

RESEARCH

Open Access



Reliable and accurate prediction of basic pK_a values in nitrogen compounds: the pK_a shift in supramolecular systems as a case study

Jackson J. Alcázar^{1*}, Alessandra C. Misad Saide² and Paola R. Campodónico¹

Abstract

This article presents a quantitative structure–activity relationship (QSAR) approach for predicting the acid dissociation constant (pK_a) of nitrogenous compounds, including those within supramolecular complexes based on cucurbiturils. The model combines low-cost quantum mechanical calculations with QSAR methodology and linear regressions to achieve accurate predictions for a broad range of nitrogen-containing compounds. The model was developed using a diverse dataset of 130 nitrogenous compounds and exhibits excellent predictive performance, with a high coefficient of determination (R^2) of 0.9905, low standard error (s) of 0.3066, and high Fisher statistic (F) of 2142. The model outperforms existing methods, such as Chemaxon software and previous studies, in terms of accuracy and its ability to handle heterogeneous datasets. External validation on pharmaceutical ingredients, dyes, and supramolecular complexes based on cucurbiturils confirms the reliability of the model. To enhance usability, a script-like tool has been developed, providing a streamlined process for users to access the model. This study represents a significant advancement in pK_a prediction, offering valuable insights for drug design and supramolecular system optimization.

Keywords Acidity, QSAR, Quantum descriptors, DFT, Cucurbiturils

*Correspondence:

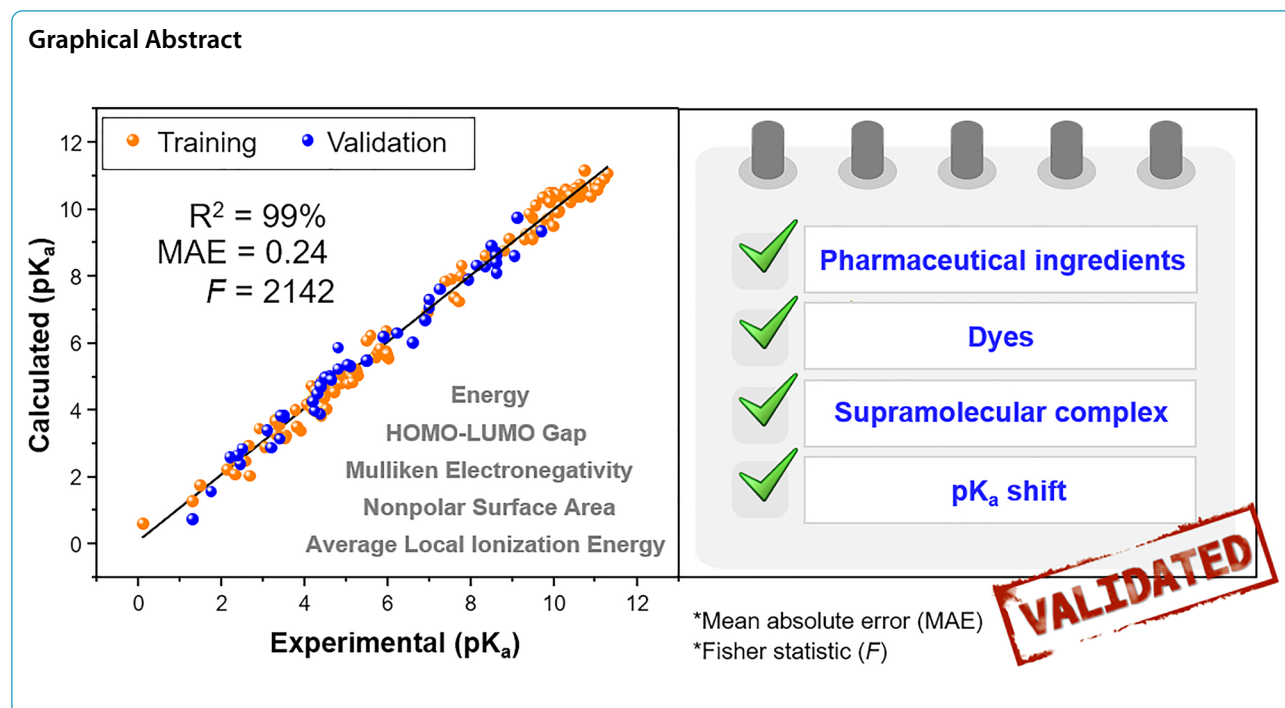
Jackson J. Alcázar

jackson.alcazar@udd.cl

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.



Introduction

The concepts of acidity and basicity are fundamental to the understanding of chemistry and have been defined by various theories throughout history [2, 18, 19, 54, 55]. One such theory was introduced by Svante Arrhenius in 1887 [2], who suggested that certain compounds can dissociate into ions in solution and identified acids as those that yield a proton (H⁺) and bases as those that yield a hydroxide ion (OH⁻). Another influential theory, the General Acid–Base Theory of Brønsted and Lowry [18, 19, 55], emerged in 1923, defining acidity and basicity in terms of the tendency to donate or accept a H⁺.

Understanding the strength of a base is crucial in comprehending its behavior and acidity plays a pivotal role in this regard. The strength of a base is commonly expressed by considering the strength of its conjugated acid, with a weaker conjugated acid indicating a stronger base. The acidity constant (K_a), which represents the equilibrium constant for the reaction between the acid and water, is used to quantitatively assess the strength of a base. For practical purposes, the pK_a value, defined as the negative logarithm of the K_a , is commonly employed [40]. The pK_a value is an essential tool that serves as an indicator of the relative acidity or basicity of a compound and enables predictions of its protonation state or protomeric forms under different pH conditions [40]. Therefore, accurate determination of pK_a holds immense significance across diverse fields, including medicinal chemistry [22, 29, 41, 60], biochemistry [11, 23, 62, 68], environmental science

[12, 15, 48, 50], chemistry of dyes [44, 65, 93] and supramolecular chemistry [10, 56, 88].

Recent research has focused on the phenomenon of supramolecular pK_a shift, which involves a significant shift in the pK_a value of nitrogenous compounds by forming supramolecular complexes with macrocyclic molecules such as cucurbiturils [7, 8, 14, 59, 64, 95]. This phenomenon is crucial for designing and optimizing supramolecular systems, with far-reaching implications in materials science [92, 98], catalysis [3, 28, 77] and the development of drugs and their delivery methods [26, 31, 37, 49, 75].

In the pharmaceutical industry, accurate pK_a values play a critical role in the design process of new drugs, as the acid/base character of a substance defines its biopharmaceutical properties, which have a direct impact on the pharmaceutical formulation of the drug [60, 61]. Nitrogen-containing heterocycles are of particular interest in the pharmaceutical industry due to their diverse biological activities [52, 53], with over 75% of FDA-approved drugs containing such structures [47]. The pK_a values of these heterocycles provide information on the absorption, distribution, metabolism, excretion, and toxicity (ADMET) of the drug, which are crucial for the design process. Poor ADMET properties have been revealed as the cause for high attrition in the development phase [84].

Experimental methodologies for pK_a determination have been extensively reviewed [74], but there are

still samples that are difficult or impossible to measure accurately. To overcome this challenge, computational approaches have emerged as a promising alternative, as they can simulate virtually any set of working conditions without requiring physical samples [81, 83, 94]. Consequently, significant efforts have been directed toward developing accurate and reliable computational methods for predicting pK_a values.

