

Article

Computation of the pK_a Values of Gallic Acid and Its Anionic Forms in Aqueous Solution: A Self-Similar Transformation Approach for Accurate Proton Hydration Free Energy Estimation

Marcin Molski 

Department of Quantum Chemistry, Faculty of Chemistry, Adam Mickiewicz University of Poznań, ul. Uniwersytetu Poznańskiego 8, 61-614 Poznań, Poland; mamolski@amu.edu.pl

Abstract: The Gibbs free energies of gallic acid (GA) and its anionic forms in aqueous solution were computed utilizing density functional theory (DFT) at the LSDA, M062X, B3LYP/QZVP levels, in conjunction with the SMD solvation model. The pK_a values corresponding to the four-step deprotonation of GA were determined through a non-linear self-similar transformation expressed as, $pK_a = a \cdot pK_a(\text{the})^c$ which establishes a link between theoretical and experimental pK_a values. This approach replaces the previously employed linear relationship, $pK_a = a \cdot pK_a(\text{the}) + b$. The proposed model demonstrates high accuracy in reproducing the experimental $pK_{a1} = 4.16 \pm 0.02$, $pK_{a2} = 8.55 \pm 0.01$, $pK_{a3} = 11.40 \pm 0.10$, $pK_{a4} = 12.8 \pm 0.40$ values of GA, with a standard error (SE) of 0.045 and a mean absolute error (MAE) of 0.019 in pK_a unit. Furthermore, it facilitates the precise determination of the Gibbs free energy of the proton hydration, yielding $\Delta G(H^+)_{aq} = 259.4272(75)$ [kcal mol^{−1}]. This result conforms acceptably with the experimental value of $\Delta G(H^+)_{aq} = -259.5$ [kcal mol^{−1}].

Keywords: DFT method; gallic acid; multi-step deprotonation; proton hydration; Gibbs energy; pK_a descriptor; selectivity; reactivity



Academic Editor: Wim Buijs

Received: 5 January 2025

Revised: 27 January 2025

Accepted: 30 January 2025

Published: 6 February 2025

Citation: Molski, M. Computation of the pK_a Values of Gallic Acid and Its Anionic Forms in Aqueous Solution: A Self-Similar Transformation Approach for Accurate Proton Hydration Free Energy Estimation. *Molecules* **2025**, *30*, 742. <https://doi.org/10.3390/molecules30030742>

Copyright: © 2025 by the author. 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

Deprotonation represents a fundamental chemical reaction that occurs across both biological [1,2] and abiotic [2–4] systems, thereby representing a significant research area focused on molecular reactivity and selectivity [3,4]. Monocarboxylic acids [5,6], along with their combinations with phenols and polyphenols [7,8], are particularly compelling subjects of study due to the presence of multiple reactive carboxyl and hydroxyl groups that engage actively in the deprotonation process. Experimental and theoretical analyses of protonation strive to assess the reactivity of each active group, identifying the sequential order of proton removal from the base molecule and evaluating the relative susceptibility of each group to deprotonation [3,4,9]. The chemical reactivity of these compounds is largely influenced by the electrophilic nature of the carboxylic moiety, as well as the resonance stabilization that occurs upon proton dissociation. These factors jointly contribute to both the acidity and predominant chemical reactivity of the carboxylic and phenolic groups. They can be numerically (quantitatively) characterized by the parameter pK_a , which stands for the negative base-10 logarithm of an acid dissociation constant in a specific solvent, usually of a hydrophilic nature facilitating the proton detachment. Experimentally and theoretically determining the pK_a values of the compounds, especially in aqueous solution, is essential for a comprehensive understanding of numerous chemical processes and provides insight into the deprotonation state of a molecule within a specific solvent environment. Theoretical

approaches to calculating pK_a parameters have garnered significant interest, particularly for applications involving molecules that have not yet been synthesized, those for which experimental determination of pK_a is challenging, and larger molecules where local environmental factors alter the intrinsic pK_a values. An example of the latter includes certain amino acids incorporated within polypeptide chains [2]. Achieving chemical accuracy in pK_a calculations is inherently challenging, as an error of $1.36 \text{ [kcal mol}^{-1}\text{]}$ in the change in free energy associated with deprotonation in a solvent corresponds to a deviation of 1 pK_a unit [2]. In this regard, gallic acid (GA, Figure 1) is among the most extensively studied compounds [3,4,9] with respect to selectivity and functional group activity, owing to its four active hydrogens in carboxyl moiety and hydroxyl groups attached to the benzene ring.

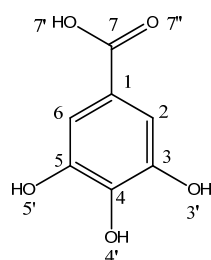


Figure 1. Gallic acid, 3,4,5-trihydroxybenzoic acid.

Research has demonstrated [3,4] that the deprotonation sequence in GA in water solution follows the order $O_{7'}\text{-H}$, $O_{4'}\text{-H}$, $O_{3'}\text{-H}$, $O_{5'}\text{-H}$, necessitating the use of distinct deprotonation descriptors: $pK_{a1}(O_{7'}\text{-H})$, $pK_{a2}(O_{4'}\text{-H})$, $pK_{a3}(O_{3'}\text{-H})$, $pK_{a4}(O_{5'}\text{-H})$ associated with the four-step deprotonation of GA presented in Figure 2.

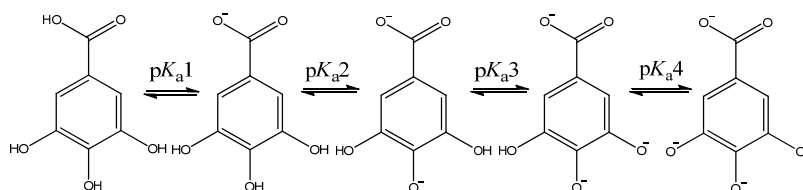


Figure 2. Four-step deprotonation of gallic acid and related pK_a parameters.

