H_2O$ | $\Delta G_s/RT \ln(10) - 2\log[H_2O]$ |
| **E3** | $HA + OH^-(3H_2O) \leftrightarrow A^-(3H_2O) + H_2O$ | $\Delta G_s/RT \ln(10) + 14 + \log[H_2O]$ |
| **F1** | $HA + 4H_2O \leftrightarrow H_3O^+(3H_2O) + A^-$ | $\Delta G_s/RT \ln(10) - 4\log[H_2O]$ |
| **F2** | $HA + 5H_2O \leftrightarrow H_3O^+(3H_2O) + A^-(H_2O)$ | $\Delta G_s/RT \ln(10) - 5\log[H_2O]$ |
| **F3** | $HA + 6H_2O \leftrightarrow H_3O^+(3H_2O) + A^-(2H_2O)$ | $\Delta G_s/RT \ln(10) - 6\log[H_2O]$ |
| **F4** | $HA + 7H_2O \leftrightarrow H_3O^+(3H_2O) + A^-(3H_2O)$ | $\Delta G_s/RT \ln(10) - 7\log[H_2O]$ |
| **CN1** | $HA(H_2O) + H_2O \leftrightarrow H_3O^+ + A^-(H_2O)$ | $\Delta G_s/RT \ln(10) - \log[H_2O]$ |
| **CN2** | $HA(H_2O) + 2H_2O \leftrightarrow H_3O^+ + A^-(2H_2O)$ | $\Delta G_s/RT \ln(10) - 2\log[H_2O]$ |
| **CN3** | $HA(H_2O) + 3H_2O \leftrightarrow H_3O^+ + A^-(3H_2O)$ | $\Delta G_s/RT \ln(10) - 3\log[H_2O]$ |
| **FN1** | $HA(H_2O) + 4H_2O \leftrightarrow H_3O^+(3H_2O) + A^-(H_2O)$ | $\Delta G_s/RT \ln(10) - 4\log[H_2O]$ |
| **FN2** | $HA(H_2O) + 5H_2O \leftrightarrow H_3O^+(3H_2O) + A^-(3H_2O)$ | $\Delta G_s/RT \ln(10) - 5\log[H_2O]$ |
| **FN3** | $HA(H_2O) + 6H_2O \leftrightarrow H_3O^+(3H_2O) + A^-(3H_2O)$ | $\Delta G_s/RT \ln(10) - 6\log[H_2O]$ |
| **EN1** | $HA(H_2O) + OH^-(3H_2O) \leftrightarrow A^-(H_2O) + 4H_2O$ | $\Delta G_s/RT \ln(10) + 14 + 4\log[H_2O]$ |
| **EN2** | $HA(H_2O) + OH^-(3H_2O) \leftrightarrow A^-(2H_2O) + 3H_2O$ | $\Delta G_s/RT \ln(10) + 14 + 3\log[H_2O]$ |
| **EN3** | $HA(H_2O) + OH^-(3H_2O) \leftrightarrow A^-(3H_2O) + 2H_2O$ | $\Delta G_s/RT \ln(10) + 14 + 2\log[H_2O]$ |

$\Delta G_{solv}(H^+)$ values have been derived from experiments. This constitutes the main disadvantage of using scheme **A** because the reported experimental values of the solvation free energy of the proton range from −259 to −264 kcal/mol.[19] Such variation is an important source of error in the calculation of p$K_a$s, i.e., it alone represents about 3 p$K_a$ units. In this work we have used $\Delta G_g(H^+)$ = −4.39 kcal/mol and $\Delta G_{solv}(H^+)$ = −265.89 kcal/mol, based on the recommendation of Camaioni and Schwerdtfeger.[46]

A simple strategy to avoid using experimental values of $\Delta G_{solv}$ is to use a water molecule as a coreactant (scheme **B**). This strategy can be extended, within the cluster-continuum model, by including more than one water molecule and modeling the proton and/or the anionic conjugated base explicitly solvated (schemes **C** and **F**). The difference between these two kinds of schemes is that in **C** schemes, the proton is solvated with only one water molecule, while in **F** schemes, it is solvated by four of them. Moreover the explicit solvation can also be applied to the acid (schemes **CN** and **FN**). However the number of charged species on both sides of the equilibrium is not conserved for any of these schemes.

Scheme **D**, on the other hand, usually lead to more reliable values of p$K_a$s since the number and the kind, of charged species is conserved on both sides of the chemical equation. This particular approach is known as the proton exchange method, the isodesmic method, or the relative method to calculate p$K_a$s. The main problem with this approach is that its outcome strongly depends on the choice of the reference acid (HRef). Therefore, the quality of the results would be determined by the structural similarity between HRef and HA, and also by the accuracy of the experimental value of p$K_a$(HRef). In this work we have used phenol as HRef, and its p$K_a$ value equal to 10.09.[47] From a theoretical point of view schemes **D** and **A** both have the same drawback, they directly depend on experimental data. Such experiment dependency weakens the predictive character of any computational methodology aiming for predicting p$K_a$s or any other chemical property. Thus schemes **E** and **EN** have the advantage of being experiment independent, and at the same time, they conserve the number and kind of charged species on both sides of the equilibrium. The variations among them (**E1**−**E3** and **EN1**−**EN3**) differentiate only on the number of the explicit water molecules and on the explicit solvation of the different species involved in the equilibrium.

It should be noticed that for those reaction schemes that explicitly includes water molecules, it is necessary to correct for the standard state of liquid water, i.e., 55.55 mol/L. In fact it has been previously demonstrated that ignoring such correction leads to systematic errors.[48,49]

## RESULTS AND DISCUSSION

**General Considerations.** Three of the nonsteroidal anti-inflammatory drugs studied in this work (profadol, tapentadol, and ketobemidone, Figure 1) have a tertiary amine site. Therefore, they may have more than one p$K_a$, and it is necessary to assess the relative acidity of these sites. To that purpose, we have modeled the protonated species and estimated the deprotonation energies from both acid sites. The formed products are the nonionic (deprotonation from the amino site) and the zwitterion (formed by deprotonation from the phenol site). The zwitterionic forms were found to be 10.63, 11.08, and 20.57 kcal/mol higher in Gibbs free energy than the nonionic form for ketobemidone, profadol, and tapentadol, respectively, in aqueous solution. Therefore the zwitterionic forms of these compounds have been ruled out. In addition, for acetaminophen, we have estimated the relative ease of deprotonation from the −NH and −OH groups and confirmed that the Gibbs free energy of

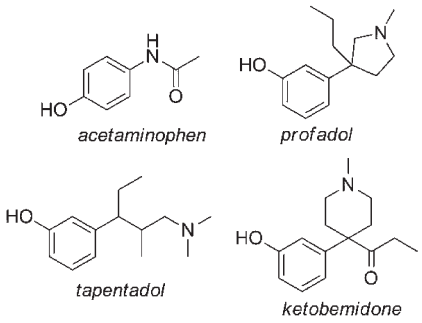


**Figure 1.** Studied nonsteroidal anti-inflammatory drugs.

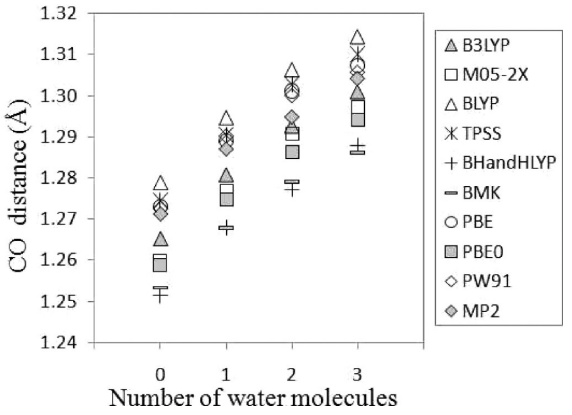


**Figure 2.** Variation of the CO distance in the phenoxide anions, with the number of explicit water molecules.

the anion formed by deprotonation of the −OH is 10.97 kcal/mol lower than that of the anion formed by deprotonation of the −NH, in aqueous solution. After these analyses it can be concluded that the processes modeled in the present work correspond to the actual phenolic deprotonations involved in their observed acid/base equilibrium.