Among the most common computational approaches are First Principles and quantitative structure–activity relationship (QSAR) methods [81]. However, the accuracy of First Principles calculations in predicting pK_a values relies heavily on the precise determination of the Gibbs free energy difference in solution, which poses a significant challenge [27, 73, 81]. This is mainly due to the difficulty in calculating the Gibbs free energy of the proton and solvation energies, which can lead to deviations in pK_a values of up to 3 units [13, 73]. One way to address these systematic errors is by using the relative pK_a approach [73, 81], which has demonstrated high accuracy and effectiveness in various solvents [83]. Nevertheless, this approach is limited by the availability of accurate pK_a values for reference systems with structural similarity to the sample.

QSAR is a time-efficient and computationally less costly approach that predicts physical properties by constructing a multiple linear regression equation for a specific physical property as a function (P) of the selected molecular descriptors (X_i). The equation, represented as Eq. 1, assigns numerical coefficients (a_i) to each molecular descriptor, which serve as weighting factors to determine the respective contributions of the predictor variables [25, 81].

$$P = a_0 + a_1X_1 + a_2X_2 \quad (1)$$

Although QSAR models are less costly than First Principles, traditional QSAR methods have been hindered by lengthy calculation times, especially when quantum-mechanical electronic descriptors are involved, particularly in large molecules. As a result, the prediction of pK_a has been impractical. However, the B97-3c method, based on density functional theory (DFT), has recently emerged as a reliable and low-cost solution [16]. This method effectively reduces computational time, thereby presenting a viable option for predicting pK_a values in large molecules and supramolecular complexes.

QSAR models have extensively employed a wide range of descriptors to predict crucial properties, including pK_a values [36, 42, 45, 46, 70, 76, 79, 81, 90]. Some of these descriptors include charge [39, 45, 79], electronic energy differences (ΔE) [5, 6, 42, 79, 97], and the highest occupied molecular orbital energy (ϵ_{HOMO}) [80, 86]. Despite the flexibility in choosing chemical descriptors and their

combinations, reported QSAR models are limited to specific datasets or structures, resulting in varying levels of accuracy and model performance across different subsets of data. Such is the case of QSAR models for predicting pK_a values [96], which show diminished accuracy when predicting pK_a of basic compounds such as nitrogenous compounds [78]. These discrepancies may arise due to the lack of appropriate relationship or representation between the selected descriptors and the structural diversity present in the complete dataset. Consequently, developing reliable QSAR models for predicting pK_a values in heterogeneous datasets has proven to be a persistent challenge. To improve the generalizability of QSAR models and achieve higher precision in predicting pK_a values, a more careful selection of chemical descriptors that effectively capture the structural variability within the entire dataset is necessary. This may involve the use of more specific descriptors or advanced variable selection methods.

The present incapacity to efficiently and accurately predict basic pK_a values from heterogeneous data highlights the need for enhanced QSAR methodologies capable of achieving high accuracy regardless of the size and structure of the compound, as well as its inclusion in supramolecular host-guest systems. In light of this challenge, we propose and validate a comprehensive QSAR approach that utilizes the B97-3C low-cost density functional theory to predict the basic pK_a values of nitrogen-containing compounds in aqueous solution at 25 °C, both independently and within cucurbituril-based supramolecular complexes. This proposed approach represents a significant advancement toward predicting pK_a shifts in supramolecular systems.

Results and discussion

The QSAR model

By employing a comprehensive approach that integrates DFT [30], Conceptual Density Functional Theory (CDFT) [30], Molecular Electron Density Theory (MEDT) [24], and quantitative analysis of molecular surface, prediction models for estimating pK_a values were evaluated through statistical analysis. The best performing model was selected based on the results of this evaluation (see “Methods” section)

The selected prediction model is represented in Eq. 2.¹ The model was trained using a diverse set of 130 nitrogenous compounds, which encompassed aromatic and non-aromatic cyclic amines as well as aliphatic amines (primary, secondary, and tertiary). A coefficient

¹ For the full equation, complete with error coefficients, please refer to Additional file 1: Equation S1.

of determination (R^2) of 0.9905 indicates an excellent fit of the data to the proposed model. The robustness of the model is further supported by a high Fisher statistic (F) of 2141.9289 and a relatively low standard error (s) of 0.3066. The root mean squared error (RMSE), 0.2982, and mean absolute error (MAE), 0.2440, provide additional evidence of the predictive accuracy of the model.

$$pK_a = 0.1074\Delta E - 0.1422\Delta HL_{Gap} - 0.9132\chi_M + 0.0151\%NPSA - 1.4887\Delta ALIE_N + 3.0608BaseT - 30.7139 \quad (2)$$

$$n = 130; R^2 = 0.9905; s = 0.3066; F = 2141.9289; RMSE = 0.2982; MAE = 0.2440$$

Our model (Eq. 2) includes a variety of descriptors calculated in a vacuum environment and demonstrates remarkable performance when applied to a diverse set of nitrogenous compounds. The descriptors included in the model are the following:

- Energy of deprotonation (ΔE) in kcal/mol: this descriptor measures the energy required to remove a proton from an acid. A higher ΔE value signifies that a greater amount of energy is needed to carry out the deprotonation, which results in a higher pK_a value.
- HOMO–LUMO gap of deprotonation (ΔHL_{Gap}) in eV: this descriptor represents the change of energy gap between the highest occupied and lowest unoccupied orbitals of the acid–base equilibrium. A higher ΔHL_{Gap} suggests a less reactive base, hence a lower pK_a value.
- Mulliken electronegativity (χ_M) in eV: this descriptor quantifies the ability of the base to donate a pair of electrons and accept a proton. A lower χ_M indicates a higher basicity, which contributes to a higher pK_a value.
- Nonpolar surface area percentage (%NPSA) of the base: this descriptor measures hydrophobicity and its influence on the solubility of the base in water. Bases with a higher %NPSA tend to have lower solubility, which leads to increased stability of the conjugated acid in a polar environment compared to the base. This reduced solubility directly affects the ability of the base to donate and accept protons, thus altering the acid–base equilibrium and resulting in an increase in the pK_a value. Notably, the %NPSA descriptor demonstrates one of the most significant individual correlations with the experimental pK_a values, as evidenced by its correlation coefficient (r) of 0.7514 (refer to Additional file 1: Table S1).
- Change in average local ionization energy ($\Delta ALIE_N$): this descriptor quantifies the energy difference (in eV) required to remove an electron from

the nitrogen atom in the acid–base reaction center. A smaller $\Delta ALIE_N$ indicates greater stabilization of the positive charge in the acid, leading to an increased pK_a value

- Base type (BaseT): this categorical descriptor takes a value of 0 for aromatic amines and 1 for aliphatic or non-aromatic amines.

By considering such a comprehensive range of independent parameters (with a correlation between parameters $\leq |0.7744|$, see Additional file 1: Table S2), our model provides valuable insights into the electronic structure, stability, solubility, hydrophobicity, and local electronic effects during proton transfer in the bases under investigation. This holistic approach contributes to the accurate prediction of pK_a values and enhances our understanding of the underlying factors governing acid–base behavior in nitrogenous compounds.

In comparison with the trading software Chemaxon, our model shows a higher R^2 (0.9905 versus 0.9583), lower s (0.3066 versus 0.6346), and lower RMSE and MAE (see Additional file 1: Table S3). These results demonstrate the superiority of the presented approach over the widely used Chemaxon software.

