

## Article

# Predicting $pK_a$ Values of Para-Substituted Aniline Radical Cations vs. Stable Anilinium Ions in Aqueous Media

 Jingxin Wang <sup>1</sup>, Hansun Fang <sup>2,\*</sup>, Zixi Zhong <sup>3</sup>, Huajun Huang <sup>2</sup> , Ximei Liang <sup>4</sup>, Yufan Yuan <sup>2</sup>, Wenwen Zhou <sup>2</sup>  and Davide Vione <sup>5,\*</sup> 
<sup>1</sup> Guangdong Provincial Engineering Technology Research Center of Public Health Detection and Assessment, School of Public Health, Guangdong Pharmaceutical University, Guangzhou 510310, China; wjxdaxue@163.com

<sup>2</sup> Key Laboratory of Poyang Lake Basin Agricultural Resource and Ecology of Jiangxi Province, College of Land Resource and Environment, Jiangxi Agricultural University, Nanchang 330045, China; huanghuajun2004@126.com (H.H.); 13767258573@163.com (Y.Y.); wenwenzhou@jxau.edu.cn (W.Z.)

<sup>3</sup> AP&A-Level Program, Guangzhou Foreign Language School, Guangzhou 511455, China; zhongzixi2007@163.com

<sup>4</sup> College of Animal Science and Technology, Jiangxi Agricultural University, Nanchang 330045, China; willie3@163.com

<sup>5</sup> Dipartimento di Chimica, Università di Torino, Via P. Giuria 5, 10125 Torino, Italy

\* Correspondence: fanghansun@163.com (H.F.); davide.vione@unito.it (D.V.)

**Abstract:** The focus of  $pK_a$  calculations has primarily been on stable molecules, with limited studies comparing radical cations and stable cations. In this study, we comprehensively investigate models with implicit solvent and explicit water molecules, direct and indirect calculation approaches, as well as methods for calculating free energy, solvation energy, and quasi-harmonic oscillator approximation for para-substituted aniline radical cations ( $R\text{-PhNH}_2^{\bullet+}$ ) and anilinium cations ( $R\text{-PhNH}_3^+$ ) in the aqueous phase. Properly including and positioning explicit  $H_2O$  molecules in the models is important for reliable  $pK_a$  predictions. For  $R\text{-PhNH}_2^{\bullet+}$ , precise  $pK_a$  values were obtained using models with one or two explicit  $H_2O$  molecules, resulting in a root mean square error (RMSE) of 0.563 and 0.384, respectively, for both the CBS-QB3 and M062X(D3)/ma-def2QZVP methods. Further improvement was achieved by adding  $H_2O$  near oxygen-containing substituents, leading to the lowest RMSE of 0.310. Predicting  $pK_a$  values for  $R\text{-PhNH}_3^+$  was more challenging. CBS-QB3 provided an RMSE of 0.349 and the M062X(D3)/ma-def2QZVP method failed to calculate  $pK_a$  accurately (RMSE > 1). However, by adopting the double-hybrid functional method and adding  $H_2O$  near the R substituent group, the calculations were significantly improved with an average absolute difference ( $\Delta pK_a$ ) of 0.357 between the calculated and experimental  $pK_a$  values. Our study offers efficient and reliable methods for  $pK_a$  calculations of  $R\text{-PhNH}_2^{\bullet+}$  (especially) and  $R\text{-PhNH}_3^+$  based on currently mature quantum chemistry software.

**Keywords:** aniline; radical cation; cation;  $pK_a$ ; quantum chemical calculation



**Citation:** Wang, J.; Fang, H.; Zhong, Z.; Huang, H.; Liang, X.; Yuan, Y.; Zhou, W.; Vione, D. Predicting  $pK_a$  Values of Para-Substituted Aniline Radical Cations vs. Stable Anilinium Ions in Aqueous Media. *Molecules* **2024**, *29*, 4522. <https://doi.org/10.3390/molecules29194522>

Academic Editors: Michal Szostak and Elena G. Bagryanskaya

Received: 8 July 2024

Revised: 9 September 2024

Accepted: 20 September 2024

Published: 24 September 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The determination of acid-base dissociation equilibrium constants ( $pK_a$  values) is essential for studying the physicochemical properties and environmental fate of compounds in the aqueous phase, as changes in their protonation forms can significantly impact reaction rates, pathways, and mechanisms [1]. Traditionally, the  $pK_a$  value is determined by electrochemical methods [2] or by analyzing the absorption spectra of a compound under different pH conditions [3]. However, conducting  $pK_a$  measurements can be influenced by various factors. For instance, handling toxic compounds can be challenging, and impurities in the target compound may interfere with the measurement of absorption spectra [4]. Moreover, compounds with multiple ionization groups may exhibit multiple  $pK_a$  values that are difficult to distinguish solely through absorption spectroscopy [5,6].

When measuring the  $pK_a$  of compounds with short lifetimes, such as organic and inorganic radicals, compared to stable chemicals, additional challenges arise. Normally, determining the  $pK_a$  values of short-lived species relies on measuring changes in transient absorption spectra, typically using methods like laser flash photolysis or pulse radiolysis [7,8]. However, these techniques can be costly, and their sensitivity depends on both the detector and the emission spectral intensity of the probe light, especially in the ultraviolet range. Insufficient probe photons may lead to neglect of important peaks in the short-wavelength range [9], which could be critical for accurate  $pK_a$  determinations.

In recent years, research has shown that  $pK_a$  values can be precisely determined by quantum chemical (QC) calculations [10,11]. The critical factor for achieving accurate calculation results primarily lies in choosing an appropriate solvent model [12,13]. Currently, the density-based solvation model (SMD) is widely employed during  $pK_a$  calculations. SMD is a continuum solvation model that relies on the quantum mechanical charge density of a solute molecule interacting with a continuum description of the solvent [14].

Several general methods in conjunction with SMD have been employed to increase the accuracy of  $pK_a$  calculations in the aqueous phase. Firstly, the utilization of direct and indirect approaches. The direct approach involves calculating  $pK_a$  values directly with solvent effects already incorporated in the molecular models, while the indirect approach relies on thermodynamic cycles between the gas and solvent phases [15,16]. Another method is the inclusion of explicit solvent molecules in the molecular models [17–19], although this may increase calculation time as the number of solvent molecules grows. Moreover, careful selection and utilization of specific calculation methods and basis sets for different forms of targeted compounds (e.g., neutral, cationic, and anionic molecules) can be helpful [20,21]. In addition, professional researchers have also developed advanced strategies, such as modifying the solute cavity [22–24] and creating new solvent models that better describe weak interaction forces between molecules [25]. However, the practical application of these advanced methods requires a profound understanding of (and experience with) the underlying theory. For example, cavity adjustments were found to be sensitive to molecular modes and the type of solvents used, and it is still uncertain whether the same rules can be applied across different systems [20]. Therefore, from the perspective of experimental chemists, and especially those focused on specific classes of compounds, a more practical and attractive approach to predict  $pK_a$  values involves utilizing pre-existing QC and data correction software, as well as the built-in solvent models, methods, and basis sets.

Aniline (-PhNH<sub>2</sub>) is a chemical group present in various agricultural, medical, and industrial compounds and natural dissolved organic matter (NDOM) [26–30]. The protonation of -PhNH<sub>2</sub> significantly influences the charge and state of the compounds. Additionally, the electron-rich nature of -PhNH<sub>2</sub> enables it to act as an electron donor and initiate redox reactions, resulting in the formation of a radical cation (R-PhNH<sub>2</sub><sup>•+</sup>, where R represents the substituent group) [31,32]. The R-PhNH<sub>2</sub><sup>•+</sup> species can either accept electrons from other reductive compounds or undergo deprotonation to form neutral radicals (R-PhNH<sup>•</sup>). R-PhNH<sub>2</sub><sup>•+</sup> tends to act as a potent electron acceptor, which can be reduced again to regenerate R-PhNH<sub>2</sub> [33], while R-PhNH<sup>•</sup> might undergo intramolecular reactions through rearrangement [31].

Pulse radiolysis studies have revealed that the average  $pK_a$  value of para-substituted R-PhNH<sub>2</sub><sup>•+</sup> is 7.3 [7], which is within the typical pH range found in natural water bodies. In contrast, the corresponding anilinium cations (R-PhNH<sub>3</sub><sup>+</sup>) have an average  $pK_a$  value of only 3.9 (Table 1) [34]. These findings suggest that, in the natural aqueous environment, the transformation pathways and kinetics of R-PhNH<sub>2</sub><sup>•+</sup> would be more sensitively influenced by the  $pK_a$  value in comparison with R-PhNH<sub>3</sub><sup>+</sup>.

**Table 1.** The reported pK<sub>a</sub> values of R-PhNH<sub>2</sub><sup>•+</sup> and R-PhNH<sub>3</sub><sup>+</sup>.

	R	H	C <sub>4</sub> H <sub>9</sub>	CF <sub>3</sub>	CH <sub>3</sub>	OCH <sub>3</sub>	CN	COCH <sub>3</sub>	I	NH <sub>2</sub>	SO <sub>3</sub> <sup>-</sup>
pK <sub>a</sub>	R-PhNH <sub>2</sub> <sup>•+</sup>	7.05 [35]	8.2 [7]	4.8 [7]	8.5 [36]	9.6 [7]	4 [7]	6.1 [7]	7.1 [7]	12 [7]	5.8 [7]
	R-PhNH <sub>3</sub> <sup>+</sup> <sup>a</sup>	4.62 [37], 4.58 [38]	4.95 [34]	2.92 [39], 2.75 [40], 2.57 [41]	5.12 [37]	5.29 [37]	1.75 [42], 1.82 [43]	2.19 [42], 2.26 [43]	3.78 [44]	5.94 [45], 6.2 [42]	3.25 [46], 2.93 [47], 3.32 [48]

<sup>a</sup>: Average pK<sub>a</sub> values were adopted if more than two values were available.

There have been studies reporting and calculating the pK<sub>a</sub> values of R-PhNH<sub>2</sub><sup>•+</sup> in organic solvents, such as DMSO [49,50]. However, there is limited information available on the calculation and prediction of pK<sub>a</sub> values of R-PhNH<sub>2</sub><sup>•+</sup> in water. Therefore, the main objective of this study is to identify a suitable QC method for accurately determining pK<sub>a</sub> based on the SMD model. Additionally, pK<sub>a</sub> values for R-PhNH<sub>3</sub><sup>+</sup> were also evaluated for comparison. Through a systematic study of pK<sub>a</sub> calculation methods for both R-PhNH<sub>2</sub><sup>•+</sup> and R-PhNH<sub>3</sub><sup>+</sup>, this research provides a practical tool for comprehensive prediction of the transformation of R-PhNH<sub>2</sub> in aqueous environments, and it also offers valuable insights into the ideas and approaches relevant to calculating pK<sub>a</sub> values of similar compounds.