In addition, it seems worthwhile to call attention on a previous report, by Liptak et al.,[10] describing that while the inclusion of the solvent has little effect on the phenols' geometries, it has a significant effect on the phenoxide anions. Since the geometrical parameter that changes the most is the CO distance, we have used it to analyze this point. In ref 10, dCO was found to be 1.251 Å in gas phase and 1.298 Å in solution, modeled using the continuum approach CPCM. In both cases the geometry optimizations were performed at HF/6-31+G(d) level of theory. In our case the solvent was modeled by including up to three explicit water molecules in the vicinity of the CO group. The dCO values obtained for each studied phenol, at every level of calculation, are reported in Table 1S, Supporting Information. The average value of dCO, including all methods of calculation and all the studied phenols, is found to be 1.266, 1.282, 1.293, and 1.301 Å for the anions with 0−3 explicit water molecules, respectively. The dCO average values, per method, have been plotted in Figure 2. The geometry optimizations performed with the BLYP functional systematically lead to the longest CO distances, while the shortest ones arise from optimizations with BHandHLYP. To our best knowledge there is no experimental data reported in the CO distance of phenoxides, but taking the value reported by Liptak et al.[10] as reference, it seems that the inclusion of one or two water molecules in the vicinity of the anions is enough to obtain good geometrical descriptions of these species.

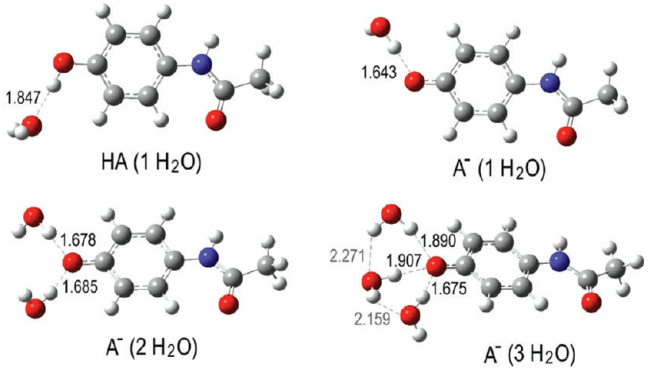


**Figure 3.** Geometries of the solute−water clusters, for the acetaminophen system, optimized at TPSS/6-311++G(d,p) level of theory.

Due to the large amount of calculations involved in the present work, it was unfeasible to perform exhaustive conformational analyses for the solute−water clusters. We have used chemical intuition and previous experience instead to construct the starting geometries in each case. To provide information on the optimized structures of the modeled clusters, the geometries corresponding to the acetaminophen system are shown in Figure 3. Since only small variations were found from changing the calculation method, we have chosen only one of them to show the geometrical distribution of the clusters. Equivalent configurations were located for all the studied phenolic systems.

**p$K_a$ Estimations.** The p$K_a$ values calculated for the studied nonsteroidal anti-inflammatory drugs (Figure 1) using the reaction schemes from Table 1 and the different levels of theory are provided as Supporting Information (Tables 2S−5S). The signed errors (SE) arising from comparison with the experimental values are reported in Tables 2−5. The SE values smaller than ±2 units of p$K_a$ have been highlighted in bold letters, since this is the limit of accuracy currently accepted for calculated p$K_a$s.[27] To facilitate rapid comparisons among all the reported data, the unsigned errors (UE) have been plotted in Figures 4−7.

**Acetaminophen.** For the phenolic deprotonation of acetaminophen, it was found that the outcomes from most of the tested reaction schemes lead to overestimated p$K_a$ values (Table 2). The exceptions are schemes **D**, **E1**, and **EN1**, which produce p$K_a$ values lower than the experimental one. In general the quality of the p$K_a$ predictions was found to be strongly dependent on the chosen scheme, while it is only slightly influenced by the method of calculation (Figure 4). The errors arising from MP2 calculations are larger than those obtained within the DFT framework for **E1** and **EN1** schemes, and smaller when the **E2** scheme is used. For all the other reaction schemes, the MP2 deviation from the experimental value was found to be intermediate, with respect to those arising from different DFT functionals. Therefore there is no use to additional computational cost arising from using MP2 instead of DFT. Reaction schemes **C**, **CN**, **F**, and **FN** lead to very large errors, regardless of the method of calculation and of the number of explicit water molecules included in the modeling. Curiously increasing the number of water molecules when using these reaction schemes does not improve but worsens the results for acetaminophen. In general the reaction schemes leading to the smallest errors are **A**,

Table 2. Signed Errors for the Calculated Values of the p$K_a$ of Acetaminophen (vs p$K_a$ = 9.5)[50]

| scheme | B3LYP | M05-2X | BLYP | TPSS | BHandHLYP | BMK | PBE | PBE0 | PW91 | MP2 |
|---|---|---|---|---|---|---|---|---|---|---|
| **A** | **0.88** | **1.47** | −2.61 | **−0.11** | 4.28 | **1.80** | −2.55 | **1.84** | −2.49 | 2.40 |
| **B** | 3.87 | 6.06 | 3.83 | **1.20** | 5.58 | 4.78 | **−0.01** | 3.64 | **0.15** | 4.24 |
| **C1** | 7.24 | 8.67 | 7.77 | 4.75 | 8.74 | 8.52 | 2.85 | 6.39 | 2.66 | 6.92 |
| **C2** | 11.32 | 11.08 | 12.69 | 8.80 | 12.09 | 12.71 | 6.66 | 9.58 | 5.95 | 10.53 |
| **C3** | 18.90 | 16.63 | 21.24 | 16.49 | 18.75 | 20.01 | 13.59 | 16.50 | 12.51 | 16.10 |
| **D** | **−0.27** | **−0.24** | **−0.97** | **−0.97** | **−0.01** | **0.10** | **−1.03** | **−0.15** | **−1.04** | 0.80 |
| **E1** | **−0.80** | **−1.98** | **−1.59** | **−0.70** | **−0.42** | **−1.91** | **−0.11** | **−0.37** | **0.87** | 2.95 |
| **E2** | 3.28 | **0.43** | 3.33 | 3.35 | 2.93 | 2.28 | 3.70 | 2.82 | 4.17 | 0.66 |
| **E3** | 10.86 | 5.98 | 11.88 | 11.04 | 9.59 | 9.58 | 10.63 | 9.74 | 10.72 | 6.24 |
| **F1** | 6.27 | 5.40 | 5.83 | 4.17 | 8.77 | 8.43 | **1.31** | 4.81 | **0.10** | 8.26 |
| **F2** | 9.63 | 8.00 | 9.76 | 7.72 | 11.92 | 12.17 | 4.16 | 7.56 | 2.60 | 10.93 |
| **F3** | 13.72 | 10.41 | 14.69 | 11.77 | 15.28 | 16.37 | 7.98 | 10.76 | 5.90 | 14.54 |
| **F4** | 21.30 | 15.96 | 23.24 | 19.46 | 21.94 | 23.67 | 14.91 | 17.68 | 12.45 | 20.12 |
| **CN1** | 4.97 | 6.96 | 4.94 | 2.38 | 6.81 | 6.36 | **1.12** | 4.72 | **1.26** | 5.13 |
| **CN2** | 9.05 | 9.37 | 9.87 | 6.44 | 10.17 | 10.56 | 4.94 | 7.92 | 4.56 | 8.74 |
| **CN3** | 16.62 | 14.92 | 18.41 | 14.11 | 16.82 | 17.85 | 11.86 | 14.83 | 11.11 | 14.31 |
| **FN1** | 7.36 | 6.29 | 6.94 | 5.35 | 10.00 | 10.02 | 2.44 | 5.90 | **1.21** | 9.14 |
| **FN2** | 11.44 | 8.70 | 11.86 | 9.40 | 13.35 | 14.21 | 6.25 | 9.09 | 4.50 | 12.75 |
| **FN3** | 19.02 | 14.25 | 20.41 | 17.08 | 20.02 | 21.51 | 13.18 | 16.01 | 11.06 | 18.33 |
| **EN1** | −3.07 | −3.69 | −4.42 | −3.07 | −2.35 | −4.07 | **−1.84** | −2.04 | **−0.53** | 4.74 |
| **EN2** | **1.01** | **−1.27** | **0.50** | **0.98** | **1.01** | **0.13** | **1.98** | **1.15** | 2.77 | 1.12 |
| **EN3** | 8.58 | 4.27 | 9.04 | 8.66 | 7.67 | 7.42 | 8.90 | 8.07 | 9.32 | 4.45 |