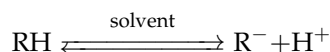
Despite the broad utility of the pK_a descriptor, the acquisition of accurate and unique experimental data for GA remains challenging. They are commonly determined through ultraviolet–visible (UV–Vis) and Raman spectroscopy [10–13], or various analytical techniques [14]. For GA, the following experimental $pK_{a1,2,3,4}$ values were reported with the standard errors if available: Set I (4.50, 7.05, 8.75, 10.25) [11], Set II (4.4 ± 0.1 , 8.8 ± 0.1 , 10.0 ± 0.1 , 11.4 ± 0.1) [12], Set III (4.44, 8.54, 10.05, 11.30) [14], Set IV (4.16 ± 0.02 , 8.55 ± 0.01 , 11.40 ± 0.10 , 12.8 ± 0.40) [13]. The availability of experimental pK_a values, measured with varying degrees of precision, facilitates a comparison with the theoretical results generated by different models and methods employed in the calculations. A straightforward approach to the theoretical simulation of solvent effects involves the application of various solvation models, with particular emphasis on the Conductor-like Polarizable Continuum Model (C-PCM) [15], the Integral Equation Formalism Polarizable Continuum Model (IEF-PCM) [16], and the universal Solvation Model Density (SMD) based on solute electron density [17]. They are widely employed to calculate the theoretical pK_a values of chemical compounds in solution through the relationship shown below [18–21]:

$$pK_a = \frac{G(R^-)_{\text{sol}} - G(RH)_{\text{sol}} + G(H^+)_{\text{sol}} - RT \ln(V)}{RT \ln(10)} = \frac{\Delta G + G(H^+)_{\text{sol}} - 1.8942}{1.3642} \quad (1)$$

in which R is the gas constant, $T = 298.15$ [K] is the temperature (which is used in all calculations), and $RT \cdot \ln(V) = 1.8942$ [kcal mol^{−1}] stands for the correction [21]

$$G^1 [\text{mol}] = G^1 [\text{atm}] - 1.8942 [\text{kcal mol}^{-1}] \quad (2)$$

for the reference state ($V = 24.46$ [L], $T = 298.15$ [K]), from 1 [atm] to 1 [mol], whereas ΔG in [kcal mol^{−1}] denotes the Gibbs free energy of the deprotonation according to the reaction



The Gibbs free energy of the solvated proton $G(\text{H}^+)_{\text{sol}}$, is related to the solvation energy of the proton $\Delta G(\text{H}^+)_{\text{sol}}$ by the relation [21]

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

in which $G(\text{H}^+)_{\text{gas}} = -6.2883$ [kcal mol^{−1}] [22] is the free energy of a proton in the gas phase under a pressure of 1 [atm]. From (1) and (2) it follows that Equation (1) is useful evaluating pK_a using the calculated (theoretical) $G(\text{H}^+)_{\text{sol}}^{\text{atm}}$ value, whereas $G(\text{H}^+)_{\text{sol}}^{\text{mol}} = G(\text{H}^+)_{\text{sol}}^{\text{atm}} - 1.8942$ [kcal mol^{−1}] can be compared to the experimental data. To estimate the Gibbs free energies of the solvated species involved in the deprotonation process, quantum-chemical computational methods at different levels of theory have been employed [5,6]. The calculations performed for the water medium revealed that the application of Equations (1)–(3) in the calculation of pK_a presents three primary challenges:

- (i) The accuracy of pK_a reproduction is contingent upon the computational method employed to determine $G(\text{RH})$ and $G(\text{R}^-)$; it varies within an MAE (mean absolute error) range of 0.51–2.86 pK_a for monoprotic acids, depending on the method used in the calculations [5].
- (ii) The value of $\Delta G(\text{H}^+)_{\text{aq}}$, as determined by experimental [23], theoretical [24–30], and mixed [30] approaches, ranges from -244.9 to -266.7 [kcal mol^{−1}], which significantly influences the predicted pK_a values.
- (iii) Equation (1) was effectively utilized to reproduce pK_a parameters for monoprotic [5–7] compounds; however, its application to polyprotic molecules undergoing multi-step proton detachment yielded [4] results that deviated significantly from the experimental ones.

To address the first challenge, calculations are conducted using a range of available semi-empirical and ab initio methods, incorporating various basis sets and solvation models to identify those that most accurately reproduce experimental pK_a values [5,6]. Additionally, a linear similarity relation [5,31–33] such as

$$pK_a = a \cdot pK_a(\text{the}) + b \quad (4)$$

is employed to enhance the accuracy of pK_a reproduction by adjusting the a and b parameters to the set of experimental $pK_a = pK_a(\text{exp})$ data and theoretical $pK_a(\text{the})$ values. This approach improves the accuracy of pK_a prediction from $\text{MAE} = 0.51\text{--}2.86$ to $\text{MAE} = 0.30\text{--}0.67$ in pK_a units for 22 monoprotic acids analyzed [5]. In the case of thiols, the $\text{MAE} = 0.57\text{--}0.96$ [32], whereas for phenols, carboxylic acids, and amines, the MAE values are lower than 0.7 pK_a units [31]. However, this level of accuracy still remains higher than the experimental error range of 0.01–0.1 in pK_a units obtained, e.g., for the multi-step deprotonation of GA [12,13]. To concurrently address the first and second challenges, Dutra

et al. [6] proposed an approach wherein the average $\langle G(H^+)_{sol} \rangle$ of the $G(H^+)_{sol}$ values could be calculated using the equation below

$$G(H^+)_{sol} = 1.3642 \cdot pK_a(\text{exp}) + G(RH)_{sol} - G(R^-)_{sol} + 1.8942 \quad (5)$$

depending on the experimental $pK_a(\text{exp})$ and the theoretical $G(RH)_{sol}$ and $G(H^-)_{sol}$ values obtained using diverse calculation methods for a specific class of compounds with similar chemical structures. This average value is subsequently incorporated into Equation (1) to form Equation (6):

$$pK_a = \frac{G(R^-)_{sol} - G(RH)_{sol} + \langle G(H^+)_{sol} \rangle - 1.8942}{1.3642} \quad (6)$$

which can be used to calculate the pK_a values for all compounds within the studied class. Calculations performed using this scheme for 22 monocarboxylic acids in an aqueous medium revealed [6] that the combination of density functional theory (DFT) with the local spin density approximation (LSDA) functional provided superior performance in nearly all calculations, achieving accuracy levels comparable to those obtained using the G4CEP (Gaussian 4 Compact Effective Pseudopotential) method employed by Silva and Custodio [5]. In this study, the calculations conducted for 22 monoprotic acids using the G4CEP method, in conjunction with (i) the SMD solvation model, (ii) SMD +1 H₂O (with one explicit water molecule), and (iii) the SMD + 1 H₂O with a linear correction based on Equation (4) and experimental $pK_a(\text{exp})$ values, yielded mean absolute errors of 0.83, 0.51, 0.30 in pK_a units, respectively. A comparative analysis with the results obtained by DFT at the B3LYP and BMK theory levels revealed significant errors reaching up to 2 pK_a units. In contrast, the largest deviations observed for the G4CEP method rarely exceeded 1 pK_a unit [5].