## 2. Results and Discussion

### 2.1. Calculation of pK<sub>a</sub>

All quantum chemical calculations were performed with Gaussian 16 [51]. M062X(D3)/6-311++g(d,p), coupled with the IEFPCM solvent model, was selected as the optimization method [52–56]. For radicals, the Unrestricted Hartree–Fock (UHF) formalism was used, while the Restricted Hartree–Fock (RHF) formalism was applied for anilines. For compounds containing heavy atoms, such as iodine (I) in this study, a mixed strategy was adopted for molecular optimization. Specifically, light atoms belonging to the first four periods of elements were optimized using the 6-311++g(d,p) basis set, while the iodine (I) atom was optimized using the ma-def2TZVP basis set [57,58]. The geometries obtained by these methods were then used for all subsequent calculations. For the deprotonation of R-PhNH<sub>2</sub><sup>•+</sup> and R-PhNH<sub>3</sub><sup>+</sup>, pK<sub>a</sub> values were calculated according to Equation (1) [17].

$$pK_a = \frac{\Delta G_{\text{deprot}}(\text{sol})}{2.303RT} \quad (1)$$

where  $\Delta G_{\text{deprot}}(\text{sol})$  represents the free energy change for the deprotonation of compounds in the aqueous phase.  $\Delta G_{\text{deprot}}(\text{sol})$  was calculated by both direct and indirect (thermodynamic) approaches, as described in Section 3.1.  $R$  denotes the ideal gas constant (8.314 J mol<sup>-1</sup> K<sup>-1</sup>), and  $T$  represents the ambient temperature set at 298.15 K.

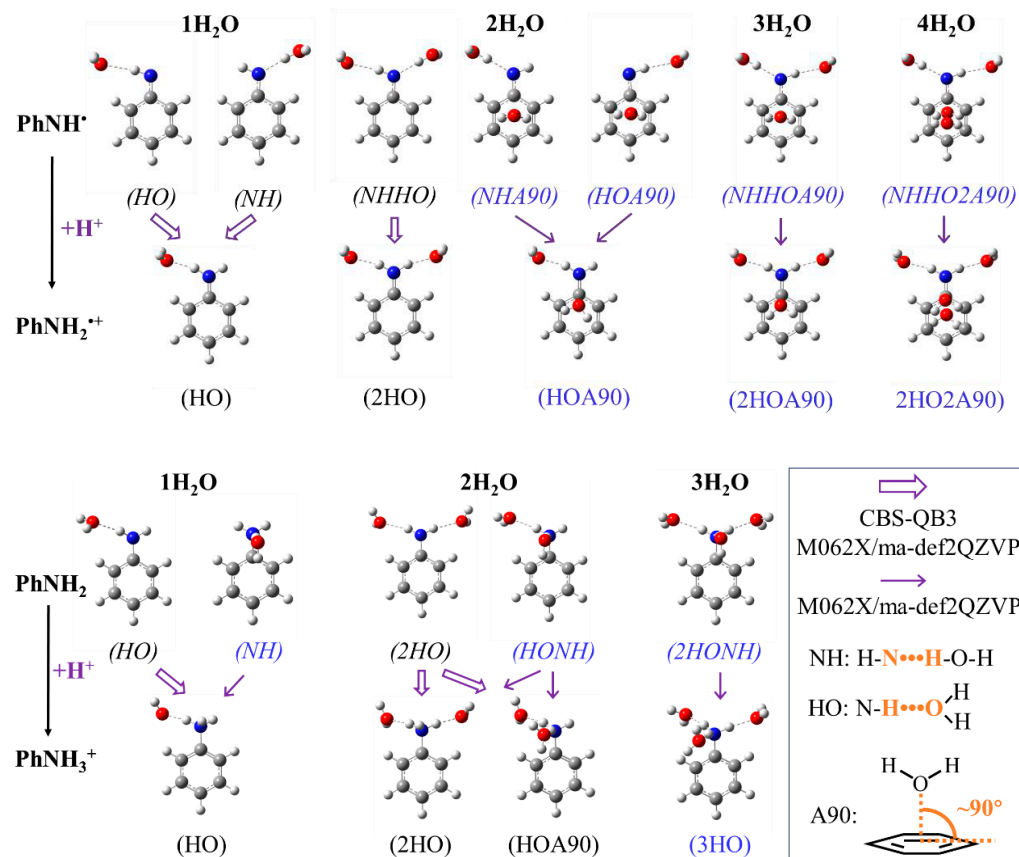
The complete basis set method CBS-QB3 is considered as an accurate and reliable method for calculating  $G(\text{sol})$  [16,59]. However, it is not applicable to compounds containing heavy atoms such as I-PhNH<sub>2</sub><sup>•+</sup> and I-PhNH<sub>3</sub><sup>+</sup>, as investigated herein. To address this limitation,  $\Delta G_{\text{deprot}}(\text{sol})$  was also calculated based on the method and basis set of M062X(D3)/ma-def2QZVP [52,57,58].

### 2.2. Numbers and Positions of Explicit Water Molecules in the Models

The effects of the addition of one to four explicit water (H<sub>2</sub>O) molecules were examined in the deprotonation of R-PhNH<sub>2</sub><sup>•+</sup> (one to four H<sub>2</sub>O molecules) and R-PhNH<sub>3</sub><sup>+</sup> (one to three H<sub>2</sub>O molecules). When constructing the molecular models, H<sub>2</sub>O molecules were positioned around the amino group (-NH<sub>x</sub>). The maximum number of added H<sub>2</sub>O molecules was determined based on the hydrogen bonds that could possibly be formed between -NH<sub>x</sub> and the H<sub>2</sub>O molecules. Some of the H<sub>2</sub>O molecules connected with -NH<sub>x</sub> were presumed to be hydrogen donors, while others served as electron-pair donors (Figure S1).

Figure 1 presents a group of optimized models illustrating the (de)protonated states of H-PhNH<sup>•</sup>/H-PhNH<sub>2</sub><sup>•+</sup> and H-PhNH<sub>2</sub>/H-PhNH<sub>3</sub><sup>+</sup> as a typical example. It is notable that some models undergo significant changes or fail to meet convergence criteria when

calculated based on the CBS-QB3 method (labeled in blue in Figure 1). This is primarily due to the readjustments made to the positions of H<sub>2</sub>O molecules during optimization with the CBS-QB3 method. Consequently, the final molecular models were significantly different from the originally optimized structures based on M062X(D3)/6-311++g(d,p).



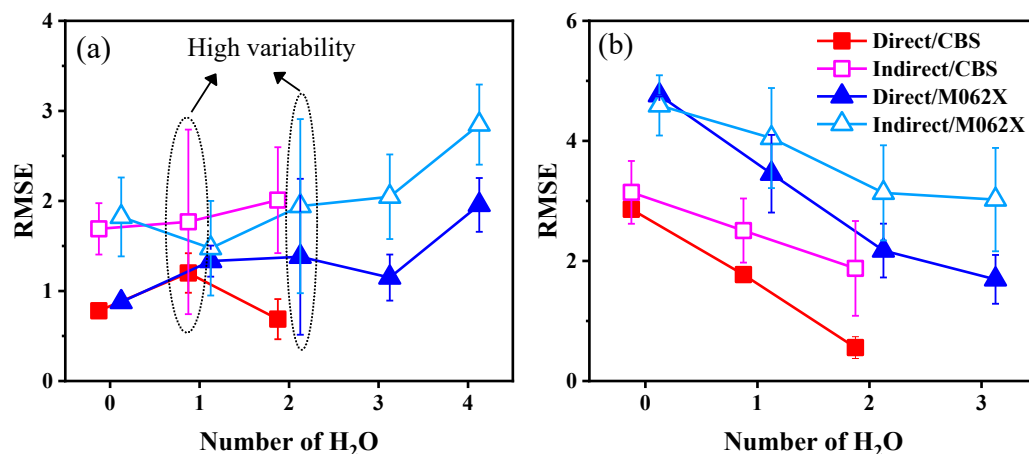
**Figure 1.** The optimized structures of H-PhNH<sup>•</sup>, H-PhNH<sub>2</sub><sup>•+</sup>, H-PhNH<sub>2</sub>, and H-PhNH<sub>3</sub><sup>+</sup> with 1–4 explicit H<sub>2</sub>O molecules at different positions. Labels HO and NH indicate that the amino group forms a hydrogen bond with H<sub>2</sub>O as a H donor and electron-pair (nitrogen) donor, respectively. A90 means one H<sub>2</sub>O molecule is positioned approximately perpendicular to the plane of the benzene ring of PhNH<sub>2</sub>. Italic font indicates the molecules in deprotonated form. Blue color indicates that the calculations of free energy were solely based on M062X(D3)/ma-def2QZVP and not on CBS-QB3.

Examples illustrating the structural changes when employing the CBS-QB3 method can be found in Figures S2 and S3. Therefore, for these models,  $G(\text{sol})$  and  $G(\text{gas})$  were solely calculated based on M062X(D3)/ma-def2QZVP.

### 2.3. Impact of the Number of Explicit H<sub>2</sub>O Molecules on pK<sub>a</sub> Calculations

All calculated energies, as well as the experimental results and calculated pK<sub>a</sub> values of R-PhNH<sub>2</sub><sup>•+</sup> and R-PhNH<sub>3</sub><sup>+</sup> in the aqueous phase, are summarized and compared in Tables S1–S6 in the Supporting Materials. The impact of the number of explicit H<sub>2</sub>O molecules on pK<sub>a</sub> calculations was assessed based on the variation of root mean square error (RMSE) as shown in Figure 2. For both R-PhNH<sub>2</sub><sup>•+</sup> and R-PhNH<sub>3</sub><sup>+</sup>, the CBS-QB3 method was found to be appropriate for models containing up to two H<sub>2</sub>O molecules.

However, if the number of H<sub>2</sub>O molecules was three or higher, using the CBS-QB3 method for further energy calculations would lead to significant structural changes from the originally optimized models (Figures S2 and S3).