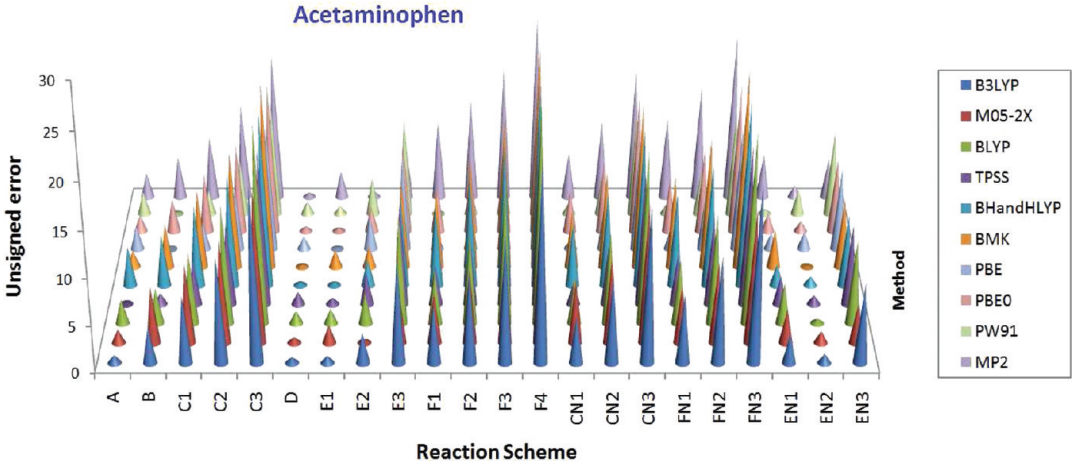


**Figure 4.** Unsigned errors for the calculated values of the phenolic p$K_a$ of acetaminophen.

**D**, **E1**, and **EN2**. With the exception of scheme **A**, these schemes are just those that conserve the number and kind of charged species at both sides of the equilibrium, and among them scheme **A** is just the one leading to largest errors.

Analyzing in more detail the results reported in Table 2, some peculiarities become evident. For the reaction scheme **A**, functionals BLYP, BHandHLYP, PBE, and PW91 produce a SE larger than 2 units of p$K_a$. PW91 also yields a SE larger than 2 within scheme **EN2**. On the other hand the SE, obtained from M05-2X calculations in conjunction with scheme **E2** and from TPSS using the scheme **B**, is significantly lower than this limit. PBE and PW91 functionals also produce very low errors when used in conjunction with the **B** scheme. In fact these are the only two functional for which the UE values arising from using scheme **B** are smaller than those obtained from scheme **A**. UE values smaller than 2 were also obtained for acetaminophen from PBE calculations using reaction schemes **F1**, **CN1**, and **EN1** and from PW91 calculations using react **F1**, **CN1**, **EN1**, and **FN1**.

**Profadol.** The general behavior of the calculated data for the phenolic deprotonation of profadol is very similar to that of acetaminophen. The reaction schemes leading to the smallest errors are **A**, **D**, **E1**, and **EN2**. Most of the tested reaction schemes yield overestimated values of p$K_a$ values (Table 3), while schemes **E1** and **EN1** lead to negative values of UE. The reaction scheme choice was found to have a much stronger influence on the quality of the results than on the particular method of calculation

Table 3. Signed Errors for the Calculated Values of the p$K_a$ of Profadol (vs p$K_a$ = 10.27)[51]

| scheme | B3LYP | M05-2X | BLYP | TPSS | BHandHLYP | BMK | PBE | PBE0 | PW91 |
|---|---|---|---|---|---|---|---|---|---|
| A | **1.45** | **1.98** | **−1.48** | **1.04** | 4.86 | 2.67 | **−1.37** | 2.34 | **−1.29** |
| B | 4.45 | 6.57 | 4.97 | 2.35 | 6.15 | 5.65 | **1.18** | 4.14 | **1.35** |
| C1 | 7.57 | 9.02 | 8.89 | 5.26 | 8.82 | 9.42 | 3.77 | 6.51 | 3.45 |
| C2 | 12.34 | 12.58 | 14.12 | 9.96 | 12.80 | 13.96 | 7.90 | 10.76 | 7.34 |
| C3 | 21.21 | 19.06 | 23.99 | 19.64 | 20.18 | 21.95 | 16.83 | 18.82 | 15.78 |
| D | **0.31** | **0.26** | **0.17** | **0.19** | **0.57** | **0.98** | **0.16** | **0.36** | **0.16** |
| E1 | **−0.46** | **−1.62** | **−0.48** | **−0.19** | **−0.33** | **−1.01** | **0.82** | **−0.26** | **1.67** |
| E2 | 4.30 | **1.93** | 4.76 | 4.51 | 3.64 | 3.53 | 4.94 | 3.99 | 5.55 |
| E3 | 13.17 | 8.41 | 14.63 | 14.19 | 11.02 | 11.52 | 13.87 | 12.06 | 14.00 |
| F1 | 6.84 | 5.90 | 6.96 | 5.32 | 9.34 | 9.31 | 2.49 | 5.32 | **1.30** |
| F2 | 9.97 | 8.35 | 10.88 | 8.22 | 12.01 | 13.07 | 5.09 | 7.68 | 3.39 |
| F3 | 14.74 | 11.91 | 16.12 | 12.93 | 15.99 | 17.62 | 9.22 | 11.93 | 7.28 |
| F4 | 23.61 | 18.39 | 25.99 | 22.61 | 23.37 | 25.61 | 18.15 | 20.00 | 15.73 |
| CN1 | 5.02 | 6.39 | 5.85 | 2.58 | 6.61 | 6.40 | **1.61** | 4.40 | **1.64** |
| CN2 | 9.79 | 9.95 | 11.08 | 7.29 | 10.59 | 10.95 | 5.75 | 8.65 | 5.54 |
| CN3 | 18.65 | 16.43 | 20.95 | 16.96 | 17.96 | 18.94 | 14.67 | 16.71 | 13.97 |
| FN1 | 7.42 | 5.73 | 7.85 | 5.55 | 9.81 | 10.06 | 2.93 | 5.58 | **1.59** |
| FN2 | 12.18 | 9.28 | 13.08 | 10.25 | 13.78 | 14.60 | 7.06 | 9.82 | 5.47 |
| FN3 | 21.05 | 15.76 | 22.95 | 19.93 | 21.16 | 22.60 | 15.99 | 17.89 | 13.92 |
| EN1 | −3.02 | −4.26 | −3.52 | −2.87 | −2.54 | −4.03 | **−1.35** | −2.37 | **−0.14** |
| EN2 | **1.75** | **−0.69** | **1.72** | **1.83** | **1.44** | **0.52** | 2.79 | **1.88** | 3.75 |
| EN3 | 10.61 | 5.78 | 11.59 | 11.51 | 8.81 | 8.51 | 11.71 | 9.94 | 12.19 |