Based on the results reported above, the primary objectives of the present study were fourfold:

- (i) To employ a modified version of the method proposed by Dutra et al. [6] to determine the $pK_{a1,2,3,4}$ parameters for the four-stage deprotonation of GA, within the bounds of experimental errors of 0.01–0.40 pK_a units.
- (ii) To calculate the Gibbs free energy of the hydrated proton $G(H^+)_{aq}$ and proton hydration $\Delta G(H^+)_{aq}$ by fitting this parameter via the improved version of Equation (4) to the experimental pK_a values reported in [11–14].
- (iii) To select the experimental data that most accurately represent the pK_a values for GA within the framework of the proposed approach.
- (iv) To demonstrate that the accurate reproduction of the pK_a of GA, within the range of experimental errors, can be achieved using the DFT method and LSDA/QZVP level of theory combined with the SMD solvation model, without the explicit inclusion of a water molecule.

2. Results and Discussion

To achieve the primary objectives of this work, the geometries of the neutral GA molecule and its four anionic forms (Figure 2) were optimized in the water medium employing the DFT method at the different theory levels (LSDA, M062X, B3LYP), and the basis set by QZVP and the SMD solvation model and the reasons for their application are set out in the Materials and Methods section. The optimized geometries of the neutral GA molecule and its anionic forms generated at the LSDA/QZVP theory level are presented in Figure 3, while Tables 1 and 2 report the corresponding Gibbs free energy values and their differences used in the determination of the $pK_{a1,2,3,4}$ parameters.

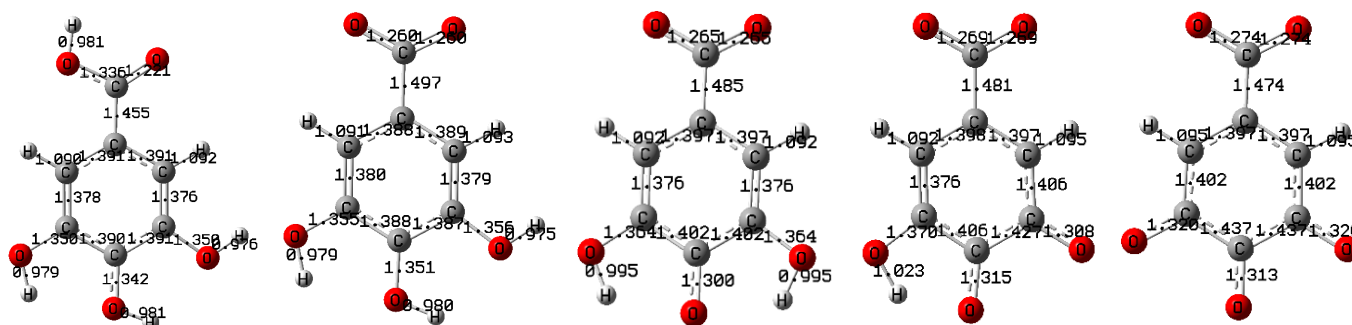


Figure 3. The optimized geometries of neutral GA molecule and its anionic forms GA^{-N} $N = 1, 2, 3, 4$. The calculations were performed in a water medium at the LSDA/QZVP theory level, using the SMD solvation model. The values of the bond lengths are presented.

Table 1. The differences ΔG_N [kcal mol^{-1}] in Gibbs free energies of neutral GA^0 molecule and its anionic forms GA^{-N} $N = 1, 2, 3, 4$ including the zero-point energy corrections from Table S1. The calculations were performed in the water medium at the LSDA, M062X, and B3LYP/QZVP theory levels, using the SMD solvation model.

	ΔG_N [kcal mol^{-1}]	LSDA	M062X	B3LYP
ΔG_1	$G(\text{GA}^{-1})_{\text{aq}} - G(\text{GA}^0)_{\text{aq}}$	267.6773	270.6210	272.7539
ΔG_2	$G(\text{GA}^{-2})_{\text{aq}} - G(\text{GA}^{-1})_{\text{aq}}$	270.8450	279.5730	280.5099
ΔG_3	$G(\text{GA}^{-3})_{\text{aq}} - G(\text{GA}^{-2})_{\text{aq}}$	283.2258	289.4136	290.8582
ΔG_4	$G(\text{GA}^{-4})_{\text{aq}} - G(\text{GA}^{-3})_{\text{aq}}$	295.9617	297.0059	298.7058

Table 2. The values of the parameters a , c (dimensionless), and $G(\text{H}^+)_{\text{aq}}$ [kcal mol^{-1}] fitted via Equation (7) to the four sets of experimental pK_a for GA reported in [11–14]. The indicators of goodness of the fit/standard error (SE) of estimation and the coefficient of determination, R^2 , are presented.

	LSDA	M062X	B3LYP	LSDA	M062X	B3LYP
Parameter	Set I			Set II		
$G(\text{H}^+)_{\text{aq}}$	−265.47(32)	−62.1(3.2)	−266.6(2.5)	−265.775(20)	−268.12(65)	−270.51(44)
a	5.87(59)	2.02(80)	2.84(94)	7.87(64)	5.36(48)	5.90(96)
c	0.177(38)	0.51(11)	0.41(10)	0.114(33)	0.247(66)	0.216(61)
SE	0.3861	0.2031	0.2630	0.4959	0.4020	0.4383
R^2	0.9918	0.9977	0.9962	0.9910	0.9941	0.9930
Parameter	Set III			Set IV		
$G(\text{H}^+)_{\text{aq}}$	−265.764(26)	−267.78(51)	−270.29(38)	−265.7155(75)	−266.81(45)	−269.63(15)
a	7.59(83)	4.91(57)	5.48(60)	7.271(57)	3.60(34)	4.33(16)
c	0.126(23)	0.275(41)	0.240(40)	0.1858(31)	0.420(31)	0.364(13)
SE	0.3327	0.2261	0.2666	0.0454	0.1453	0.0766
R^2	0.9959	0.9981	0.9973	1.0000	0.9995	0.9994

In this study, we examined two potential scenarios (Figure 4) for the deprotonation of GA, involving the dianion A, which possesses a higher energy, at $G(\text{GA}^{-2}) = -642.539452$ [Ha], than its conformer B, at $G(\text{GA}^{-2}) = -642.544071$ [Ha]. The energy disparity arises from a conformational change that contributes additional energy 2.90 [kcal mol^{-1}]. The rotation barrier for an $A \rightarrow B$ transition is equal to 2.93 [kcal mol^{-1}], whereas for $B \rightarrow A$ it is 5.61 [kcal mol^{-1}]. These results were calculated using the free energy of the transition state $[\text{AB}]^\ddagger$ $G^\ddagger = -642.534128$ [Ha]. Despite the energetic difference between dianions A and B, the reaction pathways for both molecules are isoenergetic as $\Delta G_2 + \Delta G_3 = \Delta G'_2 + \Delta G'_3$.