**Figure 2.** Impact of the numbers of H<sub>2</sub>O molecules on pK<sub>a</sub> calculations for (a) R-PhNH<sub>2</sub><sup>•+</sup> and (b) R-PhNH<sub>3</sub><sup>+</sup> based on CBS-QB3 and M062X/ma-def2QZVP methods with direct and indirect approaches.

The lowest average values of root mean square error (RMSE<sub>aver</sub>) for R-PhNH<sub>2</sub><sup>•+</sup> and R-PhNH<sub>3</sub><sup>+</sup> were 0.687 and 0.557, respectively, demonstrating that the pK<sub>a</sub> values calculations of R-PhNH<sub>2</sub><sup>•+</sup> were comparable in performance to those of R-PhNH<sub>3</sub><sup>+</sup>. Moreover, when comparing models with and without H<sub>2</sub>O molecules, the largest differences in RMSE<sub>aver</sub> for R-PhNH<sub>2</sub><sup>•+</sup> and R-PhNH<sub>3</sub><sup>+</sup> were 1.211 and 3.067, respectively, indicating that adding H<sub>2</sub>O molecules is an efficient strategy to improve the results of pK<sub>a</sub> calculations [17,60], especially for R-PhNH<sub>3</sub><sup>+</sup>. The RMSE<sub>aver</sub> value for R-PhNH<sub>3</sub><sup>+</sup> consistently decreased with an increase in the number of H<sub>2</sub>O molecules in the models, indicating a corresponding increase in calculation accuracy. Always in the case of R-PhNH<sub>3</sub><sup>+</sup>, the CBS-QB3 method exhibited notably improved performance upon addition of H<sub>2</sub>O molecules compared to the M062X(D3)/ma-def2QZVP method.

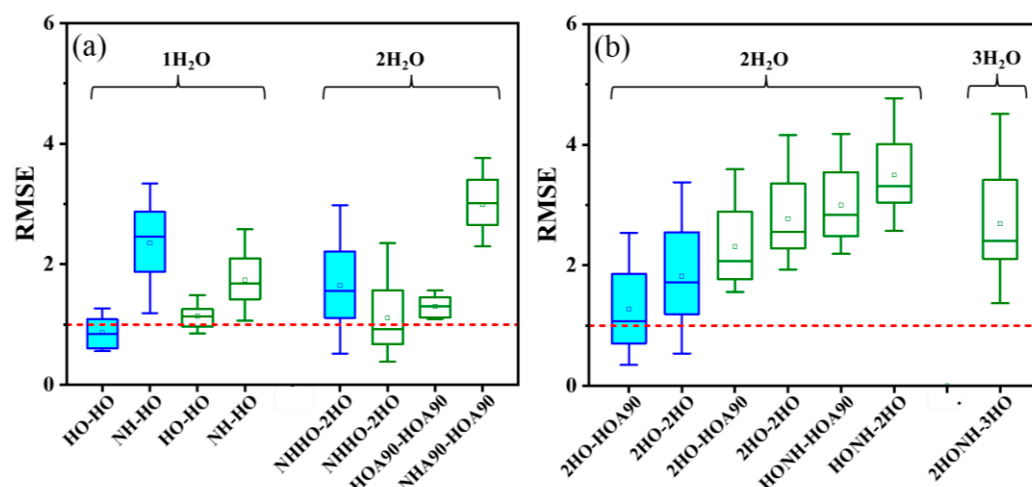
In the case of R-PhNH<sub>2</sub><sup>•+</sup>, the variation in calculation accuracy showed a more complex pattern. With the direct approach, the calculation accuracy initially decreased (i.e., RMSE<sub>aver</sub> increased) when adding only one H<sub>2</sub>O molecule. As more H<sub>2</sub>O molecules were included in the models, the calculation accuracy subsequently improved until the number of H<sub>2</sub>O molecules exceeded three, in which case the calculation accuracy decreased (RMSE<sub>aver</sub> increased) once again. Conversely, different results were obtained with the indirect approach. With the exception of adding a single H<sub>2</sub>O molecule based on the M062X(D3)/ma-def2QZVP method, the inclusion of additional H<sub>2</sub>O molecules generally decreased calculation accuracy with the indirect approach, as evidenced by an increase in RMSE<sub>aver</sub>.

When assessing the accuracy of calculation methods, it is crucial not to rely solely on RMSE<sub>aver</sub>. Instead, it is equally important to consider the standard deviation of RMSE (RMSE<sub>std</sub>). Notably, in models of R-PhNH<sub>2</sub><sup>•+</sup> (Figure 2a), there was a significant increase in RMSE<sub>std</sub> in the case of one (RMSE<sub>std</sub> = ±1.025) and two H<sub>2</sub>O molecules (RMSE<sub>std</sub> = ±0.967). The large span of RMSE<sub>std</sub> values obtained from these models of R-PhNH<sub>2</sub><sup>•+</sup> suggests that there is room for further improving pK<sub>a</sub> calculations, for instance, by further modifying the models and adjusting the calculation parameters. In contrast, much lower RMSE<sub>std</sub> values were obtained with R-PhNH<sub>3</sub><sup>+</sup> (ranging from ±0.017 to ±0.861), possibly indicating a more limited space for improvement of calculations based on the present approaches and methods.

#### 2.4. Impact of H<sub>2</sub>O Molecule Positions on pK<sub>a</sub> Calculation

The positions of explicit H<sub>2</sub>O molecules have rarely been considered in previous studies. Here, in addition to the number of H<sub>2</sub>O molecules, the influence of their relative positions to the amino group on the accuracy of pK<sub>a</sub> calculations was also studied, based on models with low RMSE<sub>aver</sub> (indicating high calculation accuracy) or with high RMSE<sub>std</sub>

(indicating potential for improvement), as obtained in the last section. Specifically, we considered R-PhNH<sub>2</sub><sup>•+</sup> models with one and two H<sub>2</sub>O molecules, as well as R-PhNH<sub>3</sub><sup>+</sup> models with two and three H<sub>2</sub>O molecules (Figure 3).



**Figure 3.** Impact of the H<sub>2</sub>O positions (x-axis and the meanings of the labels are referred to Figure 1) in models on pK<sub>a</sub> calculations of (a) R-PhNH<sub>2</sub><sup>•+</sup> and (b) R-PhNH<sub>3</sub><sup>+</sup> based on CBS-QB3 (blue and filled boxes) and M062X(D3)/ma-def2QZVP methods (green and unfilled boxes). The red dashed lines indicate RMSE = 1.

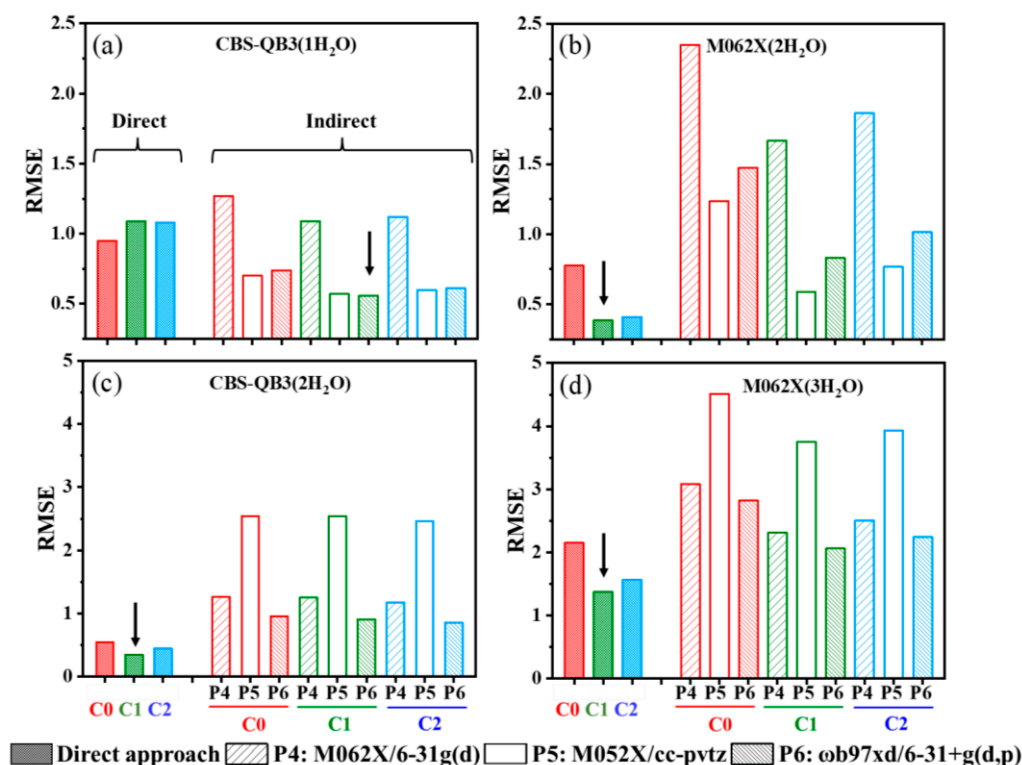
Regarding R-PhNH<sub>2</sub><sup>•+</sup> (Figure 3a), introducing a H<sub>2</sub>O molecule placed approximately perpendicular to the plane of the benzene ring would result in decreased accuracy. Based on the CBS-QB3 method, the best calculation results were achieved with models containing a single H<sub>2</sub>O molecule, which acts as an electron-couple donor for both R-PhNH<sup>•</sup> and R-PhNH<sub>2</sub><sup>•+</sup> (both in the form of HO structure as HO-HO, see Figure 1). In this model, the values of RMSE<sub>aver</sub> ± RMSE<sub>std</sub> were 0.865 ± 0.260.

As for the M062X(D3)/ma-def2QZVP method, the best results were obtained with models containing two H<sub>2</sub>O molecules. H<sub>2</sub>O and R-PhNH<sup>•</sup> were constructed according to the NHHO structure, while H<sub>2</sub>O and R-PhNH<sub>2</sub><sup>•+</sup> were built based on the 2HO structure (NHHO-2HO, see Figure 1 for these structures). In this case, we obtained RMSE<sub>aver</sub> ± RMSE<sub>std</sub> = 1.113 ± 0.616. The high ratios of RMSE<sub>std</sub> to RMSE<sub>aver</sub> (~55%) suggest that the present models with optimized numbers and positions of H<sub>2</sub>O molecules could still be significantly modified to increase calculation accuracy.