Regarding the comparison of our model with the QSAR studies previously reported [36, 42, 45, 76, 79, 90], it is necessary to note that our approach surpasses all the explored QSAR methods in terms of accuracy and versatility. Namely, Tehan's [90] semi-empirical method using AM1, Seybold's [79] method based on RM1, Juranić's [45] approach with PM6, Gross's [36] method using Hartree-Fock (HF)/6-311 G(d,p), Sandoval-Lira's [76] DFT method based on ω B97X-D/cc-pVDZ, and Holt's [42] approach with B3LYP/6-31+G(d,p). Except for Gross's [36] method, which performs similarly to ours for anilines, all methods reported lower R^2 values and higher standard errors (see Fig. 1).

While the cited works focused on a particular class of amines (non-heterogeneous data) and selected up to three parameters, we have expanded the predictive capabilities to a broader set of nitrogenous compounds by incorporating a comprehensive range of descriptors. This approach provides a more accurate estimation of the pK_a value and greatly expands the range of applicability of our model.

Yu et al. [96] presented a comparative performance of ACD and SPARC software. While the statistical results of both software are similar to those of the present work, it is important to note that these software are commercial

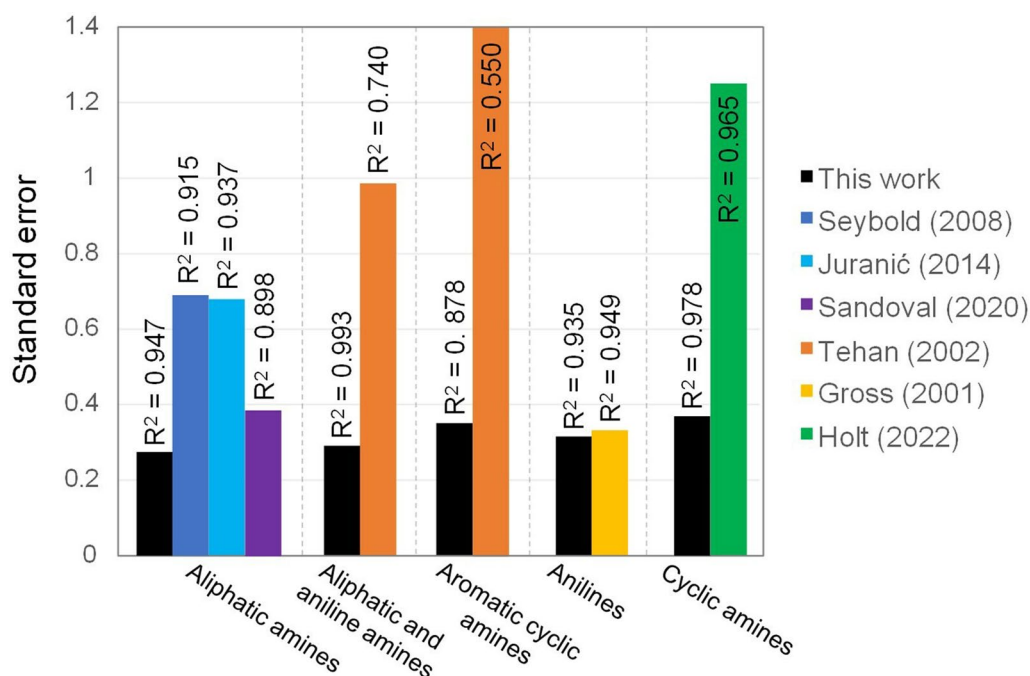


Fig. 1 Comparative performance of our model against previous QSAR methods [36, 42, 45, 76, 79, 90] based on R^2 values and standard errors (s)

and employ proprietary methodologies, which may limit their accessibility and customization. A key advantage of our model is its generalizability to diverse types of amines and its ability to handle heterogeneous data, which is a common challenge in real-world applications.

Model validation

External validation

The performance of our prediction model (Eq. 2), in estimating pK_a values, was evaluated a total of 40 compounds, which were part of an external validation dataset comprising pharmaceutical ingredients and dyes. It is important to note that the 40 compounds in the external validation dataset are not part of the training dataset of 130 compounds. Additionally, our method was tested on 6 cucurbituril-based supramolecular complexes, which were not included in either of the two previous datasets. These results are summarized in Tables 1 and 2, respectively.

According to data in Table 1, our model achieved a low external RMSE (RMSE_{ext}) of 0.32 and an external MAE (MAE_{ext}) of 0.28 when predicting pK_a values for the external validation dataset of 40 pharmaceutical ingredients and dyes. Comparing these values with the corresponding metrics obtained from Chemaxon (RMSE_{ext} of 1.09 and a MAE_{ext} of 0.79), it is evident that our model outperforms Chemaxon in terms of accuracy and precision in predicting pK_a values, especially when

estimating pK_a for supramolecular complexes, for which Chemaxon is currently incompatible.

For the dataset of 6 cucurbituril-based supramolecular complexes (Table 2), our model exhibited a slightly higher RMSE_{ext} of 0.54, indicating a moderate average deviation of the predicted pK_a values for this dataset. Similarly, the MAE_{ext} value of 0.42 suggests a moderate average error of the estimated pK_a values for the supramolecular complexes.

Internal validation

Furthermore, an internal validation assessment corroborates the excellent performance of our model. For internal validation within the training set of 130 compounds, we employed the “leave-one-out” cross-validation model. The small difference between the coefficient of determination (R^2) and the leave-one-out cross-validation correlation coefficient (Q^2_{loo}), indicated by the stability value of 0.0013, attests to the robustness of our model and suggests that it is not overfitted.

These results validate the effectiveness and reliability of our prediction model in estimating basic pK_a values of nitrogen-containing compounds, both as isolated species and as guests in cucurbiturils complexes. Further optimization and refinement of the model may enhance its performance in predicting pK_a values for diverse chemical systems, extending beyond nitrogenous compounds.

Table 1 Experimental and predicted basic pK_a values for nitrogen compounds in pharmaceutical ingredients and dyes at 25 °C in aqueous solution