However, incorporating the low-energy dianion B into the calculations yields $\Delta G'_2$ and $\Delta G'_3$ values that fail to accurately reproduce pK_{a2} and pK_{a3} descriptors. For this reason, all calculations were carried out using scheme A.

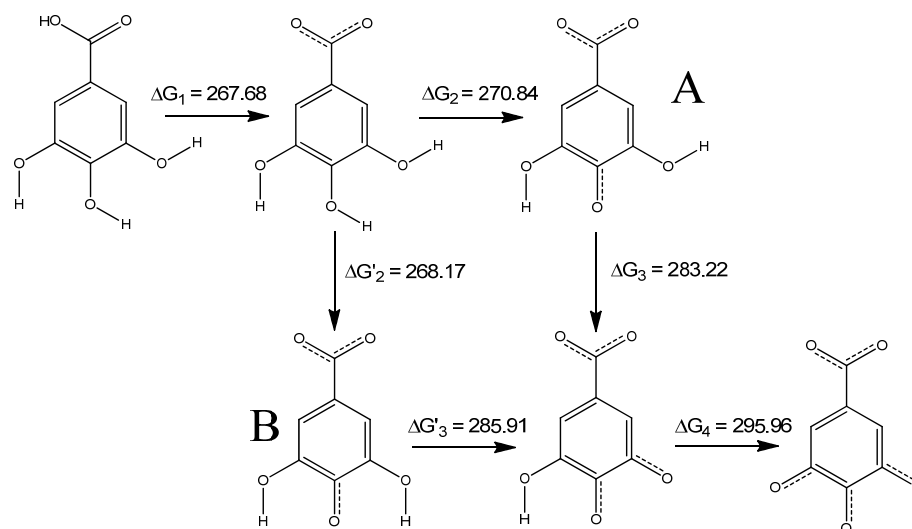


Figure 4. Potential scenarios for the deprotonation of GA, involving dianions A and B. The values (in [kcal mol^{−1}]) of Gibbs free energies for the A reaction pathway are taken from Table S1 and Table 1, whereas $\Delta G'_2 = 268.1668$ and $\Delta G'_3 = 285.9096$ [kcal mol^{−1}]. Both processes are isoenergetic as $\Delta G_2 + \Delta G_3 = \Delta G'_2 + \Delta G'_3$.

The calculations performed at the LSDA/QZVP theory level using the SMD solvation model indicated that the method based on Equations (5) and (6) yields an unsatisfactory reproduction of the experimental pK_a values of GA. For example, set IV provides the mean value $\langle G(H^+)_{aq} \rangle = -263(5)$ [kcal mol^{−1}], exhibiting a substantial standard error, and produces $pK_a = 1.77, 4.09, 13.16, 22.50$. Attempts to refine these results using Equation (4) also failed to generate satisfactory agreement with the experimental pK_a (exp) values of 4.16, 8.55, 11.40, and 12.80. For the adjusted parameters $a = 0.362$ (125) and $b = 5.47$ (1.65) obtained with a standard error (SE) of 2.0479 of estimation and a coefficient of determination (R^2) of 0.8077, the corrected pK_a values were equal to 6.11, 6.95, 10.24, 13.62. This discrepancy suggests that the relationship between the theoretical and experimental pK_a values is more complex than the linear relation (4) utilized so far. Accordingly, in this study, the linear similarity relation (4) was replaced with a non-linear self-similar formula like that shown below

$$pK_{aN} = a \cdot pK_a(\text{the})^c = a \left[\frac{\Delta G_N + G(H^+)_{aq} - 1.8942}{1.3642} \right]^c \quad (7)$$

$$\Delta G_N = G(GA^{-N})_{aq} - G(GA^{-(N-1)})_{aq} \quad N = 1, 2, 3, 4$$

which simplifies to the linear relation (4) with $b = 0$ when c approaches 1. The parameters a , c , and $G(H^+)_{aq}$ in Equation (7) fitted to the four sets of available experimental GA data have subsequently been employed to reproduce the $pK_{a1,2,3,4}$ parameters characterizing the four-step deprotonation of GA. The results of these calculations are reported in Tables 2 and 3.

Table 3. The values of $pK_{a1,2,3,4}$ reproduced by Equation (7) with parameters from Tables 1 and 2 evaluated for the four sets of experimental pK_a data reported in [11–14]. The mean absolute error (MAE) of the data reproduction is presented.

Parameter	pK_a (exp)		pK_a		pK_a (exp)		pK_a	
	Set I	LSDA	M062X	B3LYP	Set II	Set I	LSDA	M062X
pK_{a1}	4.50	4.52	4.52	4.52	4.4(1)	4.61	4.40	4.40
pK_{a2}	7.05	6.92	6.95	6.93	8.8(1)	8.67	8.67	8.67
pK_{a3}	8.75	9.06	8.91	8.96	10.0(1)	10.40	10.32	10.35
pK_{a4}	10.25	10.05	10.17	10.14	11.4(1)	11.13	11.20	11.17
MAE		0.165	0.090	0.115		0.253	0.163	0.178
Parameter	Set III	LSDA	M062X	B3LYP	Set IV	LSDA	M062X	B3LYP
pK_{a1}	4.44	4.55	4.44	4.44	4.16(2)	4.161	4.155	4.168
pK_{a2}	8.54	8.45	8.46	8.46	8.55(1)	8.536	8.607	8.600
pK_{a3}	10.05	10.32	10.23	10.26	11.40(10)	11.44	11.28	11.37
pK_{a4}	11.30	11.12	11.19	11.16	12.80(40)	12.78	12.87	12.87
MAE		0.163	0.118	0.108		0.019	0.063	0.060

The results obtained reveal that the approach proposed enables both the precise determination of the Gibbs free energy of the hydrated proton $G(H^+)_{aq} = -265.7155(75)$ [kcal mol^{−1}] and the accurate reproduction of the experimental pK_a values by taking advantage of the LSDA/QZVP theory level and set IV of the datapoints [13]. The results fall within the range of experimental errors, with MAE = 0.019 in pK_a units. As experimental data are typically represented by the Gibbs free energy of the proton hydration, this parameter can also be determined through a fitting procedure utilizing Equations (3) and (7). The findings presented in Table 4 demonstrate that parameters $\Delta G(H^+)_{aq} = -259.4272(75)$ [kcal mol^{−1}], $a = 7.271(57)$, and $c = 0.1858(31)$ reproduce set IV [13] of pK_a with $R^2 = 1.0000$ and SE = 0.0454.