Regarding R-PhNH<sub>3</sub><sup>+</sup> (Figure 3b), the lowest RMSE was obtained with the CBS-QB3 method with two H<sub>2</sub>O molecules. H<sub>2</sub>O and R-PhNH<sub>2</sub> were constructed according to the 2HO structure, while H<sub>2</sub>O and R-PhNH<sub>3</sub><sup>+</sup> were built based on the HOA90 structure (2HO-HOA90, see Figure 1 for these structures). In the case of the M062X(D3)/ma-def2QZVP method, the best model includes three H<sub>2</sub>O molecules positioned in the 2HONH structure for R-PhNH<sub>2</sub> and in the 3HO structure for R-PhNH<sub>3</sub><sup>+</sup> (2HONH-3HO). However, the RMSE values obtained from all models based on the M062X(D3)/ma-def2QZVP method exceeded 1.

### 2.5. Combination of Optimal Models and Methods for pK<sub>a</sub> Calculation

To improve the pK<sub>a</sub> calculation accuracy based on the current methods and basis sets, three methodological aspects were considered: (i) choosing appropriate approaches; (ii) employing suitable correction methods; and (iii) selecting proper protocols for calculating solvation energies (only for the indirect approach). To determine the best combinations of optimal models and methods, the RMSE values obtained from different combinations, based on the established optimized models, are summarized and compared in Figure 4.



**Figure 4.** The values of RMSE were calculated for different combinations of calculation approaches (direct and indirect), correction methods (C0: rigid-rotor harmonic oscillator method; C1,C2: two different quasi-rigid-rotor harmonic oscillator methods; see Section 3.1), and protocols for calculating solvation energy (P4–P6, see calculation methods in Section 3.1.2). Figures (a,b) present the results for R-PhNH<sub>2</sub><sup>•+</sup>, while figures (c,d) depict the results for R-PhNH<sub>3</sub><sup>+</sup>. The arrows indicate the combination of specific approaches and methods that achieved the highest accuracy in each case (lowest RMSE).

Overall, for both R-PhNH<sub>2</sub><sup>•+</sup> and R-PhNH<sub>3</sub><sup>+</sup>, the CBS-QB3 method emerged as the most accurate choice, which has been widely suggested for pK<sub>a</sub> calculations in several reports [16,61]. For R-PhNH<sub>2</sub><sup>•+</sup>, the quasi-rigid-rotor harmonic oscillator (QRRHO) correction method (C1 and C2, see Section 3.1) effectively improved calculation results, with C1 slightly outperforming C2 in the case of the M062X method. For calculations through the indirect approach, the choice of an appropriate protocol for calculating solvation energy played a more essential role in improving accuracy than the correction method. For R-PhNH<sub>2</sub><sup>•+</sup>, M062X/6-31G(d) (P4) produced the least accurate results, whereas for R-PhNH<sub>3</sub><sup>+</sup>, M052X/cc-pvtz (P5) was the least effective. In most cases, ωb97xd/6-31+g(d,p) (P6) delivered the best performance.

Once the correction method (C1) and protocol for calculating solvation energy (ωb97xd/6-31+g(d,p) in most cases, i.e., P6), were determined, the calculation errors were significantly minimized, regardless of whether the direct or indirect approach was adopted. Specifically, for R-PhNH<sub>2</sub><sup>•+</sup>, the best combination was the indirect approach/CBS-QB3/C1/P6 based on models with one H<sub>2</sub>O molecule (both R-PhNH<sup>•</sup> and R-PhNH<sub>2</sub><sup>•+</sup> in the HO structure), resulting in an RMSE of 0.563. For calculating pK<sub>a</sub> of R-PhNH<sub>2</sub><sup>•+</sup> containing heavy atoms, the best combination was the direct approach/M062X(D3)/ma-def2QZVP/C1 based on models with two H<sub>2</sub>O molecules (R-PhNH<sup>•</sup> in the NHHO structure and R-PhNH<sub>2</sub><sup>•+</sup> in the 2HO structure), which gave an RMSE of 0.384.

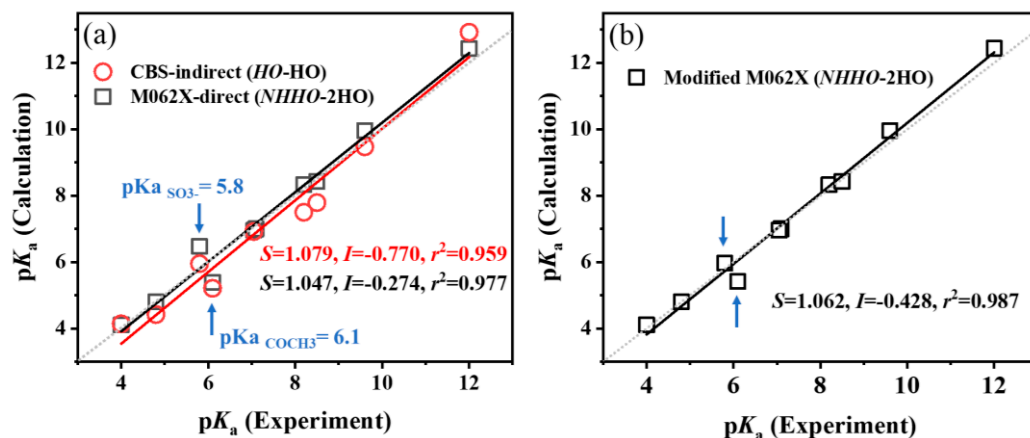
These results suggest that precise aqueous pK<sub>a</sub> predictions would be achieved for R-PhNH<sub>2</sub><sup>•+</sup> based on either the CBS-QB3 or the M062X(D3)/ma-def2QZVP method. From a practical standpoint, the direct approach combined with the M062X(D3)/ma-def2QZVP method holds advantages as it predicts the pK<sub>a</sub> of molecules containing heavy atoms without additional steps for calculating solvation energy. However, it is imperative to note that

the indirect approach utilizing CBS-QB3/C1/P6 keeps the calculation time within merely 94 min (H-PhNH<sub>2</sub><sup>•+</sup> as a typical example), while the direct approach with M062X(D3)/ma-def2QZVP/C1 requires a notably longer calculation time of 626 min (Figure S4a).

For R-PhNH<sub>3</sub><sup>+</sup>, the optimal combination of methods was the direct approach/CBS-QB3/C1 based on the model with two H<sub>2</sub>O molecules (H<sub>2</sub>O and R-PhNH<sub>2</sub> were constructed according to the 2HO structure, while H<sub>2</sub>O and R-PhNH<sub>3</sub><sup>+</sup> were built based on the HOA90 structure), resulting in an RMSE of 0.349 that is comparable to the RMSE of R-PhNH<sub>2</sub><sup>•+</sup> (0.563). Moreover, the RMSE obtained with the M062X(D3)/ma-def2QZVP method was 1.375 (direct approach/C1, models with three H<sub>2</sub>O molecules), indicating an inability to accurately predict the pK<sub>a</sub> values of molecules containing heavy atoms. These results suggest the pK<sub>a</sub> calculation is more difficult for R-PhNH<sub>3</sub><sup>+</sup> than for R-PhNH<sub>2</sub><sup>•+</sup>. New strategies will thus be explored to predict the pK<sub>a</sub> of R-PhNH<sub>3</sub><sup>+</sup>, rather than just selecting and modifying the current calculation methods that proved satisfactory for R-PhNH<sub>2</sub><sup>•+</sup>. Moreover, the calculation time for the direct approach employing CBS-QB3/C1 is relatively short, taking only 69 min (H-PhNH<sub>3</sub><sup>+</sup> as a typical example), whereas the direct approach with M062X(D3)/ma-def2QZVP/C1 requires a significantly longer duration of 501 min (Figure S4b).

## 2.6. Strategies for Further Improvements of pK<sub>a</sub> Calculations

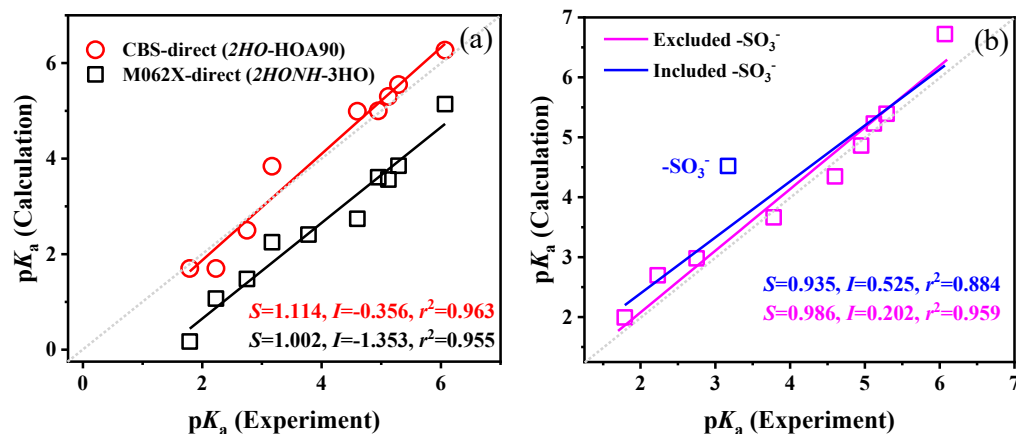
Adopting a mixed calculation strategy can potentially lead to significant improvements in the accuracy of pK<sub>a</sub> calculations. Instead of employing the same strategy for all compounds, adopting different strategies for specific molecular structures might yield better overall results [20]. To explore this possibility and identify potentially specific structures among the ten compounds we investigated, a comparison was made between the experimental and calculated pK<sub>a</sub> values in the aqueous phase, the latter obtained from the best combination of models and methods for R-PhNH<sub>2</sub><sup>•+</sup> and R-PhNH<sub>3</sub><sup>+</sup> (Figures 5 and 6).



**Figure 5.** Validation of the performance of different combinations of models and methods on pK<sub>a</sub> calculations of R-PhNH<sub>2</sub><sup>•+</sup>. Calculated pK<sub>a</sub> values (a) before and (b) after modifications. The dotted lines indicate the ideal results with S = 1 and I = 0. Arrows indicate the models modified by adding extra H<sub>2</sub>O molecules.

For R-PhNH<sub>2</sub><sup>•+</sup>, both the CBS-QB3 and M062X(D3)/ma-def2QZVP methods produced regression lines with slopes (S) close to 1, intercepts (I) close to 0, and correlation coefficients (r<sup>2</sup>) > 0.959 (Figure 5a). Therefore, both regression lines were precise for calculating the pK<sub>a</sub> values of R-PhNH<sub>2</sub><sup>•+</sup>. However, upon closer examination, it was revealed that the largest errors in the linear regressions originated from the pK<sub>a</sub> values associated with compounds containing negatively charged =O groups, like COCH<sub>3</sub>-PhNH<sub>2</sub><sup>•+</sup> and SO<sub>3</sub><sup>-</sup>-PhNH<sub>2</sub><sup>•+</sup> (Figure 5a and Table S5).