ID	Nitrogenous compounds	Experimental pK_a	Predicted pK_a^a	Error	Chemaxon pK_a^b	Error	References
131	Norephedrine	9.12	9.72	0.60	8.96	0.16	[82]
132	Procaine	9.05	8.59	0.46	8.96	0.09	[82]
133	Terbutaline	8.62	8.14	0.48	9.76	1.14	[91]
134	Quinine	8.60	8.69	0.09	8.55	0.05	[91]
135	Tetracaine	8.50	8.87	0.37	8.42	0.08	[82]
136	Astemizole	8.35	8.25	0.10	9.23	0.88	[91]
137	Nicotine	8.14	8.32	0.18	8.58	0.44	[82]
138	Clozapine	7.94	7.85	0.09	8.16	0.22	[91]
139	Ampicillin	7.25	7.60	0.35	7.23	0.02	[78]
140	Ketoconazole	6.22	6.25	0.03	6.26	0.04	[91]
141	Clotrimazole	5.89	6.17	0.28	6.42	0.53	[43]
142	Benzimidazole	5.50	5.47	0.03	5.79	0.29	[67]
143	Coumarin 7	5.10	5.26	0.16	4.21	0.89	[9]
144	Abacavir	5.04	5.35	0.31	6.87	1.83	[82]
145	Fuberidazole	4.80	5.21	0.41	3.93	0.87	[51]
146	Thiabendazole	4.60	5.00	0.40	4.08	0.52	[51]
147	Carbendazim	4.50	4.98	0.48	4.28	0.22	[51]
148	Omeprazole	4.40	4.71	0.31	4.77	0.37	[91]
149	Cerivastatin	4.38	4.79	0.41	5.58	1.20	[91]
150	Ethionamide	4.37	3.89	0.48	5.00	0.63	[91]
151	Olmesartan medoxomil	4.30	4.48	0.18	3.65	0.65	[4]
152	Lamivudine	4.24	4.01	0.23	2.00	2.24	[82]
153	Picoprazole	3.50	3.78	0.28	2.89	0.61	[63]
154	2-Amino-pentamethylbodipy	3.50	3.77	0.27	0.41	3.09	[38]
155	Diazepam	3.42	3.84	0.42	2.92	0.50	[89]
156	Prodan	3.40	3.14	0.26	4.94	1.54	[21]
157	Timoprazole	3.10	3.41	0.31	2.38	0.72	[63]
158	Benzocaine	2.45	2.39	0.06	2.78	0.33	[91]
159	Metronidazole	2.38	2.65	0.27	3.03	0.65	[82]
160	Acyclovir	2.20	2.62	0.42	2.94	0.74	[82]
161	Sulfacetamide	1.75	1.58	0.17	2.14	0.39	[67]
162	7-Dimethylazacoumarin ^c	1.30	0.73	0.57	3.67	2.37	[1]
163	CHEMBL1689126 ^c	7.00	7.07	0.07	6.54	0.46	[63]
164	CHEMBL1689112 ^c	6.90	6.68	0.22	5.94	0.96	[63]
165	CHEMBL47529 ^c	5.50	5.48	0.02	4.96	0.54	[63]
166	CHEMBL1349378 ^c	4.64	4.90	0.26	4.49	0.15	[63]
167	CHEMBL432733 ^c	4.37	4.72	0.35	7.39	3.02	[63]
168	CHEMBL271703 ^c	4.20	4.26	0.06	3.46	0.74	[63]
169	CHEMBL191553 ^c	3.20	2.89	0.31	2.80	0.40	[63]
170	CHEMBL1405150 ^c	2.50	2.86	0.36	3.64	1.14	[63]
			MAE_ext:	0.28	MAE_ext:	0.60	

^a Calculated by Eq. 2. RMSE_ext = 0.32^b Calculated by Chemaxon software. RMSE_ext = 1.09^c Does not have a common or short name. The IUPAC name is found in Additional file 1

Table 2 Predicted pK_a values of nitrogen compounds in CB7-complexed states at 25°C in aqueous solution (Predicted pK_a^{CB7}) and their respective pK_a shift

ID	Nitrogenous compounds	Exp. pK_a^{CB7}	Predicted ^a pK_a^{CB7}	Error	Exp. pK_a Shift	Predicted ^b pK_a Shift	Error	References
C1	Coumarin 7	9.70	9.36	0.34	4.60	4.06	0.54	[9]
C2	Thiabendazole	8.60	8.62	0.02	3.96	3.60	0.36	[51]
C3	Fuberidazole	8.60	8.41	0.19	3.80	3.17	0.63	[51]
C4	Carbendazim	7.00	7.31	0.31	2.50	2.35	0.15	[51]
C5	Prodan	6.60	6.02	0.58	3.20	2.84	0.36	[21]
C6	2-Amino-pentamethylbodipy	4.80	5.87	1.07	1.30	2.03	0.73	[38]
		MAE_ext:	0.42		MAE_ext:	0.46		

^a Calculated by Eq. 2. RMSE_ext = 0.54

^b Calculated pK_a Shift using the Eq. 2 for both the free and complexed substrate. RMSE_ext = 0.50

Script-like tool description

To enhance user experience with our model, we have developed a script-like tool that automates the determination of descriptors and pK_a values for nitrogenous compounds. Users can easily access the estimation process through our tool by inputting the structures of the base and the conjugate acid. The tool is available at the following link: <https://github.com/Jacksonalcazar/Basic-pKa-Estimation-Nitrogen-Compounds>. This streamlined process provided by our tool offers users a more convenient and efficient way to utilize our model.

Conclusion

The comprehensive QSAR approach presented in this article offers a powerful tool for rapidly and accurately predicting pK_a values of nitrogenous compounds, including those within supramolecular complexes based on cucurbiturils. Our model, which combines quantum mechanical calculations and QSAR methodology, exhibits excellent predictive performance and provides valuable insights into various molecular properties relevant to proton transfer. The superiority of our approach over existing methods has been demonstrated through extensive comparisons. Furthermore, we have developed a user-friendly script-like tool that streamlines the determination of descriptors and pK_a values, enhancing the accessibility and practicality of our model. This work represents a significant advancement in the field of pK_a prediction and holds great potential for applications in drug discovery, supramolecular chemistry, and other related disciplines. Through further optimization and refinement, the model can extend its predictive capabilities to diverse chemical systems beyond nitrogenous compounds.

Methods

Prediction of pK_a values of nitrogenous compounds using QSAR approach requires the calculation of relevant descriptors related to basicity or acidity. Chemical descriptors can be local or global, depending on whether they are related to a specific atom or group of atoms within the molecule or to the molecule as a whole. However, obtaining a comprehensive understanding of the electronic structure of a molecule and its acidity requires an advanced computational approach able to provide a detailed picture of the molecular parameters and properties relevant for predicting pK_a values. In this section, we describe a comprehensive methodology for predicting basic pK_a values of nitrogenous compounds.

Prediction of pK_a values using chemical descriptors

We employed a multi-faceted approach, using DFT as a fundamental computational method to determine electronic properties. DFT is widely used in quantum chemistry, since it is regarded as a reliable approach to predict molecular properties and structures [20].

However, DFT alone may not provide a complete picture of the molecular electronic structure. Therefore, additional post-processing techniques were employed. CDFT [30] was used to obtain a broader perspective of the molecular electronic structure by making use of global and local descriptors based on the conceptualization of electron density. Moreover, MEDT [24] and quantitative analysis of the molecular surface [58] were used to obtain properties that account for the electronic distribution of the molecule.

To obtain a comprehensive set of chemical descriptors, a combination of DFT, CDFT, MEDT, and quantitative analyses of molecular surface was used. Subsequently, a predictive model for pK_a values was developed, using a diverse set

of 130 training compounds in aqueous solution at 25 °C, which included cyclic amines (aromatic and non-aromatic) and aliphatic amines (primary, secondary, and tertiary).

The most relevant descriptors for the prediction of pK_a values were identified using QSARINS software for statistical analysis [32, 33]. QSARINS is a powerful tool for identifying the molecular descriptors that contribute most significantly to the predictive power of the model. QSARINS uses iterative techniques to add or remove descriptors from a multivariable linear equation, based on their statistical significance, to create a set of models from which the most suited is selected. The performance of the selected model was evaluated by testing it on a set of 40 validation compounds, which were not part of the training set of 130 compounds. This set included pharmaceutical ingredients and dyes with known pK_a values.

Determination of selected descriptors for pK_a prediction

This section outlines the methodology for calculating the parameters involved in determining the pK_a values according to Eq. 2 in the main section. The parameters include deprotonation energy (ΔE), HOMO–LUMO deprotonation gap (ΔHL_{Gap}), Mulliken electronegativity (χ_M), the percentage of nonpolar surface area (%NPSA), and the change in average local ionization energy at the nitrogen atom ($\Delta ALIE_N$).

To calculate these parameters swiftly and reliably, DFT calculations were run on optimized geometries under vacuum conditions. The B97-3c low-cost Density Functional Method [16] and the ORCA software package (Program Version 5.0.3) [66] were employed for this purpose. A more detailed description of the methodology used for each parameter is provided in the following sections.