Table 4. The values of the parameters a , c (dimensionless), and $\Delta G(H^+)_{aq}$ ([kcal mol^{−1}]) fitted to the four sets of experimental pK_a for GA via Equation (7) including Equation (3) and ΔG_N calculated at the LSDA/QZVP theory level using the SMD solvation model and a water medium. The indicators of goodness of the fit/standard error (SE) of estimation and the coefficient of determination (R^2) are presented. The mean absolute error (MAE) and the normalized mean absolute error (NMAE defined in Equation (8)) of the reproduced pK_{aN} parameters are presented.

Parameter	Set I	Set II	Set III	Set IV
$\Delta G(H^+)_{aq}$	−259.19(32)	−259.486(20)	−259.476(26)	−259.4272(75)
a	5.87(59)	7.871(39)	7.59(43)	7.271(57)
c	0.177(38)	0.114(33)	0.126(23)	0.1858(31)
R^2	0.9918	0.9910	0.9959	1.0000
SE	0.3861	0.4959	0.3327	0.0454
pK_{a1}	4.51	4.40	4.44	4.16
pK_{a2}	6.92	8.67	8.45	8.54
pK_{a3}	9.08	10.40	10.32	11.44
pK_{a4}	10.05	11.13	11.12	12.78
MAE	0.160	0.199	0.134	0.019
NMAE		1.992		0.463

The $\Delta G(H^+)_{aq} = -259.4272(75)$ calculated using the LSDA/QZVP theory level and set IV is consistent with the experimental value of -259.5 [kcal mol^{−1}] obtained by Lim et al. [23]. This value corresponds to the average -4.5 [V] of five measurements

of the standard hydrogen potential (-4.43 , -4.43 , -4.44 , -4.48 , -4.73 [V]), which allows for the determination of a range of the experimental values of $\Delta G(\text{H}^+)_{\text{aq}}$ from -254 to -261 [kcal mol $^{-1}$] [23]. The reproduction of the experimental pK_a values IV for GA using relationship (7) proved to be more accurate than that obtained using linear Equation (4). The parameters $\langle \Delta G(\text{H}^+)_{\text{aq}} \rangle = -258.6(7.4)$ [kcal mol $^{-1}$], $a = 0.36(13)$, and $b = 5.4(1.7)$ reproduce set IV of pK_a with $R^2 = 0.8079$ and $\text{SE} = 0.7118$, yielding corrected pK_a values of 5.96, 6.25, 7.38, 8.55. These results clearly indicate that, for the multivariate deprotonation of GA, the linear transformation (4) should be replaced by the self-similar formula given in Equation (7). Furthermore, the plots of the relationship linking pK_a and ΔG_N (Figure 5) generated for sets I, II, III, and IV highlight the strongly non-linear nature of this relation. The graphs of the pK_a as function of ΔG_N qualitatively have the same form for ΔG_N calculated at the LSDA, M062X, and B3LYP/QZVP theory levels.

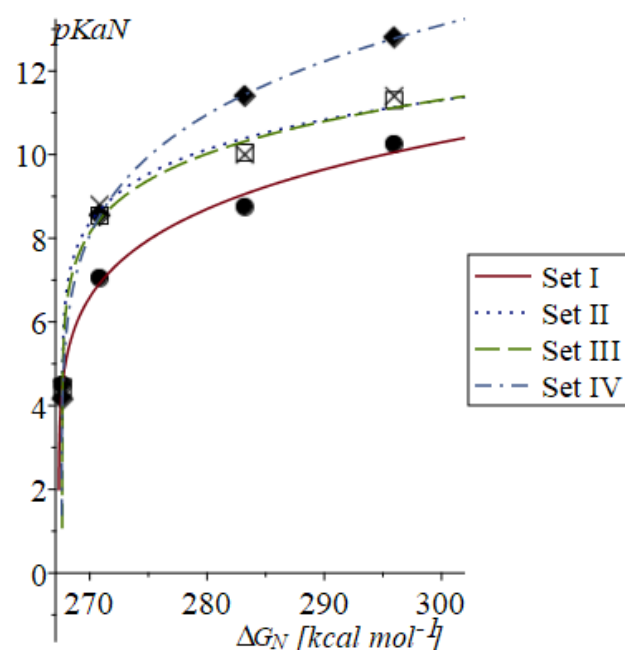


Figure 5. Plots of the self-similar relationship (7) with ΔG_N calculated at the LSDA/QZVP theory level, using the SMD model (water medium) and the four experimental datasets represented by the graphical symbols I—●; II—□; III—×; and IV—◆.

In works [5,6], it was demonstrated that the accuracy of reproducing the experimental pK_a values for monoprotic acids can be enhanced through the implementation of an SMD solvation model that incorporates an additional explicitly included water molecule. Given that the model proposed in this work effectively reproduces set IV, achieving a maximum R^2 of 1.0000, and a satisfactory $\text{MAE} = 0.019$, $\text{NMAE} = 0.463$ in pK_a units, the inclusion of the water molecule should not affect the accuracy of the calculations. To substantiate this hypothesis, the calculations were performed utilizing the scheme proposed by Kelly et al. [21], which generates $a = 6.15(12)$, $c = 0.2648(79)$, and $\Delta G(\text{H}^+)_{\text{aq}} = -260.166(44)$ [kcal mol $^{-1}$] from $\Delta G_1 = 266.4521$, $\Delta G_2 = 268.6051$, $\Delta G_3 = 278.0447$, and $\Delta G_4 = 299.3631$ [kcal mol $^{-1}$], values that reproduce the pK_a of set IV with $\text{SE} = 0.073$, $\text{MAE} = 0.030$, and $\text{NMAE} = 0.7651$ in pK_a units. These results exhibit a lower degree of accuracy compared to those obtained without the inclusion of the water molecule in the calculations at the LSDA/QZVP theory level. To investigate the impact of dispersion effects on the pK_a and $\Delta G(\text{H}^+)_{\text{aq}}$ parameters, calculations were performed employing the QZVP basis set, the SMD solvation model, and (i) the functional developed by Head-Gordon and collaborators [34], which incorporates Grimme's dispersion model D2 [34], and the B3LYP functional combined with the D3