**Figure 6.** Validation of the performance of different combinations of models and methods on  $\text{pK}_a$  calculations of  $\text{R-PhNH}_3^+$ . Calculated  $\text{pK}_a$  values were based on (a) CBS-QB3 with two  $\text{H}_2\text{O}$  molecules in models and M062X(D3)/ma-def2QZVP with three  $\text{H}_2\text{O}$  molecules in models; (b) revDSD-PBEP86-D3(BJ)/ma-def2QZVPP method with three  $\text{H}_2\text{O}$  molecules around amino and extra  $\text{H}_2\text{O}$  near R group for formation of hydrogen bonds (correction method: C0). The dotted lines indicate the ideal results with  $S = 1$  and  $I = 0$ .

It is common to introduce explicit  $\text{H}_2\text{O}$  molecules near the negative and positive charged atoms of the molecules to form hydrogen bonds and improve the calculation accuracy when studying molecules in the aqueous phase [62–64]. The purpose of adding  $\text{H}_2\text{O}$  molecules is to achieve a smooth transition between solute molecules and the implicit solvent depicted by the solvent model. This is also the main reason why placing  $\text{H}_2\text{O}$  molecules near the (de)ionization groups would improve the accuracy of  $\text{pK}_a$  calculations as observed above. However, the improvement of  $\text{pK}_a$  calculations by introducing  $\text{H}_2\text{O}$  molecules at positions beyond the (de)protonation groups of the target molecules has seldom been considered and discussed in previous reports. In this study, an additional  $\text{H}_2\text{O}$  molecule was further introduced near the  $=\text{O}$  atom of  $\text{COCH}_3\text{-PhNH}^\bullet$ ,  $\text{COCH}_3\text{-PhNH}_2^{\bullet+}$ ,  $\text{SO}_3^- \text{-PhNH}^\bullet$ , and  $\text{SO}_3^- \text{-PhNH}_2^{\bullet+}$  to better explore the impact of explicit  $\text{H}_2\text{O}$  molecules on  $\text{pK}_a$  calculations (Figure S5).

Generally, the modifications introduced specifically for these two molecules improved the  $\text{pK}_a$  calculation accuracy only when using the M062X(D3)/ma-def2QZVP method (Figure 5b). The RMSE was 0.310 (0.384 before modification),  $\Delta\text{pK}_a$  M062X ranged from 0.003 to 0.685 (average: 0.215), and  $r^2$  was 0.987 (0.977 before modification).

For  $\text{R-PhNH}_3^+$  (Figure 6a), the CBS-QB3 method yielded a regression line with a slope ( $S$ ) of 1.114 and an intercept ( $I$ ) of  $-0.356$ , enabling direct calculations of  $\text{pK}_a$ . The  $\Delta\text{pK}_a$  CBS value ranged from 0.119 to 1.042 (average: 0.486). By comparison with a previous study ( $\Delta\text{pK}_a$  CBS ranged from 0.42 to 2.90, and the average was 1.52) [20], the results obtained herein were more accurate. The differences in  $\Delta\text{pK}_a$  CBS were mainly influenced by the number of included  $\text{H}_2\text{O}$  molecules in the models. While the previous study adopted only one  $\text{H}_2\text{O}$  molecule, we incorporated two  $\text{H}_2\text{O}$  molecules and significantly improved the accuracy of  $\text{pK}_a$  calculations for  $\text{R-PhNH}_3^+$ . These findings further highlight the importance of including  $\text{H}_2\text{O}$  molecules to increase the accuracy of the  $\text{pK}_a$  calculations for  $\text{R-PhNH}_3^+$  in aqueous solution.

On the other hand, the regression line obtained by the M062X(D3)/ma-def2QZVP method had a slope ( $S$ ) close to 1, indicating a good linear relationship. However, the intercept ( $I$ ) was quite large ( $I = -1.353$ ), deviating notably from the theoretical value of 0, and the  $\Delta\text{pK}_a$  M062X values ranged from 0.922 to 1.856 (average: 1.346). These findings suggest that the errors associated with the M062X(D3)/ma-def2QZVP method appear to be systematic. It is assumed that employing a more accurate method could eliminate the deviations of the intercept ( $I$ ) by increasing all the calculated  $\text{pK}_a$  values and shifting the regression line in the positive direction of the y-axis (Figure 6a).

The same strategy of adding extra H<sub>2</sub>O molecules to the substituent groups -COCH<sub>3</sub> and -SO<sub>3</sub><sup>-</sup> was also employed to calculate the pK<sub>a</sub> values of R-PhNH<sub>3</sub><sup>+</sup>, based on the CBS-QB3 method. Moreover, the M062X(D3)/ma-def2QZVP method was replaced with the more accurate double-hybrid functional method, revDSD-PBEP86-D3(BJ), coupled with the ma-def2QZVPP basis, with the main aim of eliminating systematic errors. The calculated pK<sub>a</sub> values of R-PhNH<sub>3</sub><sup>+</sup> after incorporating these modifications are summarized in Table S7 and Figure S6.

For the pK<sub>a</sub> values obtained by the CBS-QB3 method, the addition of H<sub>2</sub>O near the -COCH<sub>3</sub> and -SO<sub>3</sub><sup>-</sup> groups failed to notably enhance, and in the case of -SO<sub>3</sub><sup>-</sup> it even decreased the accuracy of calculations (RMSE increased from 0.349 to 0.371, and *r*<sup>2</sup> of the linear fitting decreased from 0.963 to 0.865). These results suggest that adding extra H<sub>2</sub>O molecules may not always improve pK<sub>a</sub> calculations, as observed in previous studies [65]. On the other hand, the revDSD-PBEP86-D3(BJ)/ma-def2QZVPP method significantly increased the value of the intercept (*I*) from -1.353 to 0.521, while the slope remained close to 1 (*S* = 1.029), which suggests that systematic errors were minimized in this way. Additionally, the strategy of adding extra H<sub>2</sub>O molecules near the R group was further adopted when using the revDSD-PBEP86-D3(BJ)/ma-def2QZVPP method. H<sub>2</sub>O molecules were added near R groups that could form hydrogen bonds with H<sub>2</sub>O, such as negative O atoms (-COCH<sub>3</sub>, -SO<sub>3</sub><sup>-</sup>, and -OCH<sub>3</sub>), halogen atoms (F and I), and positively charged H in -NH<sub>2</sub>. The calculated pK<sub>a</sub> values are summarized in Table S7, and the modified regression line is shown in Figure 6b. The linear regression analysis yielded a line with *S* = 0.935, *I* = 0.525, and *r*<sup>2</sup> = 0.884, with an average ΔpK<sub>a</sub> value of 0.505. However, the pK<sub>a</sub> of one compound (-SO<sub>3</sub><sup>-</sup>) deviated significantly from the regression line, with a ΔpK<sub>a</sub> value of 1.351. The addition of H<sub>2</sub>O even decreased the calculation accuracy of SO<sub>3</sub><sup>-</sup>-PhNH<sub>3</sub><sup>+</sup>. These findings suggest that the revDSD-PBEP86-D3(BJ)/ma-def2QZVPP method may not be suitable for R groups with weak polarity or in anionic form. By excluding SO<sub>3</sub><sup>-</sup>-PhNH<sub>3</sub><sup>+</sup>, the calculation accuracy was significantly improved, with ΔpK<sub>a</sub> values in the range of 0.089–0.651 (average of 0.247) and a regression line with *S* = 0.986, *I* = 0.202, and *r*<sup>2</sup> = 0.959.

Although the adoption of revDSD-PBEP86-D3(BJ)/ma-def2QZVPP and extra explicit water molecules near R would improve the pK<sub>a</sub> calculation for most R-PhNH<sub>3</sub><sup>+</sup> compounds, this method is computationally quite expensive and time-consuming. Compared with M062X(D3)/ma-def2QZVP (H-PhNH<sub>3</sub><sup>+</sup> as a typical example, 501 min), the calculation time for this alternative method was notably longer (1404 min, as illustrated in Figure S4b). Therefore, unless necessary, such as when dealing with molecules containing heavy atoms, we do not recommend revDSD-PBEP86-D3(BJ)/ma-def2QZVPP for pK<sub>a</sub> calculations of compounds with high molecular weight.

### 3. Methods

Quantum chemical calculations were carried out with a computer equipped with 2 AMD EPYC 7R32 CPUs (96 cores, 192 threads total across dual CPUs, maximum clock speed of 3.3 GHz), Gigabyte MZ72-HBO motherboard with integrated graphics card, and sixteen Micron DDR4 3200 RECC random access memory cards (32 GB each).

#### 3.1. Calculation of Δ*G*<sub>deprot</sub>(sol)

##### 3.1.1. Direct Approach

###### (1) Methods and basis sets

The free energies of the deprotonated and protonated states of molecules were directly calculated based on the SMD model, and Δ*G*<sub>deprot</sub>(sol) was determined according to Equation (2) [15,16].

$$\Delta G_{\text{deprot}}(\text{sol}) = G_{\text{A}^-}(\text{sol}) - G_{\text{HA}}(\text{sol}) + \Delta G_{\text{H}^+}(\text{sol}) \quad (2)$$

where  $G_{A^-}(\text{sol})$  and  $G_{\text{HA}}(\text{sol})$  refer to the free energies in water of the deprotonated ( $A^-$ ) and protonated states (HA) of the studied compounds, respectively.  $\Delta G_{\text{H}^+}(\text{sol})$  is the free energy of  $\text{H}^+$ , which has been experimentally determined as  $-1131.44 \text{ kJ mol}^{-1}$  [4].

## (2) Corrections of $G(\text{sol})$

Corrections of  $G(\text{sol})$  were obtained based on scale factors for the Zero Point Energy (ZPE) [66], taking into account the harmonic approximation. All corrections were performed using the Shermo program (version: 2.3.5) [67]. The program incorporates three models of harmonic approximation, including the rigid-rotor harmonic oscillator (RRHO) method (labeled as C0) and two quasi-RRHO (QRRHO) methods. The first QRRHO method (QRRHO1, labeled as C1) was suggested by Truhlar, artificially raising all frequencies lower than  $100 \text{ cm}^{-1}$  to  $100 \text{ cm}^{-1}$  [68]. The second QRRHO method (QRRHO2, labeled as C2) was suggested by Grimme and was based on an interpolation method [67,69].