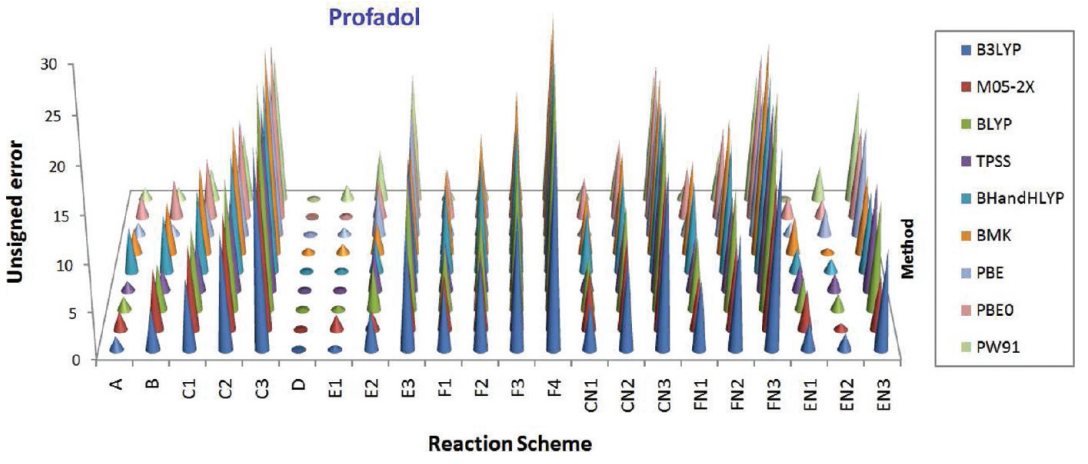


**Figure 5.** Unsigned errors for the calculated values of the phenolic p$K_a$ of profadol.

(Figure 5). The largest deviations from the experimental value arise from using reaction schemes **C**, **CN**, **F** and **FN**, and the inclusion of more water molecules in these schemes increases the UE values.

Some deviations from these general trends were found. DFT functionals BHandHLYP, BMK, and PBE0 produce a SE larger than 2 units of p$K_a$ when used in conjunction with scheme **A**. This also happens when computing scheme **EN2** with the PBE functional. As it was the case for acetaminophen, M05-2X calculations using the scheme **E2** lead to an UE < 2. However an UE value is significantly larger for profadol than for acetaminophen (1.92 vs 0.43). PBE and PW91 functionals produce UE < 2 when used in conjunction with the **B** scheme. However, in this case, their values are similar in magnitude but opposite in sign than those obtained from scheme **A**. UE values smaller than 2 were also obtained for profadol from PBE calculations using reaction schemes **CN1** and **EN1** and from PW91 calculations using reactions **F1**, **CN1**, **FN1**, and **EN1**.

**Tapentadol.** For this compound the general trends are similar to those of acetaminophen and profadol. It was found that most of the calculated values of p$K_a$ are overestimated (Table 4). The exceptions arise mainly from using schemes **E1** and **EN1**, but they are fewer than those found for acetaminophen and profadol. The quality of the p$K_a$ predictions was found to be more influenced by the chosen scheme than by the method of calculation (Figure 6). Reaction schemes **C**, **CN**, **F**, and **FN** lead to very large errors, regardless of the method of calculation and the number of explicit water molecules included in the modeling.

Table 4. Signed Errors for the Calculated Values of the p$K_a$ of Tapentadol (vs p$K_a$ = 10.09)[51]

| scheme | B3LYP | M05-2X | BLYP | TPSS | BHandHLYP | BMK | PBE | PBE0 | PW91 |
|---|---|---|---|---|---|---|---|---|---|
| **A** | 2.15 | 2.36 | **−0.88** | **1.83** | 5.45 | 2.11 | **−0.52** | 3.04 | **−0.41** |
| **B** | 5.15 | 6.95 | 5.56 | 3.14 | 6.74 | 5.09 | 2.03 | 4.84 | 2.23 |
| **C1** | 8.03 | 10.72 | 9.05 | 5.97 | 9.09 | 9.53 | 4.29 | 6.56 | 4.14 |
| **C2** | 11.98 | 15.09 | 14.00 | 10.46 | 12.84 | 13.98 | 8.39 | 11.05 | 7.78 |
| **C3** | 19.28 | 18.19 | 22.03 | 17.35 | 19.11 | 21.12 | 14.60 | 17.10 | 13.60 |
| **D** | **1.01** | **0.65** | **0.77** | **0.97** | **1.16** | **0.42** | **1.01** | **1.06** | **1.04** |
| **E1** | **−0.01** | **0.07** | **−0.32** | **0.52** | **−0.07** | **−0.90** | **1.33** | **−0.21** | 2.35 |
| **E2** | 3.94 | 4.44 | 4.64 | 5.00 | 3.68 | 3.55 | 5.43 | 4.29 | 5.99 |
| **E3** | 11.24 | 7.55 | 12.67 | 11.90 | 9.95 | 10.69 | 11.65 | 10.33 | 11.81 |
| **F1** | 7.54 | 6.29 | 7.56 | 6.11 | 9.93 | 8.75 | 3.35 | 6.01 | 2.17 |
| **F2** | 10.42 | 10.05 | 11.04 | 8.93 | 12.28 | 13.19 | 5.60 | 7.73 | 4.08 |
| **F3** | 14.38 | 14.42 | 16.00 | 13.43 | 16.03 | 17.64 | 9.70 | 12.23 | 7.73 |
| **F4** | 21.67 | 17.53 | 24.03 | 20.32 | 22.30 | 24.77 | 15.92 | 18.27 | 13.55 |
| **CN1** | 4.76 | 7.33 | 5.31 | 2.69 | 6.24 | 5.46 | **1.59** | 3.84 | **1.42** |
| **CN2** | 8.72 | 11.71 | 10.27 | 7.18 | 10.00 | 9.92 | 5.69 | 8.34 | 5.07 |
| **CN3** | 16.01 | 14.81 | 18.30 | 14.07 | 16.26 | 17.04 | 11.90 | 14.38 | 10.88 |
| **FN1** | 7.16 | 6.66 | 7.31 | 5.66 | 9.44 | 9.12 | 2.91 | 5.02 | **1.37** |
| **FN2** | 11.11 | 11.03 | 12.27 | 10.14 | 13.19 | 13.57 | 7.00 | 9.51 | 5.01 |
| **FN3** | 18.41 | 14.14 | 20.30 | 17.04 | 19.46 | 20.70 | 13.22 | 15.55 | 10.83 |
| **EN1** | −3.28 | −3.32 | −4.05 | −2.76 | −2.91 | −4.97 | **−1.37** | −2.93 | **−0.36** |
| **EN2** | **0.68** | **1.06** | **0.91** | **1.73** | **0.85** | **−0.51** | 2.73 | **1.57** | 3.28 |
| **EN3** | 7.97 | 4.16 | 8.93 | 8.62 | 7.11 | 6.61 | 8.95 | 7.61 | 9.10 |

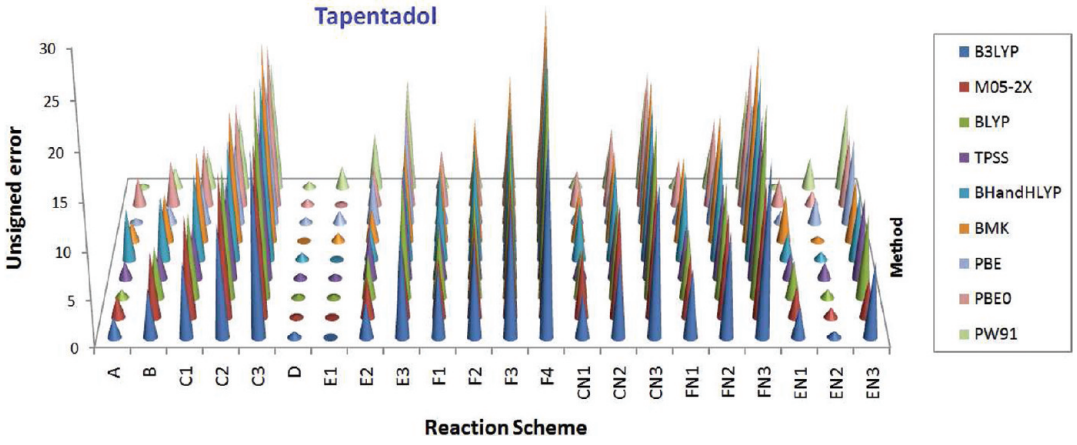


**Figure 6.** Unsigned errors for the calculated values of the phenolic p$K_a$ of tapentadol.

Moreover the inclusion of more water molecules in these schemes increases the UE values. The reaction schemes leading to the smallest errors were found to be **A**, **D**, **E1**, and **EN2**.

