

Evaluation of the pKa's of Quinazoline Derivatives : Usage of Quantum Mechanical Based Descriptors

Melisa Kiran

Bogazici University

Zeynep Pinar Haslak

Bogazici University

Halit Ates

Bogazici University

Viktorya Aviyente (✉ aviye@boun.edu.tr)

Bogazici University

Fatma Ahu Akin

Bogazici University

Research Article

Keywords: pKa, DFT descriptors, atomic charge, quinazoline derivatives

Posted Date: September 2nd, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3304183/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Evaluation of the pK_a 's of Quinazoline Derivatives : Usage of Quantum Mechanical Based Descriptors

Melisa Kiran^{1,‡}, Zeynep Pinar Haslak^{1,2,‡}, Halit Ates¹, Viktorya Aviyente^{1,*}, Fatma Ahu
Akin^{1,*}

¹ Bogazici University, Department of Chemistry, 34342, Bebek, Istanbul, Turkey

² Université de Reims Champagne-Ardenne, 51687, Reims, France

Correspondance: akin@boun.edu.tr, aviye@boun.edu.tr

[‡]These authors contributed equally.

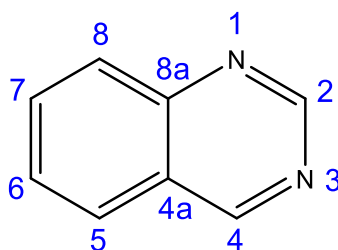
Abstract

In this study, several quantum mechanical-based computational approaches have been used in order to propose accurate protocols for predicting the pK_a 's of quinazoline derivatives, which constitute a very important class of natural and synthetic compounds in organic, pharmaceutical, agricultural and medicinal chemistry areas. Linear relationships between the experimental pK_a 's and nine different DFT descriptors (atomic charge on nitrogen atoms ($Q(N)$), ionization energy (I), electron affinity (A), chemical potential (μ), hardness (η), electrophilicity index (ω), fukui functions (f^+, f^-), condensed dual descriptor (Δf) and local hypersoftness ($s_k^{(2)}$) were considered. Several DFT methods (a combination of five DFT functionals and two basis sets) in conjunction with two different implicit solvent models were tested, and among them, M06L/6-311++G(d,p) level of theory employing the CPCM solvation model was found to give the strongest correlations between the DFT descriptors and the experimental pK_a 's of the quinazoline derivatives. The calculated atomic charge on N₁ atom ($Q(N_1)$) was shown to be the best descriptor to reproduce the experimental pK_a 's ($R^2=0.927$), whereas strong correlations were also derived for A , ω , μ , $s_k^{(2)}$ and Δf . The QM-based protocols presented in this study will enable fast and accurate high-throughput pK_a predictions of quinazoline derivatives and the relationships derived can be effectively used in data generation for successful machine learning models for pK_a predictions.

Keywords: pK_a , DFT descriptors, atomic charge, quinazoline derivatives

1. Introduction

Quinazoline (1,3-diazanaphthalene; 1,3-benzodiazine) derivatives are the privileged scaffolds in drug discovery due to their distinct and wide range of bioactivities. Anti-cancer [1–4], anti-bacterial [5–7], anti-virus [8], anti-microbial [9], anti-inflammatory [10,11], analgesic [5,9,11], anti-cytotoxic [12], anti-convulsant [13], anti-tuberculosis [14,15], anti-psychotic [16], anti-malarial [17,18], anti-obesity [19] and anti-diabetes [19] activities of quinazoline derivatives make these compounds valuable potent therapeutic agents. Recently, the Food and Drug Administration has approved several quinazoline derivatives as anticancer drugs, such as gefitinib, erlotinib and lapatinib. The functions of the quinazoline based compounds can be easily modified due to the fact that highly delocalized π electrons in N=C double bonds enhance the reactivity of quinazolines towards many types of nucleophiles (**Scheme 1**) [20, 21].



Scheme 1 2D representation of quinazoline molecule.

Acidic and basic sites within a drug molecule govern the solubility, lipophilicity and permeability in a cell membrane as well as its ADMET (absorption, distribution, metabolism, excretion and toxicity) properties. Therefore, the biological activity of a ligand is dependent on the pK_a 's of the ionizable functional groups which form acid-base interactions with the target receptor environment [22]. The pK_a of the ionizable groups can be tuned by introduction of suitable functional groups to the parent lead compound, and the ligand's ADMET properties can be easily modulated.

The two nitrogen atoms at positions 1 and 3 are the hydrogen-bond acceptors in quinazoline molecule and the protonation can take place on one of the nitrogen atoms. For the

case of substituted quinazoline derivatives, the protonation of N₁ or N₃ depends on many factors such as steric hindrance, inductive or electron donating/withdrawing effects induced by the substituents. The identification of the most favorable protonation site and the determination of the p*K_a*'s of these sites is still a challenge today. There are many experimental techniques for p*K_a* determination [23]; nevertheless the accuracy is affected when multiple tautomeric species are present in the medium, the concentration of the solution gets closer to the limits of quantification, the intermediates have short life-times, other solutes interfere and the isolation cannot be performed properly [24-26]. Besides, experimental methods are less practical for the p*K_a* determination of a large number of compounds. Therefore, the accurate p*K_a* calculations by theoretical protocols are of great interest to complement the experimental approaches.