### 3.1.2. Indirect (Thermodynamic) Approach

#### (1) Methods and basis sets

The indirect approach is based on thermodynamic cycles for calculating the value of  $G(\text{sol})$  (Equation (3)) [15,16].

$$G(\text{sol}) = G(\text{gas}) + \Delta G(\text{sol}) \quad (3)$$

where  $G(\text{gas})$  and  $\Delta G(\text{sol})$  represent, respectively, gas-phase free energy and solvation free energy of a specific compound. These two values were calculated separately using different methods and basis sets. Specifically, for compounds containing heavy atoms,  $G(\text{gas})$  was calculated using the M062X(D3)/ma-def2QZVP method [16,52,57,58]. Additionally,  $\Delta G(\text{sol})$  was calculated based on six different protocols, with and without explicit water molecules (Table 2). Then, by combining the results of  $G(\text{gas})$  and  $\Delta G(\text{sol})$ ,  $\Delta G_{\text{deprot}}(\text{sol})$  was determined according to Equations (4) and (5).

$$\Delta G_{\text{deprot}}(\text{sol}) = \Delta G(\text{gas}) + \Delta \Delta G(\text{sol}) + \Delta G_{\text{H}^+}(\text{sol}) \quad (4)$$

$$\Delta G_{\text{deprot}}(\text{sol}) = [G_{A^-}(\text{gas}) - G_{\text{HA}}(\text{gas})] + [\Delta G_{A^-}(\text{sol}) - \Delta G_{\text{HA}}(\text{sol})] + \Delta G_{\text{H}^+}(\text{sol}) \quad (5)$$

## (2) Corrections of $G(\text{gas})$

Corrections to the values of  $G(\text{gas})$  were carried out with Shermo software (version: 2.3.5), as mentioned above [48–50].

**Table 2.** Protocols for calculating  $\Delta \Delta G_{\text{HA}}(\text{sol})$  ( $\Delta G_{A^-}(\text{sol}) - \Delta G_{\text{HA}}(\text{sol})$ ), with and without explicit water molecules.

No.	Models	Methods <sup>a</sup>	Ref.
P1 <sup>b</sup>	Implicit (single)	M052X/6-31G(d)	[14]
P2 <sup>c</sup>	Implicit (mixed)	M052X/6-31+G(d,p) for neutral, e.g., R-PhNH <sup>•</sup> and R-PhNH <sub>2</sub>	[14]
		M052X/6-31+G(d,p) for cation, e.g., R-PhNH <sub>2</sub> <sup>•+</sup> and R-PhNH <sub>3</sub> <sup>+</sup>	
P3 <sup>e</sup>	Implicit (mixed)	HF/6-31G(d) for anion, e.g., SO <sub>3</sub> <sup>-</sup> -PhNH <sup>•</sup> and SO <sub>3</sub> <sup>-</sup> -PhNH <sub>2</sub>	-
		<sup>d</sup> HF/6-31G(d) for SO <sub>3</sub> <sup>-</sup> -PhNH <sup>•+</sup> and SO <sub>3</sub> <sup>-</sup> -PhNH <sub>3</sub> <sup>+</sup>	
P4	Explicit	M062X/6-31G(d)	[70]

Table 2. Cont.

No.	Models	Methods <sup>a</sup>	Ref.
P5	Explicit	M052X/cc-pVTZ	[71]
P6	Explicit	$\omega$ b97xd/6-31+g(d,p)	[17]

<sup>a</sup>: Compounds containing iodine atoms were calculated based on the ma-def2QZVP basis set [57,58]. <sup>b</sup>: M052X/6-31G(d) was the best overall average performer for aqueous solvation energies [14]. <sup>c</sup>: M052X/6-31G(d), M052X/6-31+G(d,p), and HF/6-31G(d) were the best for neutrals, cations, and anions, respectively [14]. <sup>d</sup>: Since  $-\text{SO}_3^-$  carries a negative charge, molecules in neutral and cationic form containing this group were also calculated using HF/6-31G(d), which worked best for anions [14]. Finally, the closest value to the experimental  $\text{pK}_a$  was selected. <sup>e</sup>: As  $\text{R-PhNH}_2^{\bullet+}$  is not a typically stable cation as those parameterized in the original reference [14], Protocol-3 was introduced as an imitation of Protocol-2 (P2).  $\text{R-PhNH}_2^{\bullet+}$  and  $\text{R-PhNH}^\bullet$  were calculated with M052X/6-31G(d) and HF/6-31G(d), respectively.

### 3.1.3. Data Processing and Advanced Modification Method

The root mean square error (RMSE, calculated according to Equation (6) where  $n$  indicates the total number of calculated molecules) and linear regression analysis were basically used as standards to select and improve appropriate models and combinations of calculation methods/parameters. If the attempts to achieve higher accuracy using CBS-QB3 and M062X(D3)/ma-def2QZVP methods did not succeed (RMSE exceeded 1 or the correlation coefficient  $r^2 < 0.9$ ), the doubly hybrid functional revDSD-PBEP86-D3(BJ) in conjunction with ma-def2QZVPP was further employed to improve the accuracy of the calculations [57,58,72].

$$\text{RMSE} = \sqrt{\frac{(\text{pK}_{a\text{Calculation}} - \text{pK}_{a\text{Experiment}})^2}{n}} \quad (6)$$

## 4. Conclusions

In this study, we constructed various molecular models by altering the number and positions of  $\text{H}_2\text{O}$  molecules towards  $\text{pK}_a$  calculations for  $\text{R-PhNH}_2^{\bullet+}$  and  $\text{R-PhNH}_3^+$  in the aqueous phase. Then, based on the appropriate models, we evaluated the performance of direct and indirect approaches, CBS-QB3 and M062X(D3)/ma-def2QZVP methods, along with methods for correcting free energy and calculating solvation energy.

For  $\text{R-PhNH}_2^{\bullet+}$ , both CBS-QB3 and M062X(D3)/ma-def2QZVP produced accurate  $\text{pK}_a$  values using models with one or two  $\text{H}_2\text{O}$  molecules. Additional  $\text{H}_2\text{O}$  near oxygen-containing substituents further improved the results. CBS-QB3 performed best with the indirect approach, C1 correction method, and  $\omega$ b97xd/6-31+g(d,p) method for solvation energy calculation, resulting in an average  $\Delta\text{pK}_a = 0.376$ . M062X(D3)/ma-def2QZVP performed best with the direct approach and C1 correction method, yielding an average  $\Delta\text{pK}_a = 0.215$ . These methods highlight a potential to predict  $\text{pK}_a$  values of other environmentally important radical cations, such as protonated phenoxyl radicals.

Predicting  $\text{pK}_a$  values for  $\text{R-PhNH}_3^+$  was more challenging compared to  $\text{R-PhNH}_2^{\bullet+}$ . CBS-QB3 resulted in an average  $\Delta\text{pK}_a = 0.486$  with two  $\text{H}_2\text{O}$  molecules in the model, while this method was useless for molecules with heavy atoms. M062X(D3)/ma-def2QZVP failed to accurately calculate  $\text{pK}_a$  values, with an average  $\Delta\text{pK}_a = 1.351$  based on models with three  $\text{H}_2\text{O}$  molecules. Although the adoption of the revDSD-PBEP86-D3/ma-def2QZVPP method and the inclusion of extra  $\text{H}_2\text{O}$  near the R group improved the results (average  $\Delta\text{pK}_a = 0.357$ ), it is worth noting that the computational time requirements limit the method's application when dealing with compounds having high molecular weight.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/molecules29194522/s1>, Table S1: The calculated energies of  $\text{R-PhNH}^\bullet$  ( $\text{A}^-$ ) and  $\text{R-PhNH}_2^{\bullet+}$  (AH) based on direct method (unit: a.u.); Table S2: The calculated energies of  $\text{R-PhNH}^\bullet$  ( $\text{A}^-$ ) and  $\text{R-PhNH}_2^{\bullet+}$  (AH) based on indirect method (unit: a.u.); Table S3: The calculated energies of  $\text{R-PhNH}_2$  ( $\text{A}^-$ ) and  $\text{R-PhNH}_3^+$  (AH) based on direct method (unit: a.u.); Table S4: The calculated energies of  $\text{R-PhNH}_2$  ( $\text{A}^-$ ) and  $\text{R-PhNH}_3^+$  (AH) based on indirect method (unit: a.u.);

Table S5: The experimental and calculated  $pK_a$  values of  $\text{PhNH}_2^{\bullet+}$ ; Table S6: The experimental and calculated  $pK_a$  values of  $\text{PhNH}_3^+$ ; Table S7:  $pK_a$  values of  $\text{R-PhNH}_3^+$  calculated by revDSD-PBEP86-D3(BJ)/ma-def2QZVPP with models included three  $\text{H}_2\text{O}$  combined with amino group; Figure S1: The potential positions of  $\text{H}_2\text{O}$  molecules in the models.  $\text{H-PhNH}^\bullet$ ,  $\text{H-PhNH}_2^{\bullet+}$ ,  $\text{H-PhNH}_2$  and  $\text{H-PhNH}_3^+$  are presented as examples; Figure S2: Comparison of  $\text{H-PhNH}_2$  models with three  $\text{H}_2\text{O}$  molecules optimized by M062X(D3)/6-311++g(d,p) and calculated based on CBS-QB3; Figure S3: The changes in molecular structures after the calculation of  $G(\text{gas})$  and  $G(\text{sol})$  based on the CBS-QB3 method; Figure S4: The calculation time for (a)  $\text{H-PhNH}_2^{\bullet+}$  (indirect approach/CBS-QB3/C1/P6/1H<sub>2</sub>O and direct approach/M062X(D3)/ma-def2QZVP/C1/2H<sub>2</sub>O), and (b)  $\text{H-PhNH}_3^+$  (direct approach/CBS-QB3/C1/2H<sub>2</sub>O, direct approach/M062X(D3)/ma-def2QZVP/C1/3H<sub>2</sub>O, and direct approach/M062X(D3)/revDSD-PBEP86-D3(BJ)/ma-def2QZVPP/3H<sub>2</sub>O). The number of processors is 20; Figure S5: The optimized structures of (a)  $\text{COCH}_3\text{-PhNH}^\bullet$ , (b)  $\text{COCH}_3\text{-PhNH}_2^{\bullet+}$ , (c)  $\text{SO}_3^-\text{-PhNH}^\bullet$  and (d)  $\text{SO}_3^-\text{-PhNH}_2^{\bullet+}$  with an additional  $\text{H}_2\text{O}$  near =O; Figure S6: The performance of modified CBS-QB3 and revDSD-PBEP86-D3(BJ)/ma-def2QZVPP methods on  $pK_a$  calculations of  $\text{R-PhNH}_3^+$ .