In contrast to these general trends, some peculiarities were found. DFT functionals BHandHLYP, BMK, and PBE0 produce a SE larger than 2 units of p$K_a$ when used in conjunction with scheme **A**, as it was the case for profadol. Computing scheme **EN2** with the PW91 functional also lead UE > 2. PBE and PW91 functionals produce UE < 2, when used in conjunction with the **B** scheme. However, in this case their values are similar in magnitude but opposite in sign than those obtained from scheme **A**, also in line with the results for profadol. UE values smaller than 2 were also obtained for profadol from PBE calculations using reaction schemes **CN1** and **EN1** and from PW91 calculations using reactions **F1**, **CN1**, and **EN1**. The outcomes from PBE and PW91 calculations are similar to what was described for acetaminophen and profadol.

**Ketobemidone.** In this case the general behavior of the calculated data is congruent with what was described for the other compounds in the tested set. The reaction schemes leading to the smallest errors are **A**, **D**, **E1**, and **EN2**. Most of the tested reaction schemes yield overestimated values of p$K_a$ values (Table 5), while schemes **D**, **E1**, and **EN1** lead to negative values of UE. The reaction scheme choice was found to have a much stronger influence on the quality of the results than the particular method of calculation (Figure 7). The largest deviations from the experimental value arise from using reaction schemes **C**, **CN**, **F** and **FN**, and the inclusion of more water molecules in these schemes increases the UE values.

Table 5. Signed Errors for the Calculated Values of the p$K_a$ of Ketobemidone (vs p$K_a$ = 9.96)[51]

| scheme | B3LYP | M05-2X | BLYP | TPSS | BH&HLYP | BMK | PBE | PBE0 | PW91 |
|---|---|---|---|---|---|---|---|---|---|
| **A** | **1.38** | **1.59** | **−1.60** | **0.77** | 4.65 | 2.00 | **−1.60** | 2.13 | **−1.57** |
| **B** | 4.37 | 6.17 | 4.84 | 2.08 | 5.94 | 4.98 | **0.95** | 3.93 | **1.07** |
| **C1** | 8.12 | 9.48 | 9.19 | 5.76 | 9.04 | 8.86 | 4.34 | 7.10 | 4.14 |
| **C2** | 12.13 | 12.86 | 13.62 | 10.05 | 12.66 | 12.93 | 7.82 | 10.55 | 7.25 |
| **C3** | 21.62 | 18.75 | 24.09 | 19.85 | 20.86 | 22.87 | 17.10 | 19.13 | 15.88 |
| **D** | **0.23** | **−0.13** | **0.05** | **−0.08** | **0.36** | **0.31** | **−0.07** | **0.15** | **−0.12** |
| **E1** | **0.08** | **−1.16** | **−0.18** | **0.31** | **−0.12** | **−1.57** | **1.38** | **0.34** | 2.35 |
| **E2** | 4.09 | 2.22 | 4.25 | 4.59 | 3.51 | 2.50 | 4.87 | 3.78 | 5.46 |
| **E3** | 13.58 | 8.11 | 14.73 | 14.40 | 11.71 | 12.44 | 14.14 | 12.36 | 14.09 |
| **F1** | 6.77 | 5.51 | 6.84 | 5.05 | 9.14 | 8.64 | 2.27 | 5.11 | **1.01** |
| **F2** | 10.51 | 8.81 | 11.18 | 8.73 | 12.23 | 12.51 | 5.65 | 8.27 | 4.08 |
| **F3** | 14.52 | 12.20 | 15.62 | 13.02 | 15.86 | 16.58 | 9.14 | 11.73 | 7.19 |
| **F4** | 24.02 | 18.09 | 26.09 | 22.82 | 24.06 | 26.53 | 18.42 | 20.30 | 15.82 |
| **CN1** | 4.62 | 6.74 | 5.19 | 2.34 | 6.10 | 6.15 | **1.38** | 4.29 | **1.53** |
| **CN2** | 8.63 | 10.12 | 9.63 | 6.63 | 9.73 | 10.22 | 4.87 | 7.75 | 4.65 |
| **CN3** | 18.12 | 16.01 | 20.10 | 16.43 | 17.92 | 20.16 | 14.13 | 16.32 | 13.27 |
| **FN1** | 7.01 | 6.07 | 7.19 | 5.31 | 9.30 | 9.80 | 2.70 | 5.47 | **1.48** |
| **FN2** | 11.02 | 9.45 | 11.63 | 9.60 | 12.92 | 13.87 | 6.18 | 8.92 | 4.59 |
| **FN3** | 20.52 | 15.34 | 22.10 | 19.40 | 21.12 | 23.81 | 15.45 | 17.49 | 13.22 |
| **EN1** | −3.42 | −3.91 | −4.17 | −3.11 | −3.05 | −4.29 | **−1.58** | −2.47 | **−0.25** |
| **EN2** | **0.59** | **−0.52** | **0.27** | **1.18** | **0.58** | **−0.21** | **1.91** | **0.98** | 2.86 |
| **EN3** | 10.08 | 5.36 | 10.74 | 10.98 | 8.77 | 9.73 | 11.18 | 9.55 | 11.48 |

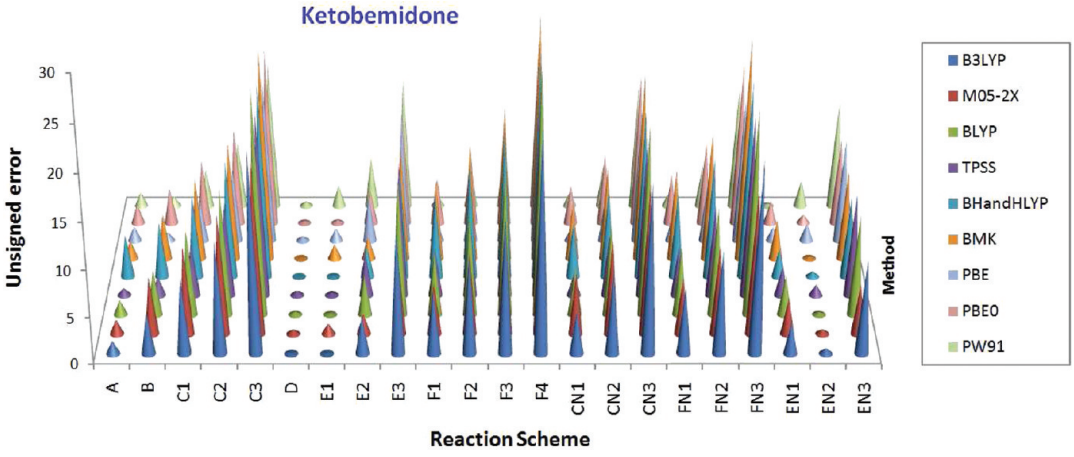


**Figure 7.** Unsigned errors for the calculated values of the phenolic p$K_a$ of ketobemidone.

The deviations from these general trends that were found for ketobemidone are described next. It is interesting to notice that even deviations are similar, though not identical, for all the tested cases. DFT functionals BHandHLYP, BMK, and PBE0 produce a SE larger than 2 units of p$K_a$, when used in conjunction with scheme **A**. This also happens when computing scheme **EN2** with the PW91 functional. PBE and PW91 functionals produce UE < 2 when used in conjunction with the **B** scheme. The UE values are similar in magnitude but opposite in sign than those obtained from scheme **A**. UE values smaller than 2 were also obtained for ketobemidone from PBE calculations using reaction schemes **CN1** and **EN1** and from PW91 calculations using reaction **F1**, **CN1**, **FN1**, and **EN1**.