dispersion model. Such a combination has been recommended in [35]. The computational results, summarized in Table S2, indicate that dispersion effects have a negligible influence on the deprotonation of GA. It is noteworthy, however, that the value of the parameter $\Delta G(H^+)_{aq} = -264.10(23)$ [kcal mol^{−1}] is consistent with $\Delta G(H^+)_{aq} = -264.29$ [kcal mol^{−1}] reported by Zhan and Dixon [27]. To evaluate the impact of basis sets on the accuracy of the reproduction of the pK_a and $\Delta G(H^+)_{aq}$ parameters for GA, calculations were conducted at the 6311++G(d,p), aug-cc-pVQZ, QZVP/LSDA theory levels in conjunction with the SMD solvation model. The results obtained are presented in Table 5. The findings indicate that employing the QZVP basis set is not essential for accurately calculating the pK_a parameters in the scheme proposed, as the less extensive 6-311++G(d,p) basis set also yields results consistent with the experimental data. However, an analysis of the $\Delta G(H^+)_{aq}$ parameter values reveals that the use of larger, more comprehensive basis sets is necessary to achieve result that align closely with the experimental one of $\Delta G(H^+)_{aq} = -259.5$ [kcal mol^{−1}] reported by Lim et al. [23].

Table 5. The effect of basis set on the theoretical reproduction of proton hydration energy $\Delta G(H^+)_{aq}$ and pK_a parameter values. The calculations were performed in the water medium at the LSDA/QZVP theory levels, using the SMD solvation model. The experimental value of $\Delta G(H^+)_{aq} = -259.5$ [kcal mol^{−1}] was reported by Lim et al. [23].

	6311++G(d,p)	aug-cc-pVQZ	QZVP
$\Delta G(H^+)_{aq}$	−257.1671(51)	−259.2358(72)	−259.4272(75)
a	7.156(34)	7.268(54)	7.271(57)
c	0.1934(19)	0.1869(29)	0.1858(31)
R ²	1.0000	1.0000	1.0000
SE	0.0265	0.0430	0.0454
pK_{a1}	4.159	4.160	4.161
pK_{a2}	8.542	8.537	8.536
pK_{a3}	11.423	11.435	11.436
pK_{a4}	12.788	12.778	12.776
MAE	0.011	0.017	0.019
NMAE	0.268	0.427	0.463

3. Materials and Methods

For monocarboxylic acids, the best reproduction of the pK_a values was achieved [5,6] using the composite G4CEP and DFT method in conjunction with the LSDA [36], M06-2X [37], and B3LYP [38,39] functionals, and the SMD solvation model as well as the Dunning's [40] basis sets aug-cc-pVDZ and aug-cc-pVTZ. In light of the findings presented in [5,6], in this work, calculations were conducted utilizing the DFT method implemented in Gaussian vs. 16 software [40] in conjunction with the LSDA, M062X, and B3LYP functionals. To identify a basis set capable of accurately reproducing the Gibbs energy for GA and its anionic forms, preliminary calculations were performed on the energy of GA in an aqueous environment, employing the DFT/LSDA method and the SMD solvation model, utilizing basis sets cc-pVQZ, aug-cc-pVQZ [40], and QZVP [41]. The resulting energy values (in [Ha]: $E_1(\text{cc-pVQZ}) = -643.496286$; $E_2(\text{aug-cc-pVQZ}) = -643.501176$; $E_3(\text{QZVP}) = -643.511057$), and their corresponding differences (in [kcal mol^{−1}]: $\Delta E_{12} = 3.0685$; $\Delta E_{23} = 6.2004$; $\Delta E_{13} = 9.2689$), guided the selection of the quadruple zeta valence polarized (QZVP) basis set [41] for subsequent calculations, as it yielded the lowest energy value for the GA. A similar situation was observed for the M062X and B3LYP functionals; consequently, the DFT method at the LSDA, M062X, and B3LYP/QZVP level of theory, combined with the SMD solvation model, was employed

for the primary calculations. The input structures were constructed by taking advantage of the Gauss View-6.1 graphical interface [42], and calculations were carried out using the Gaussian version 16 A software [42,43]. Since the computations for the compounds studied at the LSDA,M062X,B3LYP/QZVP level are time consuming and their convergence depends on the initial geometry input, the optimization was divided into two stages: (i) the determination of an approximate geometry at the LSDA,M062X,B3LYP/cc-pVQZ theory levels in the gas phase, and (ii) the final calculation at the LSDA,M062X,B3LYP/QZVP theory levels in the water medium using the SMD solvation model. Thus, optimization was performed at each stage, with the geometry determined at the lower level serving as the starting point for optimization at the higher level. The optimized structures of the compounds considered in this work are displayed in Figure 3. The parameters a , c , $G(H^+)_{aq}$, and $\Delta G(H^+)_{aq}$ were fitted to the four sets of experimental pK_a values for GA, including set I [4.50, 7.05, 8.75, 10.25] [11], set II [4.4(1), 8.8(1), 10.0(1), 11.4(1)] [12], set III [4.44, 8.54, 10.05, 11.3] [14], and set IV [4.16(2), 8.55(1), 11.4(1), 12.8(4)] [13]. The calculations were performed by taking advantage of Sigma Plot vs. 11 software—the results of calculations together with indicators of goodness of the fit, R^2 , coefficient of determination, and SE, standard error of estimation, are presented in Tables 2 and 4. The fitted parameters were utilized to calculate the pK_aN values for GA, as presented in Tables 3 and 4, along with the mean absolute errors (MAEs) and the normalized mean absolute errors (NMAEs) associated with the reproduction of the experimental $pK_aN(\text{exp})$ data, defined as follows

$$MAE = \frac{1}{4} \sum_{N=1}^4 |pK_aN - pK_aN(\text{exp})| \quad NMAE = \frac{1}{4} \sum_{N=1}^4 \frac{|pK_aN - pK_aN(\text{exp})|}{u_N} \quad (8)$$

Here, u_N denotes experimental errors in the measurements provided for datasets II ($u_N = 0.1$) [12] and IV ($u_1 = 0.02$, $u_2 = 0.01$, $u_3 = 0.10$, $u_4 = 0.40$) [13]. A value of $NMAE \leq 1$ indicates the reproduction of experimental data within the accuracy of the measurement error.