**Author Contributions:** Conceptualization, D.V.; methodology, H.F., J.W. and H.H.; software, J.W. and W.Z.; validation, Z.Z. and X.L.; data curation, Y.Y. and J.W.; writing—original draft preparation, H.F. and D.V.; writing—review and editing, D.V.; supervision, D.V.; funding acquisition, H.F. and D.V. All authors have read and agreed to the published version of the manuscript.

**Funding:** J.W., H.F., Z.Z., H.H., X.L., Y.Y., and W.Z. gratefully acknowledge financial support from the National Natural Science Foundation of China (No. 42367062), the Natural Science Foundation of Jiangxi Province (No. 20212BAB203020, No. 20212BAB213017), the China Scholarship Council (No. 202208360073), the Special Innovation Fund for Graduate Students of Jiangxi Province (No. YC2023-S361). DV acknowledges support from the Project CH4.0, under the MUR program “Dipartimenti di Eccellenza 2023-2027” (CUP: D13C22003520001).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Data are contained within the article or Supplementary Materials.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Fang, H.S.; Gao, Y.P.; Wang, H.H.; Yin, H.; Li, G.Y.; An, T.C. Photo-induced oxidative damage to dissolved free amino acids by the photosensitizer polycyclic musk tonalide: Transformation kinetics and mechanisms. *Water Res.* **2017**, *115*, 339–346. [[CrossRef](#)] [[PubMed](#)]
2. Kolthoff, I.M.; Chantooni, M.K., Jr. Calibration of the glass electrode in acetonitrile. Shape of potentiometric titration curves. Dissociation constant of picric acid1. *J. Am. Chem. Soc.* **1965**, *87*, 4428–4436. [[CrossRef](#)]
3. Zalipsky, J.J.; Patel, D.M.; Darnowski, R.J.; Reaveycantwell, N.H.  $pK_a$  determination of methaqualone. *J. Pharm. Sci.* **1976**, *65*, 460–461. [[CrossRef](#)] [[PubMed](#)]
4. Thapa, B.; Raghavachari, K. Accurate  $pK_a$  evaluations for complex bio-organic molecules in aqueous media. *J. Chem. Theory Comput.* **2019**, *15*, 6025–6035. [[CrossRef](#)]
5. Ruano, G.; Pedano, M.L.; Albornoz, M.; Fuhr, J.D.; Martiarena, M.L.; Zampieri, G. Deprotonation of the amine group of glyphosate studied by xps and dft. *Appl. Surf. Sci.* **2021**, *567*, 150753. [[CrossRef](#)]
6. Zuo, W.L.; Li, N.; Chen, B.; Zhang, C.; Li, Q.; Yan, M. Investigation of the deprotonation of tetracycline using differential absorbance spectra: A comparative experimental and dft/td-dft study. *Sci. Total Environ.* **2020**, *726*, 138432. [[CrossRef](#)] [[PubMed](#)]
7. Jonsson, M.; Lind, J.; Eriksen, T.E.; Merenyi, G. Redox and acidity properties of 4-substituted aniline radical cations in water. *J. Am. Chem. Soc.* **1994**, *116*, 1423–1427. [[CrossRef](#)]
8. Qian, T.T.; Kun, L.; Gao, B.; Zhu, R.; Wu, X.; Wang, S. Photo-ionization and photo-excitation of curcumin investigated by laser flash photolysis. *Spectrochim. Acta A* **2013**, *116*, 6–12. [[CrossRef](#)]
9. Lang, B.; Mosquera-Vázquez, S.; Lovy, D.; Sherin, P.; Markovic, V.; Vauthey, E. Broadband ultraviolet-visible transient absorption spectroscopy in the nanosecond to microsecond time domain with sub-nanosecond time resolution. *Rev. Sci. Instrum.* **2013**, *85*, 73107. [[CrossRef](#)]
10. Gangarapu, S.; Marcelis Antonius, T.M.; Zuilhof, H. Accurate  $pK_a$  calculation of the conjugate acids of alkanolamines, alkaloids and nucleotide bases by quantum chemical methods. *Chem. Phys. Chem.* **2013**, *14*, 990–995. [[CrossRef](#)]

11. Pracht, P.; Wilcken, R.; Udvarhelyi, A.; Rodde, S.; Grimme, S. High accuracy quantum-chemistry-based calculation and blind prediction of macroscopic pKa values in the context of the sampl6 challenge. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 1139–1149. [[CrossRef](#)] [[PubMed](#)]
12. Guerard, J.J.; Arey, J.S. Critical evaluation of implicit solvent models for predicting aqueous oxidation potentials of neutral organic compounds. *J. Chem. Theory Comput.* **2013**, *9*, 5046–5058. [[CrossRef](#)] [[PubMed](#)]
13. Klamt, A. The cosmo and cosmo-rs solvation models. *WIREs Comput. Mol. Sci.* **2018**, *8*, 1338. [[CrossRef](#)]
14. Marenich, V.A.; Cramer, J.C.; Truhlar, D.G. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J. Phys. Chem. B* **2009**, *113*, 6378–6396. [[CrossRef](#)]
15. Ho, J.; Ertem, M.Z. Calculating free energy changes in continuum solvation models. *J. Phys. Chem. B* **2016**, *120*, 1319–1329. [[CrossRef](#)]
16. Xu, T.; Chen, J.; Chen, X.; Xie, H.; Wang, Z.; Xia, D.; Tang, W.; Xie, H. Prediction models on pKa and base-catalyzed hydrolysis kinetics of parabens: Experimental and quantum chemical studies. *Environ. Sci. Technol.* **2021**, *55*, 6022–6031. [[CrossRef](#)] [[PubMed](#)]
17. Thapa, B.; Schlegel, H.B. Density functional theory calculation of pKa's of thiols in aqueous solution using explicit water molecules and the polarizable continuum model. *J. Phys. Chem. A* **2016**, *120*, 5726–5735. [[CrossRef](#)]
18. Haworth, N.L.; Wang, Q.; Coote, M.L. Modeling flexible molecules in solution: A pKa case study. *J. Phys. Chem. A* **2017**, *121*, 5217–5225. [[CrossRef](#)] [[PubMed](#)]
19. Hebert, S.P.; Schlegel, H.B. Computational study of the pH-dependent competition between carbonate and thymine addition to the guanine radical. *Chem. Res. Toxicol.* **2019**, *32*, 195–210. [[CrossRef](#)]
20. Xu, L.K.; Coote, M.L. Methods to improve the calculations of solvation model density solvation free energies and associated aqueous pKa values: Comparison between choosing an optimal theoretical level, solute cavity scaling, and using explicit solvent molecules. *J. Phys. Chem. A* **2019**, *123*, 7430–7438. [[CrossRef](#)]
21. Xu, L.K.; Coote, M.L. Improving the accuracy of pcm-uahf and pcm-uaks calculations using optimized electrostatic scaling factors. *J. Chem. Theory Comput.* **2019**, *15*, 6958–6967. [[CrossRef](#)] [[PubMed](#)]
22. Barone, V.; Carnimeo, I.; Scalmani, G. Computational spectroscopy of large systems in solution: The dftb/pcm and td-dftb/pcm approach. *J. Chem. Theory Comput.* **2013**, *9*, 2052–2071. [[CrossRef](#)]
23. Provorse, L.M.R.; Isborn, C.M. Combining explicit quantum solvent with a polarizable continuum model. *J. Phys. Chem. B* **2017**, *121*, 10105–10117. [[CrossRef](#)] [[PubMed](#)]
24. Behjatmanesh-Ardakani, R.; Karimi, M.A.; Ebady, A. Cavity shape effect on pKa prediction of small amines. *J. Mol. Struct. THEOCHEM* **2009**, *910*, 99–103. [[CrossRef](#)]
25. Vyboishchikov, S.F.; Voityuk, A.A. Fast non-iterative calculation of solvation energies for water and non-aqueous solvents. *J. Comput. Chem.* **2021**, *42*, 1184–1194. [[CrossRef](#)]
26. Trovó, A.G.; Nogueira, R.F.P.; Agüera, A.; Sirtori, C.; Fernández-Alba, A.R. Photodegradation of sulfamethoxazole in various aqueous media: Persistence, toxicity and photoproducts assessment. *Chemosphere* **2009**, *77*, 1292–1298. [[CrossRef](#)]
27. Shi, C.; Qin, L.; Wu, S.; Kang, S.; Li, X. Highly sensitive sers detection and photocatalytic degradation of 4-aminothiophenol by a cost-effective cobalt metal-organic framework-based sandwich-like sheet. *Chem. Eng. J.* **2021**, *422*, 129970. [[CrossRef](#)]
28. Zhang, C.J.; Chen, H.; Xue, G.; Liu, Y.; Chen, S.; Jia, C. A critical review of the aniline transformation fate in azo dye wastewater treatment. *J. Cleaner Prod.* **2021**, *321*, 128971. [[CrossRef](#)]
29. İhsan, H.M.; İnce, U.; Gündüz, M.G.; Coşkun, G.P.; Birgül, K.; Doğan, Ş.D.; Küçükgülzel, Ş.G. Synthesis, antimicrobial evaluation and molecular modeling studies of novel thiosemicarbazides/semicarbazides derived from p-aminobenzoic acid. *J. Mol. Struct.* **2022**, *1261*, 132907. [[CrossRef](#)]
30. Wang, D.; Lan, Y.; Chen, W.; Han, X.; Liu, S.; Cao, D.; Cheng, X.; Wang, Q.; Zhan, Z.; He, W. The six-year biochar retention interacted with fertilizer addition alters the soil organic nitrogen supply capacity in bulk and rhizosphere soil. *J. Environ. Manag.* **2023**, *338*, 117757. [[CrossRef](#)]
31. Tentscher, P.R.; Eustis, S.N.; McNeill, K.; Arey, J.S. Aqueous oxidation of sulfonamide antibiotics: Aromatic nucleophilic substitution of an aniline radical cation. *Chem.-Eur. J.* **2013**, *19*, 11216–11223. [[CrossRef](#)] [[PubMed](#)]
32. Hu, J.; Wang, J.; Nguyen, T.H.; Zheng, N. The chemistry of amine radical cations produced by visible light photoredox catalysis. *Beilstein J. Org. Chem.* **2013**, *9*, 1977–2001. [[CrossRef](#)]
33. Wenk, J.; Canonica, S. Phenolic antioxidants inhibit the triplet-induced transformation of anilines and sulfonamide antibiotics in aqueous solution. *Environ. Sci. Technol.* **2012**, *46*, 5455–5462. [[CrossRef](#)] [[PubMed](#)]
34. Yang, J.D.; Xue, X.S.; Ji, P. Internet Bond-Energy Databank (pKa and BDE): iBond Home Page. Available online: <http://ibond.nankai.edu.cn/> (accessed on 10 January 2024).
35. Land, E.J.; Porter, G. Primary photochemical processes in aromatic molecules. Part 8—Absorption spectra and acidity constants of anilino radicals. *Trans. Far. Soci.* **1963**, *59*, 2027–2037. [[CrossRef](#)]
36. Jovanovic, S.V.; Steenken, S.; Tomic, M.; Marjanovic, B.; Simic, M.G. Flavonoids as antioxidants. *J. Am. Chem. Soc.* **1994**, *116*, 4846–4851. [[CrossRef](#)]
37. Hall, N.F.; Sprinkle, M.R. Relations between the structure and strength of certain organic bases in aqueous solution. *J. Am. Chem. Soc.* **1932**, *54*, 3469–3485. [[CrossRef](#)]