**Other Phenols.** To further analyze the performance of the tested protocols, they have been applied to the p$K_a$ calculations of other phenols. They are: 3-methoxyphenol (*m*-$OCH_3$), 4-methylphenol (*p*-$CH_3$), 3-cyanophenol (*m*-CN), 4-methylthiophenol (*p*-$SCH_3$), 3-hydroxybenzaldehyde (*m*-CHO), and 2,4-dimethylphenol (*op*-$2CH_3$). The calculated p$K_a$s are provided as Supporting Information (Tables 6S−11S).

The average UEs, including these six compounds, are shown in Figure 8. The general behavior of the calculated p$K_a$s is in agreement with what was found for the nonsteroidal anti-inflammatory drugs. The reaction schemes leading to the smallest errors are **A**, **D**, **E1**, and **EN2**, which supports the good performance of these schemes. The choice of the method of calculation was found


dx.doi.org/10.1021/ct2001864 |J. Chem. Theory Comput. 2011, 7, 2528–2538

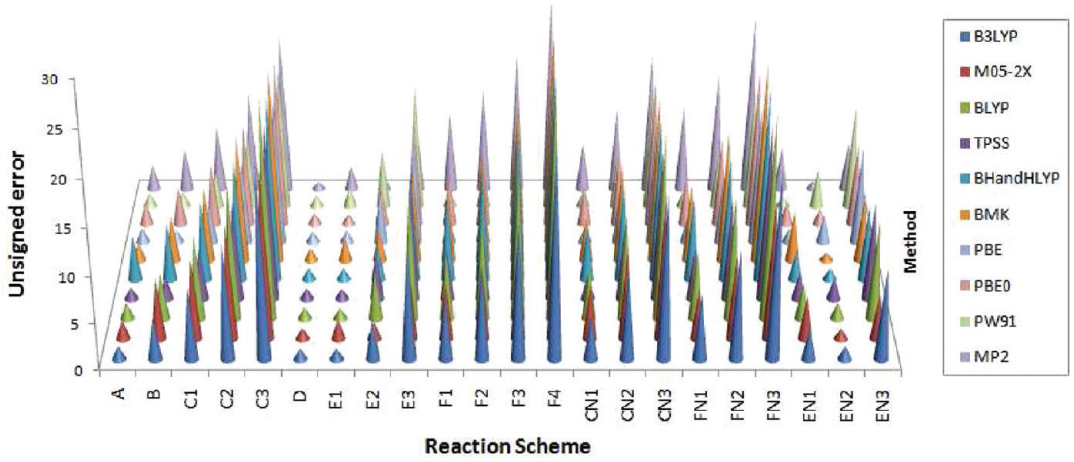


**Figure 8.** Unsigned errors for the calculated p$K_a$ values of phenolics (average, including *m*-$OCH_3$, *p*-$CH_3$, *m*-CN, *p*-$SCH_3$, *m*-CHO, and *op*-$2CH_3$).

**Table 6. Mean Unsigned Errors for the Calculated Values of Phenolic p$K_a$s within the Different Approaches**

| scheme | B3LYP | M05-2X | BLYP | TPSS | BHandHLYP | BMK | PBE | PBE0 | PW91 | MP2 |
|---|---|---|---|---|---|---|---|---|---|---|
| A | **1.46** | **1.97** | **1.69** | **1.11** | 4.88 | 2.03 | **1.58** | 2.35 | **1.47** | 2.67 |
| B | 4.45 | 6.55 | 4.89 | 2.33 | 6.17 | 5.01 | **1.24** | 4.14 | **1.43** | 4.51 |
| C1 | 7.67 | 9.00 | 8.88 | 5.50 | 8.75 | 8.52 | 4.01 | 6.67 | 3.81 | 7.18 |
| C2 | 12.18 | 12.55 | 14.19 | 10.34 | 12.88 | 13.65 | 8.22 | 10.71 | 7.78 | 10.91 |
| C3 | 20.48 | 18.18 | 23.16 | 18.69 | 19.90 | 21.22 | 15.90 | 18.12 | 14.90 | 16.96 |
| D | **0.88** | **0.82** | **0.94** | **0.90** | **0.96** | **1.00** | **0.91** | **0.90** | **0.94** | **0.80** |
| E1 | **0.86** | **1.77** | **0.95** | **0.82** | **0.82** | 2.02 | **1.15** | **0.77** | 2.02 | 2.69 |
| E2 | 4.14 | 2.01 | 4.83 | 4.89 | 3.72 | 3.21 | 5.26 | 3.94 | 5.99 | 1.04 |
| E3 | 12.44 | 7.54 | 13.79 | 13.24 | 10.74 | 10.79 | 12.94 | 11.35 | 13.11 | 7.09 |
| F1 | 6.85 | 5.89 | 6.89 | 5.30 | 9.36 | 8.67 | 2.55 | 5.32 | **1.38** | 8.53 |
| F2 | 10.06 | 8.33 | 10.87 | 8.46 | 11.94 | 12.17 | 5.32 | 7.84 | 3.75 | 11.19 |
| F3 | 14.58 | 11.88 | 16.19 | 13.31 | 16.07 | 17.30 | 9.54 | 11.88 | 7.72 | 14.92 |
| F4 | 22.88 | 17.52 | 25.16 | 21.66 | 23.09 | 24.88 | 17.22 | 19.30 | 14.84 | 20.98 |
| CN1 | 4.73 | 6.35 | 5.42 | 2.55 | 6.12 | 5.40 | **1.57** | 4.29 | **1.67** | 5.23 |
| CN2 | 9.25 | 9.90 | 10.74 | 7.40 | 10.25 | 10.52 | 5.73 | 8.34 | 5.61 | 8.96 |
| CN3 | 17.54 | 15.52 | 19.70 | 15.73 | 17.26 | 18.09 | 13.40 | 15.75 | 12.72 | 15.00 |
| FN1 | 7.13 | 5.69 | 7.42 | 5.52 | 9.31 | 9.05 | 2.83 | 5.47 | **1.63** | 9.24 |
| FN2 | 11.64 | 9.23 | 12.73 | 10.36 | 13.43 | 14.18 | 7.04 | 9.51 | 5.55 | 12.97 |
| FN3 | 19.94 | 14.87 | 21.70 | 18.70 | 20.46 | 21.75 | 14.72 | 16.92 | 12.66 | 19.02 |
| EN1 | 3.31 | 4.29 | 3.94 | 2.91 | 3.04 | 5.04 | **1.53** | 2.47 | **0.80** | 4.64 |
| EN2 | **1.27** | **1.01** | **1.40** | **1.94** | **1.12** | **0.65** | 2.77 | **1.59** | 3.82 | **0.91** |
| EN3 | 9.51 | 4.88 | 10.33 | 10.28 | 8.11 | 7.66 | 10.44 | 8.98 | 10.93 | 5.14 |

to have less influence on the quality of the result than the reaction scheme, also for these six phenols.

**Generalizations.** According to all the gathered data, several generalizations and recommendations can be made regarding the accuracy of phenolic p$K_a$s obtained from calculations. Table 6 shows the mean unsigned errors (MUE) for each scheme−method pair, including the 10 tested systems. The MUE values <2 units of p$K_a$ have been highlighted in bold letters. This limit has been chosen based on the accuracy currently accepted for calculated p$K_a$s.[27]

It should be kept in mind that these generalizations and recommendations are valid within the cluster-continuum model, using the PCM model and the strategy based on thermodynamic cycles. Changes in any of these aspects to the used methodology can lead to different outcomes.