Energy of deprotonation (ΔE)

The electronic structure optimization and energy calculation for determining the energy of deprotonation (ΔE) were carried out using the ORCA software package [66] with the B97-3c low-cost functional to obtain the most stable conformation at a local minimum of the base and conjugate acid, swiftly and reliably [16]. Protocol tightSCF was employed to ensure convergence. The total energy of the molecule was calculated, taking into account the Becke–Johnson dispersion damping (DFT-D3BJ) [34, 35] and short-range basis incompleteness SRB correction of the basis set [16, 87]. The energy change during the deprotonation process of the conjugate acid was calculated as the difference between the total energy of the base and the total energy of the conjugate acid:

$$\Delta E = \text{Total energy of base} - \text{Total energy of conjugate acid} \quad (3)$$

It is important to note that this methodology requires the base to have a net charge of zero and the conjugate acid to have a net charge of +1. If the base is an ion, it should be neutralized with its respective counterion (Cl^- or Na^+).

HOMO–LUMO gap of deprotonation (ΔHL_{Gap})

The HOMO–LUMO gap (HL_{Gap}) is a crucial parameter for characterizing the electronic properties of a system. The HOMO and LUMO energies were automatically determined at the end of the electronic structure optimization in the previous step. Specifically, the HOMO and LUMO energies were obtained from the eigenvalues of the highest occupied and lowest unoccupied molecular orbitals, respectively. Subsequently, the HOMO–LUMO gap energy was calculated by subtracting the LUMO energy from the HOMO energy.

$$HL_{Gap} = \epsilon_{LUMO} - \epsilon_{HOMO} \quad (4)$$

Thus, the variation of the HL_{Gap} in the deprotonation process of the conjugate acid was calculated as:

$$\Delta HL_{Gap} = HL_{Gap} \text{ of base} - HL_{Gap} \text{ of conjugate acid} \quad (5)$$

Mulliken electronegativity (χ_M)

Quantifies the ability of the base to donate a pair of electrons and accept a proton. χ_M was determined by Eq. 6 [69, 72]

Thus, the variation of the HL_{Gap} in the deprotonation process of the conjugate acid was calculated as:

$$\chi_M = \frac{1}{2}(VIP + VEA) \quad (6)$$

Where VIP and VEA are the vertical ionization potential and vertical electron affinity of the base, respectively. The energy of the neutral molecule was calculated by single point from optimized base using the same quantum chemistry software package and level of theory employed in the preceding calculations. Next, the energy of the cation ($N-1$) was obtained by removing an electron from the neutral molecule (N) using the same software package and level of theory, setting the charge of the molecule to +1 in the input file. The VIP was calculated as the energy difference between the cation (E_{N-1}) and the neutral molecule (E_N) [17],

$$VIP = E_{N-1} - E_N \quad (7)$$

The energy of the anion ($N+1$) was obtained by adding an electron to the neutral molecule (N) using the same

software package and level of theory, setting the charge of the molecule to -1 in the input file. The VEA was calculated as the energy difference between the neutral molecule (E_N) and the anion (E_{N+1}) [17],

$$VEA = E_N - E_{N+1} \quad (8)$$

Nonpolar surface area percentage (%NPSA)

The percentage of nonpolar surface area of the base was calculated using Multiwfn version 3.8 [57], a software package for post-processing wavefunction analysis, with an improved Marching Tetrahedra algorithm [58]. The molecular structure of the base was loaded into the software in Gaussian Binary Wavefunction format (.gbw) and analyzed using the “quantitative analysis of molecular surface” function with electrostatic potential (ESP) as the mapped function. The analysis was conducted under default settings, with an electron density contour value of 0.00100 used to define the isovalue of the electron density surface. The grid point spacing of 0.250000 was selected for generating the molecular surface, and the ratio of van der Waals radius was set to 1.7000 to extend the spatial region of cubic grids, which determines the size of the molecular surface by expanding the van der Waals radii of the atoms in the molecule.

Change in average local ionization energy at nitrogen atom ($\Delta ALIE_N$)

The reactivity of the acid–base reaction center was investigated using the concept of Average Local Ionization Energy (ALIE) [71, 85]. To compute the ALIE values for the nitrogen atoms in both the base and conjugate acid, Eq. 9 was employed.

$$ALIE_N = \sum_i \rho_i(N) \frac{|\epsilon_i|}{\rho(N)} \quad (9)$$

where $\rho_i(N)$ denotes the density of the i -th orbital of the nitrogen atom, ϵ_i refers to the corresponding orbital energy, and $\rho(N)$ denotes the total electron density on the nitrogen atom. The calculations were performed using Multiwfn software (version 3.8) [57] by importing the optimized molecular structures in.gbw format.

The difference between the $ALIE_N$ of the conjugate acid and the $ALIE_N$ of the base was calculated to determine $\Delta ALIE_N$, which provides a quantitative measure of the change in the electronic structure and potential energy of the nitrogen atom upon the acid–base reaction. The parameter $\Delta ALIE_N$ was calculated using the following equation:

$$\Delta ALIE_N = ALIE_N \text{ of base} - ALIE_N \text{ of acid} \quad (10)$$

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13321-023-00763-3>.

Additional file 1. Optimized structures of the investigated compounds, equations, tables and additional figures.

Acknowledgements

JJA. thanks the Vicerectoría de Investigación y Doctorado (VRID), Universidad de Desarrollo, for a postdoctoral fellowship. The authors acknowledge FONDEQUIP EQM150093 for computational resources and the Instituto de Ciencias e Innovación en Medicina (ICIM), Facultad de Medicina, Universidad de Desarrollo, for general support.

Author contributions

JJA Formal analysis, investigation, project administration and lead of contribution; JJA and PRC Validation of results; JJA, ACMS and PRC Data curation; JJA and ACMS Writing of original draft. All authors reviewed the manuscript.

Funding

Not applicable.

Availability of data and materials

We have made the inputs used in this study publicly available, as well as a script that can estimate the basic pKa based on the structure of the base and the conjugate acid. You can access the inputs and the script at the following GitHub repository: <https://github.com/Jacksonalcazar/Basic-pKa-Estimation-Nitrogen-Compounds>.

Code availability

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Centro de Química Médica, Universidad del Desarrollo, Av.Plaza 680, 7780272 Santiago, RM, Chile. ²Santiago, RM, Chile.

Received: 29 June 2023 Accepted: 20 September 2023

Published online: 28 September 2023

References

- Alcázar JJ, Márquez E, García-Río L et al (2022) Changes in protonation sites of 3-styryl derivatives of 7-(dialkylamino)-aza-coumarin dyes induced by cucurbit[7]uril. *Front Chem*. <https://doi.org/10.3389/fchem.2022.870137>
- Arrhenius S (1887) Über die Dissociation der in Wasser gelösten Stoffe. *Zeitschrift für Physikalische Chemie* 1(1):631–648. <https://doi.org/10.1515/zpch-1887-0164>
- Assaf KI, Nau WM (2015) Cucurbiturils: from synthesis to high-affinity binding and catalysis. *Chem Soc Rev* 44(2):394–418. <https://doi.org/10.1039/c4cs00273c>
- Bajerski L, Rossi RC, Dias CL et al (2010) Development and validation of a discriminating in vitro dissolution method for a poorly soluble drug, Olmesartan Medoxomil: comparison between commercial tablets. *AAPS PharmSciTech* 11(2):637–644. <https://doi.org/10.1208/s12249-010-9421-0>