38. Shapiro, S.L.; Isaacs, E.S.; Bandurco, V.; Freedman, L. Apparent dissociation constants of haloalkylamines. *J. Med. Pharm. Chem.* **1962**, *5*, 793–799. [CrossRef]
39. Sheppard, W.A. The electrical effect of the sulfur pentafluoride group. *J. Am. Chem. Soc.* **1962**, *84*, 3072–3076. [CrossRef]
40. Sheppard, W.A. The electronic properties of fluoroalkyl groups. Fluorine p- $\pi$  interaction. *J. Am. Chem. Soc.* **1965**, *87*, 2410–2420. [CrossRef]
41. Roberts, J.D.; Webb, R.L.; McElhill, E.A. The electrical effect of the trifluoromethyl group. *J. Am. Chem. Soc.* **1950**, *72*, 408–411. [CrossRef]
42. Vandenberg, J.M.; Henrich, C.; Vanden, B.S.G. Comparison of  $pK_a$  values determined by electrometric titration and ultraviolet absorption methods. *Anal. Chem.* **1954**, *26*, 726–727. [CrossRef]
43. Garcia, B.; Leal, J.M. The  $pK_b$  calculation of strong bases—A revision of various methods. *Collect. Czech. Chem. C* **1987**, *52*, 299–307. [CrossRef]
44. Biggs, A.I.; Robinson, R.A. The ionisation constants of some substituted anilines and phenols: A test of the Hammett relation. *J. Chem. Soc.* **1961**, 388–393. [CrossRef]
45. Kalopissis, G. Structure-activity relationships of aromatic amines in the ames *salmonella typhimurium* assay. *Mutat. Res.-Fund. Mol. M.* **1991**, *246*, 45–66. [CrossRef] [PubMed]
46. Garcia, B.; Domingo, I.; Domingo, P.L.; Leal, J.M. Determination of limiting molar conductivities of weak organic-acids in aqueous-solutions. *Collect. Czech. Chem. C* **1991**, *56*, 1184–1192. [CrossRef]
47. Hargreaves, M.K.; Stevinson, E.A.; Evans, J. The apparent dissociation constants of various weak acids in mixed aqueous solvents. *J. Chem. Soc.* **1965**, 4582–4583. [CrossRef]
48. McCoy, R.D.; Swinehart, D.F. The ionization constant of metanilic acid from 0 to 50° by means of e.M.F. Measurements. *J. Am. Chem. Soc.* **1954**, *76*, 4708–4710. [CrossRef]
49. Yu, A.; Liu, Y.H.; Li, Z.C.; Cheng, J.P. Computation of  $pK_a$  values of substituted aniline radical cations in dimethylsulfoxide solution. *J. Phys. Chem. A* **2007**, *111*, 9978–9987. [CrossRef]
50. Bordwell, F.G.; Zhang, X.M.; Cheng, J.P. Bond dissociation energies of the nitrogen-hydrogen bonds in anilines and in the corresponding radical anions. Equilibrium acidities of aniline radical cations. *J. Org. Chem.* **1993**, *58*, 6410–6416. [CrossRef]
51. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Petersson, G.A.; Nakatsuji, H.; et al. *Gaussian 16 Rev. C.01*; Fox, Gaussian, Inc.: Wallingford, CT, USA, 2016.
52. Zhao, Y.; Truhlar, D.G. The m06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four m06-class functionals and 12 other functionals. *Theor. Chem. Acc.* **2008**, *120*, 215–241. [CrossRef]
53. Clark, T.; Chandrasekhar, J.; Spitznagel, G.W.; Schleyer, P.V.R. Efficient diffuse function-augmented basis sets for anion calculations. Iii. The 3-21+g basis set for first-row elements, li–f. *J. Comput. Chem.* **1983**, *4*, 294–301. [CrossRef]
54. Krishnan, R.; Binkley, J.S.; Seeger, R.; Pople, J.A. Self-consistent molecular orbital methods. Xx. A basis set for correlated wave functions. *J. Chem. Phys.* **2008**, *72*, 650–654. [CrossRef]
55. Cancès, E.; Mennucci, B.; Tomasi, J. A new integral equation formalism for the polarizable continuum model: Theoretical background and applications to isotropic and anisotropic dielectrics. *J. Chem. Phys.* **1997**, *107*, 3032–3041. [CrossRef]
56. Mennucci, B.; Tomasi, J. Continuum solvation models: A new approach to the problem of solute's charge distribution and cavity boundaries. *J. Chem. Phys.* **1997**, *106*, 5151–5158. [CrossRef]
57. Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for h to rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305. [CrossRef] [PubMed]
58. Zheng, J.J.; Xu, X.F.; Truhlar, D.G. Minimally augmented karlsruhe basis sets. *Theor. Chem. Acc.* **2011**, *128*, 295–305. [CrossRef]
59. Montgomery, J.A.; Frisch, M.J.; Ochterski, J.W.; Petersson, G.A. A complete basis set model chemistry. Vi. Use of density functional geometries and frequencies. *J. Chem. Phys.* **1999**, *110*, 2822–2827. [CrossRef]
60. Teranishi, K.; Ishikawa, A.; Sato, H.; Nakai, H. Systematic investigation of the thermodynamic properties of amine solvents for CO<sub>2</sub> chemical absorption using the cluster-continuum model. *B. Chem. Soc. Jpn.* **2017**, *90*, 451–460. [CrossRef]
61. Banerjee, S.; Bhanja, S.K.; Kanti, C.P. Quantum chemical predictions of aqueous  $pK_a$  values for oh groups of some  $\alpha$ -hydroxycarboxylic acids based on ab initio and dft calculations. *Comput. Theor. Chem.* **2018**, *1125*, 29–38. [CrossRef]
62. Casanovas, R.; Frau, J.; Ortega-Castro, J.; Salvà, A.; Donoso, J.; Muñoz, F. Absolute and relative  $pK_a$  calculations of mono and diprotic pyridines by quantum methods. *J. Mol. Struct.-Theochem.* **2009**, *912*, 5–12. [CrossRef]
63. Najgebauer, P.; Staś, M.; Wrzaliak, R.; Broda, M.A.; Wieczorek, P.P.; Andrushchenko, V.; Kupka, T. Muscimol hydration and vibrational spectroscopy—The impact of explicit and implicit water. *J. Mol. Liq.* **2022**, *363*, 119870. [CrossRef]
64. Kato, M.; Izuka, S.; Fujihara, T.; Nagasawa, A.; Kawai, S.; Tanaka, T.; Takayanagi, T. Electronic structure calculation study of metal complexes with a phytosiderophore mugineic acid. *Inorg. Chim. Acta.* **2011**, *370*, 304–310. [CrossRef]
65. Zhang, S.M. A reliable and efficient first principles-based method for predicting  $pK_a$  values. Iii. Adding explicit water molecules: Can the theoretical slope be reproduced and  $pK_a$  values predicted more accurately? *J. Comput. Chem.* **2012**, *33*, 517–526. [CrossRef] [PubMed]
66. Bao, J.W.; Lucas, Z.J.; Alecu, I.M.; Lynch, B.J.; Zhao, Y.; Truhlar, D.G. Database of Frequency Scale Factors for Electronic Model Chemistries. Available online: <https://comp.chem.umn.edu/freqscale/version3b2.htm> (accessed on 16 July 2023).

67. Lu, T.; Chen, Q.X. Shermo: A general code for calculating molecular thermochemistry properties. *Comput. Theor. Chem.* **2021**, *1200*, 113249. [[CrossRef](#)]
68. Ribeiro, R.F.; Marenich, A.V.; Cramer, C.J.; Truhlar, D.G. Use of solution-phase vibrational frequencies in continuum models for the free energy of solvation. *J. Phys. Chem. B* **2011**, *115*, 14556–14562. [[CrossRef](#)]
69. Grimme, S. Supramolecular binding thermodynamics by dispersion-corrected density functional theory. *Chem.-Eur. J.* **2012**, *18*, 9955–9964. [[CrossRef](#)] [[PubMed](#)]
70. Marenich, A.V.; Ding, W.D.; Cramer, C.J.; Truhlar, D.G. Resolution of a challenge for solvation modeling: Calculation of dicarboxylic acid dissociation constants using mixed discrete–continuum solvation models. *J. Phys. Chem. Lett.* **2012**, *3*, 1437–1442. [[CrossRef](#)]
71. Sutton, C.C.R.; Franks, G.V.; da Silva, G. First principles  $pK_a$  calculations on carboxylic acids using the smd solvation model: Effect of thermodynamic cycle, model chemistry, and explicit solvent molecules. *J. Phys. Chem. B* **2012**, *116*, 11999–12006. [[CrossRef](#)]
72. Santra, G.; Sylvetsky, N.; Martin, J.M.L. Minimally empirical double-hybrid functionals trained against the gmtn55 database: Revdsc-d4, revdod-pbe-d4, and dod-scan-d4. *J. Phys. Chem. A* **2019**, *123*, 5129–5143. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.