The quality of the p$K_a$ predictions was found to be strongly dependent on the chosen scheme while only moderately influenced by the method of calculation. The reaction schemes leading to the smallest errors are **A**, **D**, **E1**, and **EN2**, while those producing the largest errors are **C**, **CN**, **F**, and **FN**. The MUE obtained when using reaction schemes **A**, **D**, **E1**, and **EN2** is within or below the currently accepted accuracy. In particular, schemes **D** and **E1** produce MUE values significantly lower than 2 units of p$K_a$, with most of the used methods of calculation (Figure 9). Therefore we recommend these approaches to predict reliable p$K_a$ values, for phenolic deprotonations of relative large systems, within the DFT

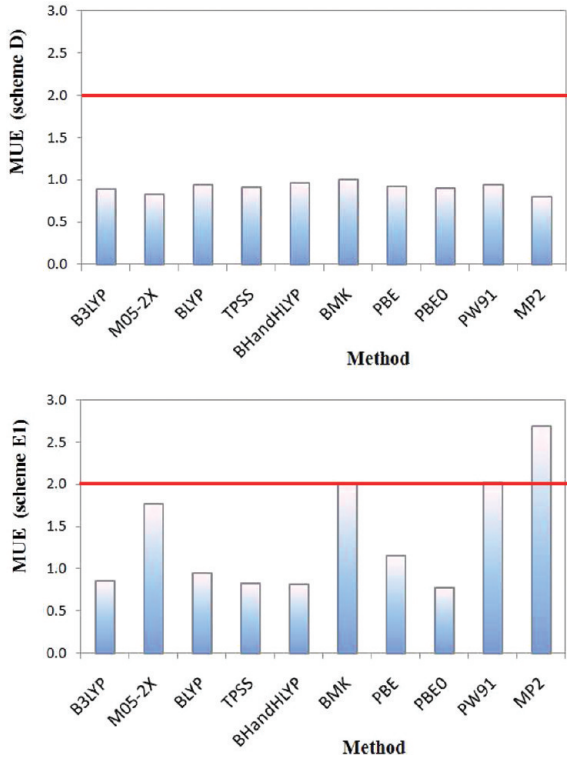


**Figure 9.** Mean unsigned errors for reaction schemes **D** and **E1**, including the 10 tested systems.

framework and using relatively modest basis sets, i.e., at very reasonable computational costs. We particularly recommend the **E1** reaction scheme, since it has the additional advantage of being experiment independent.

According to the above discussion, it can be stated that the schemes conserving the number and kind of charged species at both sides of the equilibrium lead to more accurate values of p$K_a$, since these approaches promote the cancelation of errors. So this seems to be the main problem to address for succeeding in the challenging task of predicting reliable acid constants. Among the reaction schemes leading to the smallest errors (**A**, **D**, **E1**, and **EN2**), all but scheme **A** conserve the number and the kind of charges at both sides of the equilibrium, and among them scheme **A** is just the one leading to largest errors. The reaction schemes producing the largest MUEs (**C**, **CN**, **F**, and **FN**) have all the charged species at the products side.

Even though the reaction scheme **A** is among those leading to the smallest errors, it is not recommended to be used in conjunction with BHandHLYP, MP2, or PBE0 methods (Table 6). On the contrary, reaction schemes **B**, **CN1**, and **EN1** seem to be good approaches for predicting phenolic p$K_a$s, provided that the calculations are performed with PBE or PW91 functionals. In particular the combination **EN1**−PW91 produced a very small value of MUE.

## CONCLUSIONS

Twenty-two reaction schemes have been tested, within the cluster-continuum model and including up to seven explicit water molecules, using the PCM and the strategy based on thermodynamic cycles [Gibbs free energy of reaction in solution $\Delta G_s$ is obtained as the sum of the Gibbs free energy of reaction in vacuum ($\Delta G_g$) and the solvation free energies ($\Delta\Delta G_{solv}$)].

The quality of the p$K_a$ predictions was found to be strongly dependent on the chosen scheme, while is only moderately influenced by the method of calculation. The reaction schemes leading to the smallest errors are **A**, **D**, **E1**, and **EN2**, while those producing the largest errors are **C**, **CN**, **F**, and **FN**.

In particular, schemes **D** and **E1** produce MUE values significantly lower than the currently accepted accuracy for p$K_a$ calculations. Therefore we recommend these approaches to predict reliable p$K_a$ values for phenolic deprotonations of relative large systems, within the DFT framework and using relatively modest basis sets, i.e., at very reasonable computational costs. We particularly recommend the **E1** reaction scheme, since it has the additional advantage of being experiment independent.

The best p$K_a$ values obtained from the **E1** reaction scheme (MUE < 1 units of p$K_a$) are those involving calculations with PBE0 (MUE = 0.77), TPSS (MUE = 0.82), BHandHLYP (MUE = 0.82), and B3LYP (MUE = 0.86) functionals. The best results when using the **D** reaction schemes correspond to calculations performed with MP2 (MUE = 0.80), M05-2X (MUE = 0.82), B3LYP (MUE = 0.88), PBE0 and TPSS (MUE = 0.90), and PBE (MUE = 0.91).

It is important to notice that these conclusions are valid within the cluster-continuum model, using the PCM model in conjunction with the UAHF cavity, and the strategy based on thermodynamic cycles. Changes in any of these aspects to the used methodology can lead to different results.

## ASSOCIATED CONTENT

**Supporting Information.** CO distances in the phenoxide anions, p$K_a$ values calculated using all the tested reaction schemes and levels of theory. This material is available free of charge via the Internet at http://pubs.acs.org.

## AUTHOR INFORMATION

**Corresponding Author**

*E-mail: agalano@xanum.uam.mx.

## ACKNOWLEDGMENT

The authors thank Laboratorio de Visualización y Cómputo Paralelo at UAM - Iztapalapa for the access to its computer facilities. T.C.-H. thanks CONACyT for Master fellowship and M.T.R.-S. for partial financial support through project 82932. We also acknowledge PROMEP for partial financial support through SIATA network and for the stipend to A.M.R.-Z.

## REFERENCES

(1) Riccardi, D.; Schaefer, P.; Cui, Q. *J. Phys. Chem. B* **2005**, *109*, 17715.
(2) Brown, T. N.; Mora-Diez, N. *J. Phys. Chem. B* **2006**, *110*, 9270.
(3) Thomas, G. *Medicinal Chemistry: An Introduction*; John Wiley & Sons: West Sussex, U.K., 2000.
(4) Albert, A.; Serjeant, E. P. *The Determination of Ionization Constants*; Chapman and Hall: New York, 1984.
(5) Pliego, J. R., Jr.; Riveros, J. M. *J. Phys. Chem. A* **2002**, *106*, 7434.
(6) Schuurmann, G. S.; Cossi, M.; Barone, V.; Tomasi, J. *J. Phys. Chem. A* **1998**, *102*, 6706.
(7) Silva, C. O.; Silva, M. A.; Nascimento, M. A. C. *J. Phys. Chem. A* **1999**, *103*, 11194.