5. Baldasare CA, Seybold PG (2020) Computational estimation of the gas-phase and aqueous acidities of carbon acids. *ACS Appl Mater Interfaces*. <https://doi.org/10.1021/acs.jpca.9b11964>
6. Baldasare CA, Seybold PG (2021) Computational estimation of the aqueous acidities of alcohols, hydrates, and enols. *J Phys Chem A* 125(17):3600–3605. <https://doi.org/10.1021/acs.jpca.1c01330>
7. Barooh N, Mohanty J, Pal H et al (2012) Stimulus-responsive supramolecular pKa tuning of cucurbit[7]uril encapsulated coumarin 6 dye. *J Phys Chem B* 116(12):3683–3689. <https://doi.org/10.1021/jp212459r>
8. Barooh N, Mohanty J, Pal H et al (2014) Cucurbituril-induced supramolecular pKa shift in fluorescent dyes and its prospective applications. *Proc Natl Acad Sci India Sect A Phys Sci* 84(1):1–17. <https://doi.org/10.1007/s40010-013-0101-9>
9. Barooh N, Sundararajan M, Mohanty J et al (2014) Synergistic effect of intramolecular charge transfer toward supramolecular pKa shift in cucurbit[7]uril encapsulated coumarin dyes. *J Phys Chem B* 118(25):7136–7146. <https://doi.org/10.1021/jp501824p>
10. Barooh N, Mohanty J, Bhasikuttan AC (2022) Cucurbituril-based supramolecular assemblies: prospective on drug delivery, sensing, separation, and catalytic applications. *Langmuir*. <https://doi.org/10.1021/acs.langmuir.2c00556>
11. Berg JM, Tymoczko JL, Gatto GJ et al (2019) *Biochemistry*, 9th edn. W.H. Freeman and Company, New York
12. Bernhardsen IM, Knuutila HK (2017) A review of potential amine solvents for CO₂ absorption process: absorption capacity, cyclic capacity and pKa. *Int J Greenhouse Gas Control* 61:27–48. <https://doi.org/10.1016/j.jggc.2017.03.021>
13. Bodnarchuk MS, Heyes DM, Dini D et al (2014) Role of deprotonation free energies in pKa prediction and molecule ranking. *J Chem Theory Comput* 10(6):2537–2545. <https://doi.org/10.1021/ct400914w>
14. Bojesomo RS, Saleh N (2022) Photoinduced electron transfer in encapsulated heterocycles by cavitands. *Photochem Photobiol* 98(4):754–762. <https://doi.org/10.1111/php.13571>
15. Bond T, Templeton MR, Graham N (2012) Precursors of nitrogenous disinfection by-products in drinking water—a critical review and analysis. *J Hazard Mater* 235–236:1–16. <https://doi.org/10.1016/j.jhazmat.2012.07.017>
16. Brandenburg JG, Banwarth C, Hansen A et al (2018) B97–3c: a revised low-cost variant of the B97–D density functional method. *J Chem Phys* 148(6):064104. <https://doi.org/10.1063/1.5012601>
17. Bredas JL (2014) Mind the gap! *Mater Horizons* 1(1):17–19. <https://doi.org/10.1039/c3mh00098b>
18. Brønsted JN (1934) Zur Theorie der Säuren und Basen und der protolytischen Lösungsmittel. *Zeitschrift für Physikalische Chemie* 169(1):52–74. <https://doi.org/10.1515/zpch-1934-16906>
19. Brønsted JN, Pedersen K (1924) Die katalytische Zersetzung des Nitramids und ihre physikalisch-chemische Bedeutung. *Zeitschrift für Physikalische Chemie* 108(1):185–235. <https://doi.org/10.1515/zpch-1924-10814>
20. Burke K (2012) Perspective on density functional theory. *J Chem Phys* 136(15):150901. <https://doi.org/10.1063/1.4704546>
21. Chandra F, Pal K, Lathwal S et al (2016) Supramolecular guest relay using host-protein nanocavities: an application of host-induced guest protonation. *Mol BioSyst* 12(9):2859–2866. <https://doi.org/10.1039/c6mb00423g>
22. Das B, Baidya AT, Mathew AT et al (2022) Structural modification aimed for improving solubility of lead compounds in early phase drug discovery. *Bioorg Med Chem* 56:116614. <https://doi.org/10.1016/j.bmc.2022.116614>
23. Di Costanzo L, Panunzi B (2021) Visual pH sensors: from a chemical perspective to new bioengineered materials. *Molecules* 26(10):2952. <https://doi.org/10.3390/molecules26102952>
24. Domingo LR (2016) Molecular electron density theory: a modern view of reactivity in organic chemistry. *Molecules* 21(10):1319. <https://doi.org/10.3390/molecules21101319>
25. Dörgö G, Péter Hamadi O, Varga T et al (2020) Mixtures of QSAR models: learning application domains of pKa predictors. *J Chemometr* 34(4):e3223
26. El-Sheshtawy HS, Chatterjee S, Assaf KI et al (2018) A supramolecular approach for enhanced antibacterial activity and extended shelf-life of fluoroquinolone drugs with cucurbit[7]uril. *Sci Rep* 8(1):1–10. <https://doi.org/10.1038/s41598-018-32312-6>
27. Fujiki R, Matsui T, Shigeta Y et al (2021) Recent developments of computational methods for pKa prediction based on electronic structure theory with solvation models. *J* 4(4):849–864. <https://doi.org/10.3390/j4040058>
28. Funk S, Schatz J (2020) Cucurbiturils in supramolecular catalysis. *J Incl Phenomena Macroscopic Chem* 96(1–2):1–27. <https://doi.org/10.1007/s10847-019-00956-0>
29. Gaohua L, Miao X, Dou L (2021) Crosstalk of physiological pH and chemical pKa under the umbrella of physiologically based pharmacokinetic modeling of drug absorption, distribution, metabolism, excretion, and toxicity. *Expert Opin Drug Metab Toxicol* 17(9):1103–1124. <https://doi.org/10.1080/17425255.2021.1951223>
30. Geerlings P, De Proft F, Langenaeker W (2003) Conceptual density functional theory. *Chem Rev* 103(5):1793–1873. <https://doi.org/10.1021/cr99029p>
31. Ghosh I, Nau WM (2012) The strategic use of supramolecular pKa shifts to enhance the bioavailability of drugs. *Adv Drug Deliv Rev* 64(9):764–783. <https://doi.org/10.1016/j.addr.2012.01.015>
32. Gramatica P (2020) Principles of QSAR modeling. *Int J Quant Struct Prop Relationships* 5(3):61–97
33. Gramatica P, Chirico N, Papa E et al (2013) QSARINS: a new software for the development, analysis, and validation of QSAR MLR models. *J Comput Chem* 34(24):2121–2132. <https://doi.org/10.1002/jcc.23361>
34. Grimme S, Antony J, Ehrlich S et al (2010) A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H–Pu. *J Chem Phys* 132(15):154104. <https://doi.org/10.1063/1.3382344>
35. Grimme S, Ehrlich S, Goerigk L (2011) Effect of the damping function in dispersion corrected density functional theory. *J Comput Chem* 32(7):1456–1465. <https://doi.org/10.1002/jcc.21759>
36. Gross KC, Seybold PG, Peralta-Inga Z et al (2001) Comparison of quantum chemical parameters and Hammett constants in correlating pKa values of substituted anilines. *J Org Chem* 66(21):6919–6925. <https://doi.org/10.1021/jo010234g>
37. Gu A, Wheate NJ (2021) Macrocycles as drug-enhancing excipients in pharmaceutical formulations. *J Incl Phenomena Macroscopic Chem* 100(1–2):55–69. <https://doi.org/10.1007/s10847-021-01055-9>
38. Gupta M, Parvathi K, Mula S et al (2017) Enhanced fluorescence of aqueous BODIPY by interaction with cavitand cucurbit[7]uril. *Photochem Photobiol Sci* 16(4):499–506. <https://doi.org/10.1039/C6PP00325G>
39. Haslak ZP, Zareb S, Dogan I et al (2021) Using atomic charges to describe the pKa of carboxylic acids. *J Chem Inf Model*. <https://doi.org/10.1021/acs.jcim.1c00059>
40. Himmel D, Radtke V, Butschke B et al (2018) Basic Remarks on Acidity. *Angew Chem Int Ed* 57(16):4386–4411. <https://doi.org/10.1002/anie.201709057>
41. Holovach S, Melnykov KP, Skreminskiy A et al (2022) Effect of gem-difluorination on the key physicochemical properties relevant to medicinal chemistry: the case of functionalized cycloalkanes. *Chem A Eur J* 28(19):e202200331. <https://doi.org/10.1002/chem.202200331>
42. Holt RA, Seybold PG (2022) Computational estimation of the acidities of pyrimidines and related compounds. *Molecules* 27(2):385. <https://doi.org/10.3390/molecules27020385>
43. Hsieh YL, Ilevbare GA, Van Erdenbrugh B et al (2012) PH-induced precipitation behavior of weakly basic compounds: determination of extent and duration of supersaturation using potentiometric titration and correlation to solid state properties. *Pharm Res* 29(10):2738–2753. <https://doi.org/10.1007/s11095-012-0759-8>
44. Hunger K (2002) *Industrial dyes*, 1st edn. Wiley-VCH, Weinheim. <https://doi.org/10.1002/3527602011>
45. Juranić I (2014) Simple method for the estimation of pKa of amines. *Croatica Chem Acta* 87(4):343–347. <https://doi.org/10.5562/cca2462>
46. Karelson M, Lobanov VS, Katritzky AR (1996) Quantum-chemical descriptors in QSAR/QSPR studies. *Chem Rev* 96(3):1027–1043. <https://doi.org/10.1021/cr950202r>
47. Kerru N, Gummidi L, Maddila S et al (2020) A review on recent advances in nitrogen-containing molecules and their biological applications. *Molecules* 25(8):1909. <https://doi.org/10.3390/molecules25081909>
48. Khalili F, Rayer AV, Henni A et al (2012) Kinetics and dissociation constants (pKa) of polyamines of importance in post-combustion carbon dioxide (CO₂) capture studies. *ACS Symp Ser* 1097:43–70. <https://doi.org/10.1021/bk-2012-1097.ch003>