4. Conclusions

The proposed method for determining the pK_a parameters of GA enables the simultaneous validation of both the method used and the experimental data employed in the calculations. Among the four datasets analyzed, the non-linear model demonstrated the best agreement with dataset IV [13], which exhibited the highest statistical consistency when reproduced using the DFT method at the LSDA/QZVP theory level and the SMD solvation model. The parameters presented in Table 3 reproduce experimental dataset IV with a standard error, SE, of 0.0454 and a coefficient of determination, R^2 , of 1.0000, achieving reproduction of the experimental pK_a values [13] for GA within the limits of experimental errors as $NMAE = 0.463$ pK_a units. These findings confirm the validity of the self-similar Formula (7) in particular and the overall calculation methodology in general. Additionally, the reproduction of the experimental pK_a values for GA using relationship (7) has been proved to be more accurate than that obtained with Equation (4). The results obtained indicate that, for the multivariate deprotonation of GA, the linear transformation (4) should be replaced by the self-similar formula given in Equation (7). The application of Equation (7) in conjunction with the ΔG_N values calculated using the DFT method at the LSDA/QZVP theory level, alongside the SMD solvation model, is sufficient to achieve a satisfactory reproduction of the measured pK_a values of GA. Thus, the utilization of the computationally time-consuming G4CEP method and the SMD model with the inclusion of an additional water molecule, as suggested in works [5,6], is deemed unnecessary for this objective. It is acknowledged that the proposed approach represents a significant simplification of the real deprotonation process due to the complex reactions that occur between GA and water.

These reactions result in substantial modifications to the structure of GA [44] and its anionic forms, which in turn have a pronounced effect on the kinetics and thermodynamics of GA deprotonation. The proposed model is not derived from the first principles but constitutes a quasi-phenomenological framework, wherein the semi-empirical parameters a and c encapsulate all *hidden variables* that are not explicitly incorporated into the analysis. A first-principles model should, at the molecular level, account for all possible structures of GA, including polymorphs, rotamers, and tautomers and their anionic forms, as well as their interactions with solvent molecules and the formation of complex polymolecular systems. It will be essential to undertake research in this regard in the near future.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/molecules30030742/s1>, Table S1. The Gibbs and zero-point energies [Ha] of neutral GA^0 molecule and its anionic forms GA^{-N} $N = 1, 2, 3, 4$ calculated in the water medium at the LSDA, M062X, B3LYP/QZVP theory levels, using the SMD solvation model. Table S2. The effect of dispersion on the theoretical reproduction of proton hydration energy $\Delta G(\text{H}^+)_{\text{aq}}$ and $\text{pK}_{\text{a}N}$ parameter values. The calculations were performed in a water medium at the wB97XD/D2//QZVP, B3LYP/D3/QZVP theory levels, using the SMD solvation model. The theoretical value of $\Delta G(\text{H}^+)_{\text{aq}} = -264.29 \text{ [kcal mol}^{-1}\text{]}$ is reported by Zhan and Dixon [27]. For comparison, the results obtained at the B3LYP/QZVP level of theory without taking the dispersion effect into account are shown.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All new data created are reported in this work.

Acknowledgments: This research was supported by the computing grant PL0293-01 from Supercomputing and Networking Center (PCSS) in Poznań, Poland. I would like to thank Piotr Kopta from PCSS for technical support and Julian Molski for his constructive comments.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Gregory, M.T.; Gao, Y.; Cui, Q.; Yang, W. Multiple deprotonation paths of the nucleophile 3'-OH in the DNA synthesis reaction. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2103990118. [CrossRef] [PubMed]
2. Alongi, K.S.; Shields, G.S. Theoretical Calculations of Acid Dissociation Constants: A Review Article. *Annu. Rep. Comput. Chem.* **2010**, *6*, 113–138. [CrossRef]
3. Marino, T.; Galano, A.; Russo, N. Radical scavenging ability of gallic acid toward OH and OOH radicals. Reaction mechanism and rate constants from the density functional theory. *J. Phys. Chem. B* **2014**, *118*, 10380–10389. [CrossRef] [PubMed]
4. Molski, M. Theoretical study on the radical scavenging activity of gallic acid. *Heliyon* **2023**, *9*, e12806. [CrossRef]
5. De Souza Silva, C.d.S.; Custodio, R. Assessment of pK_{a} determination for monocarboxylic acids with an accurate theoretical composite method: G4CEP. *J. Phys. Chem. A* **2019**, *123*, 8314–8320. [CrossRef]
6. Dutra, F.R.; Silva, C.d.S.; Custodio, R. On the accuracy of the direct method to calculate pK_{a} from electronic structure calculations. *J. Phys. Chem. A* **2021**, *125*, 65–73. [CrossRef]
7. Walton-Raaby, M.; Floen, T.; García-Díez, G.; Mora-Díez, N. Calculating the aqueous pK_{a} of phenols: Predictions for antioxidants and cannabinoids. *Antioxidants* **2023**, *12*, 1420. [CrossRef]
8. Morency, M.; Néron, S.; Iftimie, R.; Wuest, J.D. Predicting pK_{a} values of quinols and related aromatic compounds with multiple OH groups. *J. Org. Chem.* **2021**, *86*, 14444–14460. [CrossRef]
9. Badhani, B.; Kakkar, R. Influence of intrinsic and extrinsic factors on the antiradical activity of gallic acid: A theoretical study. *Struct. Chem.* **2018**, *29*, 359–373. [CrossRef]
10. Masoud, M.S.; Ali, A.E.; Haggag, S.S.; Nasr, N.M. Spectroscopic studies on gallic acid and its azo derivatives and their iron(III) complexes. *Spectrochim. Acta A Mol. Biomol. Spectrosc.* **2014**, *120*, 505–511. [CrossRef]