(8) Toth, A. M.; Liptak, M. D.; Phillips, D. L.; Shields, G. C. *J. Chem. Phys.* **2001**, *114*, 4595.
(9) Liptak, M. D.; Shields, G. C. *J. Am. Chem. Soc.* **2001**, *123*, 7314.
(10) Liptak, M. D.; Gross, K. C.; Seybold, P. G.; Feldgus, S.; Shields, G. C. *J. Am. Chem. Soc.* **2002**, *124*, 6421.
(11) Chipman, D. M. *J. Phys. Chem. A* **2002**, *106*, 7413.
(12) Adam, K. R. *J. Phys. Chem. A* **2002**, *106*, 11963.
(13) Saracino, G. A. A.; Improta, R.; Barone, V. *Chem. Phys. Lett.* **2003**, *373*, 411.
(14) Klamt, A.; Eckert, F.; Diedenhofen, M.; Beck, M. E. *J. Phys. Chem. A* **2003**, *107*, 9380.
(15) Almerindo, G. I.; Tondo, D. W.; Pliego, J. R., Jr. *J. Phys. Chem. A* **2004**, *108*, 166.
(16) Magill, A. M.; Cavell, K. J.; Yates, B. F. *J. Am. Chem. Soc.* **2004**, *126*, 8717.
(17) Vianello, R.; Maksic, Z. B. *Eur. J. Org. Chem.* **2005**, *16*, 3571.
(18) Namazian, M.; Zakery, M.; Noorbala, M. R.; Coote, M. L. *Chem. Phys. Lett.* **2008**, *451*, 163.
(19) Ho, J.; Coote, M. L. *J. Chem. Theory Comput.* **2009**, *5*, 295.
(20) Song, Y.; Mao, J.; Gunner, M. R. *J. Comput. Chem.* **2009**, *30*, 2231.
(21) Delgado, E. J. *Chem. Phys. Lett.* **2009**, *471*, 133.
(22) Casasnovas, R.; Frau, J.; Ortega-Castro, J.; Salvà, A.; Donoso, J.; Muñoz, F. *J. Mol. Struct. (THEOCHEM)* **2009**, *912*, 5.
(23) Dissanayake, D. P.; Senthilnithy, R. *J. Mol. Struct. (THEOCHEM)* **2009**, *910*, 93.
(24) Uudsemaa, M.; Kanger, T.; Lopp, M.; Tamm, T. *Chem. Phys. Lett.* **2010**, *485*, 83.
(25) Rayne, S.; Forest, K. *J. Mol. Struct. (THEOCHEM)* **2010**, *949*, 60.
(26) Casasnovas, R.; Fernandez, D.; Ortega-Castro, J.; Frau, J.; Donoso, J.; Muñoz, F. *Theor. Chem. Acc.* **2011**, in press DOI 10.1007/s00214-011-0945-5.
(27) Ho, J.; Coote, M. L. *Theor. Chem. Acc.* **2010**, *125*, 3.
(28) (a) Curtiss, L. A.; Raghavachari, K.; Redfern, P. C.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1998**, *109*, 7764. (b) Baboul, A. G.; Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **1999**, *110*, 7650. (c) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K.; Rassolov, V.; Pople, J. A. *J. Chem. Phys.* **1999**, *110*, 4703. (d) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *127*, 124105. (e) Curtiss, L. A.; Redfern, P. C.; Raghavachari, K. *J. Chem. Phys.* **2007**, *126*, 84108.
(29) (a) Ochterski, J. W.; Petersson, G. A.; Montgomery, J. A *J. Chem. Phys.* **1996**, *104*, 2598. (b) Montgomery, J. A., Jr.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **1999**, *110*, 2822. (c) Montgomery, J. A.; Frisch, M. J.; Ochterski, J. W.; Petersson, G. A. *J. Chem. Phys.* **2000**, *112*, 6532.
(30) Claverie, P.; Daudey, J. P.; Langlet, J.; Pullman, B.; Piazzola, D.; Huron, M. J. *J. Phys. Chem.* **1978**, *82*, 405.
(31) Freitas, L. C. G.; Longo, R. L.; Simas, A. M. *J. Chem. Soc., Faraday Trans.* **1992**, *88*, 189.
(32) Pliego, J. R., Jr.; Riveros, J. M. *J. Phys. Chem. A* **2001**, *105*, 7241.
(33) (a) Rice-Evans, C. A.; Miller, N. J.; Bolwell, P. G.; Bramley, P. M.; Pridham, J. B. *Free Radical Res.* **1995**, *22*, 375. (b) Cook, N. C.; Samman, S. *J. Nutr. Biochem.* **1996**, *7*, 66. (c) Rice-Evans, C.; Miller, N. J.; Paganga, G. *Free Radical Biol. Med.* **1996**, *20*, 933. (d) Cao, G.; Sofic, E.; Prior, R. L. *Free Radical Biol. Med.* **1997**, *22*, 749. (e) Kahkonen, M. P.; Hopia, A. I.; Vuorela, H. J.; Rauha, J. P.; Pihlaja, K.; Kujala, T. S.; Heinonen, M. *J. Agric. Food Chem.* **1999**, *47*, 3954. (f) Lemanska, K.; Szymusiak, H.; Tyrakowsla, B.; Zielinski, R.; Soffers, A. E. M.; Rietens, I. M. C. M. *Free Radical Biol. Med.* **2001**, *31*, 869. (g) Materska, M.; Perucka, I. *J. Agric. Food Chem.* **2005**, *53*, 1750. (h) Fernandez-Panchon, M. S.; Villano, D.; Troncoso, A. M.; Garcia-Parrilla, M. C. *Crit. Rev. Food Sci.* **2008**, *48*, 649.
(34) (a) Renaud, S.; de Lorgeril, M. *Lancet* **1992**, *339*, 1523. (b) Morton, L. W.; Caccetta, R. A. A.; Puddey, I. B.; Croft, K. D. *Clin. Exp. Pharmacol. Physiol.* **2000**, *27*, 152. (c) Corder, R.; Mullen, W.; Khan, K. Q.; Marks, S. C.; Wood, E. G.; Carrier, M. J.; Crozier, A. *Nature* **2006**, *30*, 444. (d) Manach, C.; Mazur, A.; Scalbert, A. *Curr. Opin. Lipidol.* **2005**, *16*, 77.
(35) (a) Saito, M.; Hosoyama, H.; Ariga, T. *J. Agric. Food Chem.* **1998**, *46*, 1460. (b) Jiang, F.; Dusting, G. J. *Curr. Vasc. Pharmacol.* **2003**, *1*, 135.
(36) Sokmen, M.; Serkedjieva, J.; Daferera, D.; Gulluce, M.; Polissiou, M.; Tepe, B.; Akpulat, H. A.; Sahin, F.; Sokmen, A. *J. Agric. Food Chem.* **2004**, *52*, 3309.
(37) (a) Stavric, B. *Food Chem. Toxicol.* **1994**, *32*, 79. (b) Brown, D. M.; Kelly, G. E.; Husband, A. J. *Mol. Biotechnol.* **2005**, *30*, 253.
(38) (a) Ramassamy, C. *Eur. J. Pharmacol.* **2006**, *545*, 51. (b) Hamaguchi, T.; Ono, K.; Murase, A.; Yamada, M. *Am. J. Pathol.* **2009**, *175*, 2557.
(39) Hsu, S. *J. Am. Acad. Dermatol.* **2005**, *52*, 1049.
(40) Cornwell, T.; Cohick, W.; Raskin, I. *Phytochemistry* **2004**, *65*, 995.
(41) Hsu, C. L.; Yen, G. C. *Mol. Nutr. Food Res.* **2008**, *52*, 53.
(42) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; 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, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03*, revision E.01; Gaussian, Inc.: Wallingford, CT, 2004.
(43) (a) Cances, M. T.; Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3032. (b) Mennucci, B.; Tomasi, J. *J. Chem. Phys.* **1997**, *106*, 5151. (c) Mennucci, B.; Cances, E.; Tomasi, J. *J. Phys. Chem. B* **1997**, *101*, 10506. (d) Tomasi, J.; Mennucci, B.; Cances, E. *J. Mol. Struct (THEOCHEM)* **1999**, *464*, 211.
(44) Barone, V.; Cossi, M.; Tomasi, J. *J. Chem. Phys.* **1997**, *107*, 3210.
(45) Frisch, A. E.; Frisch, M.; Trucks, G. W. *Gaussian 03 User's Reference*; Gaussian, Inc.: Wallingford, CT, 2003.
(46) Camaioni, D. M.; Schwerdtfeger, C. A. *J. Phys. Chem. A* **2005**, *109*, 10795.
(47) Torres-Lapasio, J. R.; Garcia-Alvarez-Coque, M. C.; Bosch, E.; Roses, M. *J. Chromatogr. A* **2005**, *1089*, 170.
(48) Pliego, J. R. *Chem. Phys. Lett.* **2003**, *367*, 145.
(49) Bryantsev, V. S.; Diallo, M. S.; Goddard, W. A., III *J. Phys. Chem. B* **2008**, *112*, 9709.
(50) Hasemann, P.; Balk, M.; Preu, L.; Wätzig, H. *Electrophoresis* **2007**, *28*, 1779.
(51) Calculated using Advanced Chemistry Development (ACD/Laboratories) Software V11.02 ( 1994−2011 ACD/Laboratories).