49. Khurana R, Barooah N, Bhasikuttan AC et al (2017) Modulation in the acidity constant of acridine dye with cucurbiturils: stimuli-responsive pKa tuning and dye relocation into live cells. *Org Biomol Chem* 15(39):8448–8457. <https://doi.org/10.1039/c7ob02135f>
50. Kim MK, Zoh KD (2016) Occurrence and removals of micropollutants in water environment. *Environ Eng Res* 21(4):319–332. <https://doi.org/10.4491/eer.2016.115>
51. Koner AL, Ghosh I, Saleh N et al (2011) Supramolecular encapsulation of benzimidazole-derived drugs by cucurbit[7]uril. *Can J Chem* 89(2):139–147. <https://doi.org/10.1139/V10-079>
52. Kumar A, Singh AK, Singh H et al (2023) Nitrogen containing heterocycles as anticancer agents: a medicinal chemistry perspective. *Pharmaceuticals* 16(2):299. <https://doi.org/10.3390/ph16020299>
53. Lang DK, Kaur R, Arora R et al (2020) Nitrogen-containing heterocycles as anticancer agents: an overview. *Anti-Cancer Agents Med Chem* 20(18):2150–2168. <https://doi.org/10.2174/1871520620666200705214917>
54. Lewis GN (1916) The atom and the molecule. *J Am Chem Soc* 38(4):762–785. <https://doi.org/10.1007/s12045-019-0841-1>
55. Lowry TM (1923) The uniqueness of hydrogen. *J Soc Chem Ind* 42(3):43–47. <https://doi.org/10.1002/jctb.5000420302>
56. Loya JD, Li SJ, Nuruh DK et al (2019) Application of the pKa rule to synthesize salts of bezafibrate. *Supramol Chem* 31(8):558–564. <https://doi.org/10.1080/10610278.2019.1635695>
57. Lu T, Chen F (2012) Multiwfn: a multifunctional wavefunction analyzer. *J Comput Chem* 33(5):580–592. <https://doi.org/10.1002/jcc.22885>
58. Lu T, Chen F (2012) Quantitative analysis of molecular surface based on improved Marching Tetrahedra algorithm. *J Mol Graph Model* 38:314–323. <https://doi.org/10.1016/j.jmgm.2012.07.004>
59. Macartney DH (2018) Cucurbit[n]uril host-guest complexes of acids, photoacids, and super photoacids. *Israel J Chem* 58(3):230–243. <https://doi.org/10.1002/ijch.201700096>
60. Manalack DT, Prankerd RJ, Yuriev E et al (2013) The significance of acid/base properties in drug discovery. *Chem Soc Rev* 42(2):485–496. <https://doi.org/10.1039/c2cs35348b>
61. Manjooran G (2020) pKa and ka (Acid dissociation constant). *Southern Afr J Anaesth Analgesia* 26(6):108. <https://doi.org/10.36303/SAJAA.2020.26.6.S3.2552>
62. Marunaka Y (2021) Roles of interstitial fluid pH and weak organic acids in development and amelioration of insulin resistance. *Biochem Soc Trans* 49(2):715–726. <https://doi.org/10.1042/BST20200667>
63. Mendez D, Gaulton A, Bento AP et al (2019) ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res* 47(D1):D930–D940. <https://doi.org/10.1093/nar/gky1075>
64. Mohanty J, Barooah N, Bhasikuttan AC (2021) Effect of confinement on the physicochemical properties of chromophoric dyes/drugs with cucurbit[n]uril: prospective applications. *Chemical reactivity in confined systems: theory, modelling and applications*. John Wiley & Sons, Ltd, pp 371–393. <https://doi.org/10.1002/9781119683353.ch19>
65. More KN, Mun SK, Kang J et al (2021) Molecular design of fluorescent pH sensors based on reduced rhodol by structure-pKa relationship for imaging of lysosome. *Dyes and pigments* 184:108785. <https://doi.org/10.1016/j.dyepig.2020.108785>
66. Neese F, Wennmohs F, Becker U et al (2020) The ORCA quantum chemistry program package. *J Chem Phys* 152(22):224108. <https://doi.org/10.1063/5.0004608>
67. O'Neil MJ (2013) *The Merck index: an encyclopedia of chemicals, drugs, and biologicals*. RSC Publishing
68. Pahari S, Sun L, Alexov E (2019) PKAD: a database of experimentally measured pKa values of ionizable groups in proteins. *Database*. <https://doi.org/10.1093/database/baz024>
69. Parr RG, Donnelly RA, Levy M et al (1977) Electronegativity: the density functional viewpoint. *J Chem Phys* 68(8):3801–3807. <https://doi.org/10.1063/1.436185>
70. Patel HM, Noolvi MN, Sharma P et al (2014) Quantitative structure–activity relationship (QSAR) studies as strategic approach in drug discovery. *Med Chem Res* 23(12):4991–5007. <https://doi.org/10.1007/s00044-014-1072-3>
71. Politzer P, Murray JS, Bulat FA (2010) Average local ionization energy: a review. *J Mol Model* 16(11):1731–1742. <https://doi.org/10.1007/s00894-010-0709-5>
72. Putz MV, Russo N, Sicilia E (2005) About the Mulliken electronegativity in DFT. *Theor Chem Acc* 114(1–3):38–45. <https://doi.org/10.1007/s00214-005-0641-4>
73. Rebollar-Zepeda AM, Galano A (2012) First principles calculations of pKa values of amines in aqueous solution: application to neurotransmitters. *Int J Quant Chem* 112(21):3449–3460. <https://doi.org/10.1002/qua.24048>
74. Reijenga J, van Hoof A, van Loon A et al (2013) Development of methods for the determination of pKa values. *Anal Chem Insights* 8(1):53–71. <https://doi.org/10.4137/ACI.S12304>
75. Saleh N, Koner AL, Nau WM (2008) Activation and stabilization of drugs by supramolecular pKa shifts: drug-delivery applications tailored for cucurbiturils. *Angew Chem Int Ed* 47(29):5398–5401. <https://doi.org/10.1002/anie.200801054>
76. Sandoval-Lira J, Mondragón-Solórzano G, Lugo-Fuentes LI et al (2020) Accurate estimation of pKb values for amino groups from surface electrostatic potential (VS, min) calculations: the isoelectric points of amino acids as a case study. *J Chem Inf Model* 60(3):1445–1452. <https://doi.org/10.1021/acs.jcim.9b01173>
77. Sashuk V, Butkiewicz H, Fiałkowski M et al (2016) Triggering autocatalytic reaction by host-guest interactions. *Chem Commun* 52(22):4191–4194. <https://doi.org/10.1039/c5cc10063a>
78. Settimo L, Bellman K, Knegtel RM (2014) Comparison of the accuracy of experimental and predicted pKa values of basic and acidic compounds. *Pharm Res* 31(4):1082–1095. <https://doi.org/10.1007/s11095-013-1232-z>
79. Seybold PG (2008) Analysis of the pKas of aliphatic amines using quantum chemical descriptors. *Int J Quant Chem* 108(15):2849–2855. <https://doi.org/10.1002/qua.21809>
80. Seybold PG, Kreye WC (2012) Theoretical estimation of the acidities of alcohols and azoles in gas phase, DMSO, and water. *Int J Quant Chem* 112(24):3769–3776. <https://doi.org/10.1002/qua.24216>
81. Seybold PG, Shields GC (2015) Computational estimation of pKa values. *Wiley Interdiscip Rev Comput Mol Sci* 5(3):290–297. <https://doi.org/10.1002/wcms.1218>
82. Shalaeva M, Kenseth J, Lombardo F et al (2008) Measurement of dissociation constants (pKa values) of organic compounds by multiplexed capillary electrophoresis using aqueous and cosolvent buffers. *J Pharm Sci* 97(7):2581–2606. <https://doi.org/10.1002/jps.21287>
83. Shields GC, Seybold PG (2013) *Computational approaches for the prediction of pKa values*, 1st edn. CRC Press, Boca Raton. <https://doi.org/10.1201/b16128>
84. Silakari O, Singh PK (2021) ADMET tools: prediction and assessment of chemical ADMET properties of NCEs. *Concepts and experimental protocols of modelling and informatics in drug design*. Elsevier, Amsterdam, pp 299–320. <https://doi.org/10.1016/b978-0-12-820546-4.00014-3>
85. Sjöberg P, Murray JS, Brinck T et al (1990) Average local ionization energies on the molecular surfaces of aromatic systems as guides to chemical reactivity. *Can J Chem* 68(8):1440–1443. <https://doi.org/10.1139/v90-220>
86. Soscún Machado HJ, Hinchliffe A (1995) Relationships between the HOMO energies and pKa values in monocyclic and bicyclic azines. *J Mol Struct THEOCHEM* 339(1–3):255–258. [https://doi.org/10.1016/0166-1280\(94\)04108-5](https://doi.org/10.1016/0166-1280(94)04108-5)
87. Sure R, Grimme S (2013) Corrected small basis set Hartree–Fock method for large systems. *J Comput Chem* 34(19):1672–1685. <https://doi.org/10.1002/jcc.23317>
88. Swebocki T, Niedziałkowski P, Cirocka A et al (2020) In pursuit of key features for constructing electrochemical biosensors-electrochemical and acid-base characteristic of self-assembled monolayers on gold. *Supramol Chem* 32(4):256–266. <https://doi.org/10.1080/10610278.2020.1739685>
89. Tam KY, Takács-Novák K (2001) Multi-wavelength spectrophotometric determination of acid dissociation constants: a validation study. *Anal Chim Acta* 434(1):157–167. [https://doi.org/10.1016/S0003-2670\(01\)00810-8](https://doi.org/10.1016/S0003-2670(01)00810-8)
90. Tehan BG, Lloyd EJ, Wong MG et al (2002) Estimation of pKa using semiempirical molecular orbital methods. Part 2: application to amines, anilines and various nitrogen containing heterocyclic compounds. *Quant Struct Activity Relationships* 21(5):473–485. [https://doi.org/10.1002/1521-3838\(200211\)21:5<473::AID-QSAR473>3.0.CO;2-D](https://doi.org/10.1002/1521-3838(200211)21:5<473::AID-QSAR473>3.0.CO;2-D)
91. Wan H, Holmén AG, Wang Y et al (2003) High-throughput screening of pKa values of pharmaceuticals by pressure-assisted capillary electrophoresis and mass spectrometry. *Rapid Commun Mass Spectrometry* 17(23):2639–2648. <https://doi.org/10.1002/rcm.1229>