11. Agrawal, M.D.; Bhandari, C.S.; Dixit, M.K.; Sogani, N.C. 3,4,5-Trihydroxybenzoic acid as Chelatebildner, 1. Mitt.: Praseodym. *Monatsh. Chem.* **1976**, *713*, 75–82. [[CrossRef](#)]
12. Huguenin, J.; Hamady, S.O.S.; Bourson, P. Monitoring deprotonation of gallic acid by Raman spectroscopy. *J. Raman Spectrosc.* **2015**, *46*, 1062–1066. [[CrossRef](#)]
13. Kipton, H.; Powell, J.; Taylor, M.C. Interactions of iron(II) and iron(III) with gallic acid and its homologues: A potentiometric and spectrophotometric study. *Aust. J. Chem.* **1982**, *35*, 739–756. [[CrossRef](#)]
14. Loginova, L.F.; Medyntsev, V.V.; Khomutov, B.I. Acidic properties of gallic acid and stability constants of iron complexes. *Zh. Obshch. Him.* **1972**, *42*, 739–742.
15. Cossi, M.; Rega, N.; Scalmani, G.; Barone, V. Energies, structures, and electronic properties of molecules in solution with the C-PCM solvation model. *J. Comput. Chem.* **2003**, *24*, 669–681. [[CrossRef](#)]
16. Tomasi, J.; Mennucci, B.; Cancès, E. The IEF version of the PCM solvation method: An overview of a new method addressed to study molecular solutes at the QM ab initio level. *J. Mol. Struct. THEOCHEM* **1999**, *464*, 211–226. [[CrossRef](#)]
17. Marenich, A.V.; Cramer, C.J.; 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](#)]
18. Xu, L.; Coote, M.L. Methods To Improve the Calculations of Solvation Model Density Solvation Free Energies and Associated Aqueous pK_a 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](#)]
19. Pezzola, S.; Venanzi, M.; Galloni, P.; Conte, V.; Sabuzi, F. Towards the “Eldorado” of pK_a Determination: A Reliable and Rapid DFT Model. *Molecules* **2024**, *29*, 1255. [[CrossRef](#)]
20. Bryantsev, V.S.; Diallo, M.S.; Goddard, W.A. Calculation of solvation free energies of charged solutes using mixed cluster/continuum models. *J. Phys. Chem. B* **2008**, *112*, 9709–9719. [[CrossRef](#)]
21. Kelly, C.P.; Cramer, C.J.; Truhlar, D.G. Adding explicit solvent molecules to continuum solvent calculations for the calculation of aqueous acid dissociation constants. *J. Phys. Chem. A* **2006**, *110*, 2493–2499. [[CrossRef](#)] [[PubMed](#)]
22. Fifen, J.J.; Dhaouadi, Z.; Nsangou, M. Revision of the thermodynamics of the proton in gas phase. *J. Phys. Chem. A* **2014**, *118*, 11090–11097. [[CrossRef](#)]
23. Lim, C.; Bashford, D.; Karplus, M. Absolute pK_a calculations with continuum dielectric methods. *J. Phys. Chem.* **1991**, *95*, 5610–5620. [[CrossRef](#)]
24. Tissandier, M.D.; Cowen, K.A.; Feng, W.Y.; Gundlach, E.; Cohen, M.H.; Earhart, A.D.; Coe, J.V.; Tuttle, T.R. The Proton’s Absolute Aqueous Enthalpy and Gibbs Free Energy of Solvation from Cluster-Ion Solvation Data. *J. Phys. Chem. A* **1998**, *102*, 7787–7794. [[CrossRef](#)]
25. Klots, C.E. Solubility of protons in water. *J. Phys. Chem.* **1981**, *85*, 3585–3588. [[CrossRef](#)]
26. Tawa, G.J.; Topol, I.A.; Burt, S.K.; Caldwell, R.A.; Rashin, A.A. Calculation of the aqueous solvation free energy of the proton. *J. Chem. Phys.* **1998**, *109*, 4852–4863. [[CrossRef](#)]
27. Zhan, C.-G.; Dixon, D.A. Absolute Hydration Free Energy of the Proton from First-Principles Electronic Structure Calculations. *J. Phys. Chem. A* **2001**, *105*, 11534–11540. [[CrossRef](#)]
28. Marković, Z.; Tošović, J.; Milenković, D.; Marković, S. Revisiting the solvation enthalpies and free energies of the proton and electron in various solvents. *Comput. Theor. Chem.* **2016**, *1077*, 11–17. [[CrossRef](#)]
29. Rimarčík, J.; Lukeš, V.; Klein, E.; Ilčin, M. Study of the solvent effect on the enthalpies of homolytic and heterolytic N–H bond cleavage in p-phenylenediamine and tetracyano-p-phenylenediamine. *J. Mol. Struct. THEOCHEM* **2010**, *952*, 25–30. [[CrossRef](#)]
30. Mejías, A.; Lago, S. Calculation of the absolute hydration enthalpy and free energy of H^+ and OH^- . *J. Chem. Phys.* **2000**, *113*, 7306–7316. [[CrossRef](#)]
31. Galano, A.; Pérez-González, A.; Castañeda-Arriaga, R.; Muñoz-Rugeles, L.; Mendoza-Sarmiento, G.; Romero-Silva, A.; Ibarra-Escutia, A.; Rebollar-Zepeda, A.M.; León-Carmona, J.R.; Hernández-Olivares, M.A.; et al. Empirically Fitted Parameters for Calculating pK_a Values with Small Deviations from Experiments Using a Simple Computational Strategy. *J. Chem. Inf. Model.* **2016**, *56*, 1714–1724. [[CrossRef](#)] [[PubMed](#)]
32. Pérez-González, A.; Castañeda-Arriaga, R.; Verastegui, B.; Carreón-González, M.; Alvarez-Idaboy, J.R.; Galano, A. Estimation of empirically fitted parameters for calculating pK_a values of thiols in a fast and reliable way. *Theor. Chem. Acc.* **2018**, *137*, 5. [[CrossRef](#)]
33. Miguel, E.L.M.; Silva, P.L.; Pliego, J.R. Theoretical Prediction of pK_a in Methanol: Testing SM8 and SMD Models for Carboxylic Acids, Phenols, and Amines. *J. Phys. Chem. B* **2014**, *118*, 5730–5739. [[CrossRef](#)] [[PubMed](#)]
34. Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620. [[CrossRef](#)]
35. Schröder, H.; Hühnert, J.; Schwabe, T. Evaluation of DFT-D3 dispersion corrections for various structural benchmark sets. *J. Chem. Phys.* **2017**, *146*, 044115. [[CrossRef](#)]

36. Parr, R.G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, NY, USA; Oxford, UK, 1989.
37. 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](#)]
38. Becke, A.D. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.* **1993**, *98*, 5648–5652. [[CrossRef](#)]
39. Lee, C.; Yang, W.; Parr, R.G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Phys. Rev. B* **1988**, *37*, 785–789. [[CrossRef](#)]
40. Dunning, T.H. Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007–1023. [[CrossRef](#)]
41. 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](#)]
42. 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, et al. *Gaussian 16*; Revision A.03; Gaussian, Inc.: Wallingford, CT, USA, 2016.
43. Glendening, E.D.; Reed, A.E.; Carpenter, J.E.; Weinhold, F. NBO Version 3.1. 2017.
44. Braun, D.E.; Bhardwaj, R.M.; Florence, A.J.; Tocher, D.A.; Price, S.L. Complex Polymorphic System of Gallic Acid—Five Monohydrates, Three Anhydrides, and over 20 Solvates. *Cryst. Growth Des.* **2013**, *13*, 19–23. [[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.