92. Wang Z, Sun C, Yang K et al (2022) Cucurbituril-based supramolecular polymers for biomedical applications. *Angew Chem* 134(38):e202206763. <https://doi.org/10.1002/ange.202206763>
93. Watwe V, Kulkarni S, Kulkarni P (2023) Development of dried uncharred leaves of *Ficus benjamina* as a novel adsorbent for cationic dyes: kinetics, isotherm, and batch optimization. *Ind Crops Prod* 195:116449. <https://doi.org/10.1016/j.indcrop.2023.116449>
94. Yang Q, Li Y, Yang J et al (2020) Holistic prediction of the pKa in diverse solvents based on a machine-learning approach. *Angew Chem* 132(43):19444–19453. <https://doi.org/10.1002/ange.202008528>
95. Yin T, Zhang S, Li M et al (2019) Macrocyclic encapsulation triggered supramolecular pKa shift: a fluorescence indicator for detecting octreotide in aqueous solution. *Sens Actuat B Chem* 281:568–573. <https://doi.org/10.1016/j.snb.2018.10.136>
96. Yu H, Kühne R, Ebert RU et al (2010) Comparative analysis of QSAR models for predicting pKa of organic oxygen acids and nitrogen bases from molecular structure. *J Chem Inf Model* 50(11):1949–1960. <https://doi.org/10.1021/ci100306k>
97. Zhang S (2012) A reliable and efficient first principles-based method for predicting pKa values. 4. Organic bases. *J Comput Chem* 33(31):2469–2482. <https://doi.org/10.1002/jcc.23068>
98. Zhang YM, Yang Y, Zhang YH et al (2016) Polysaccharide nanoparticles for efficient siRNA targeting in cancer cells by supramolecular pKa shift. *Sci Rep* 6(1):1–11. <https://doi.org/10.1038/srep28848>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