The acid dissociation constant can be calculated by the use of the thermodynamic cycle which takes into account the free energy of solvation, desolvation and deprotonation in the gas phase of the acidic species. However, large uncertainties emerging from the solvation free energy of the proton (H⁺), instability of the species in the gas phase and large conformational differences between the gas and the solvent calculations make this method less applicable [27]. In order to improve the accuracy of estimated p*K_a* values, there is a tremendous effort based on computational approaches [27-50]. The isodesmic reaction scheme appeared as an alternative protocol to the thermodynamic cycle in which the errors due to gas phase calculations are minimized [25]. Applications of isodesmic reactions for p*K_a* estimations of small organic molecules (pyridines, alcohols, carboxylic acids, amines, phenols, benzoic acids) [51], nitrogen-containing organic superbases [52], amino acids and peptides [27], N-acyl sulfonamide based drug-like compounds [53] and aromatic arsonic acids [52] have been reported to present successful correlations between experimental and calculated p*K_a*'s. On the other hand, quantitative structure property relationship (QSPR) techniques offer faster and accurate p*K_a* calculations by linking the molecular descriptors to the p*K_a*'s of organic

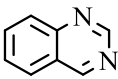
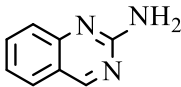
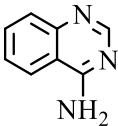
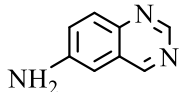
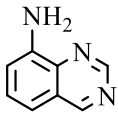
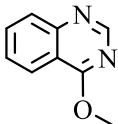
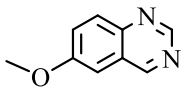
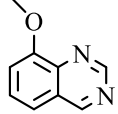
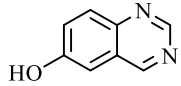
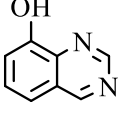
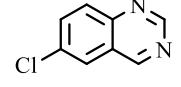
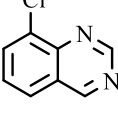
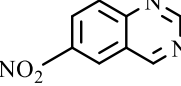
molecules. An accurate protocol was developed by Ugur *et al.* by computing the atomic charges for the prediction of pK_a 's of thiols and alcohols by testing several DFT functionals, basis sets, semiempirical methods, solvation and charge models [54]. The same approach was used by Haslak *et al.* in order to calculate the pK_a 's of carboxylic acids [55]. Considering its strong accuracy and high computational efficiency, these pK_a prediction calculations using atomic charges have also indicated a promising method for predicting amino acids' pK_a in a protein environment. Conceptual DFT descriptors, such as ionization energy (I), electron affinity (A), chemical potential (μ), global hardness (η), electrophilicity index (ω) and Fukui functions (f), can be successfully applied in order to identify the protonation sites and to predict the pK_a 's of molecules. Recently, by using the reactivity descriptors that arise from Conceptual DFT, different equations were derived for predicting the pK_a 's of several peptides [27,36,56].

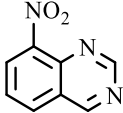
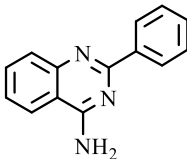
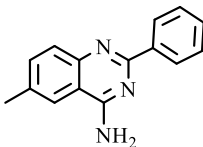
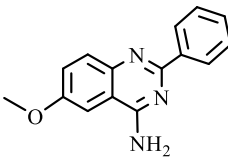
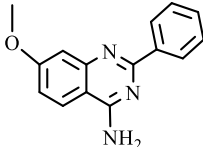
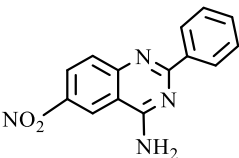
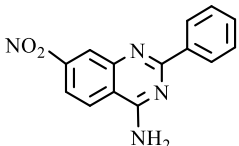
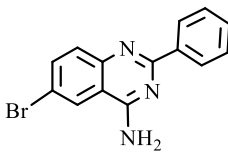
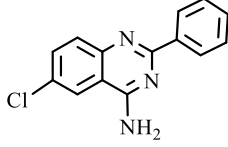
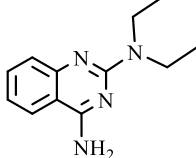
To the best of our knowledge, there is no QM-based protocol that yields precise pK_a values for the quinazoline derivatives. Inspired by the successful applications of DFT descriptors, in this study we aim to suggest fast and accurate protocols for the prediction of pK_a 's of quinazoline derivatives, which constitute a very important class of natural and synthetic compounds in organic, pharmaceutical, agricultural and medicinal chemistry areas.

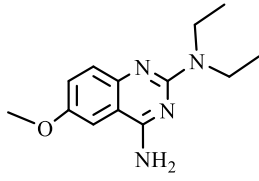
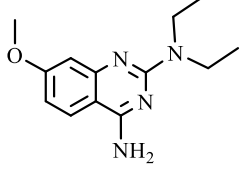
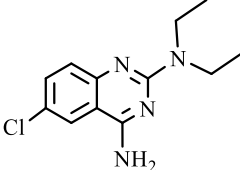
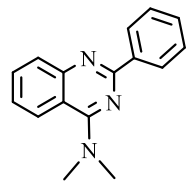
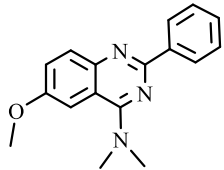
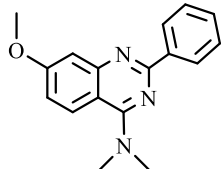
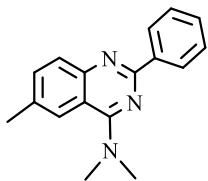
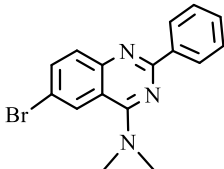
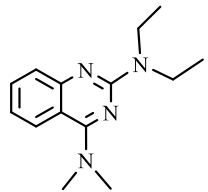
2. Methodology

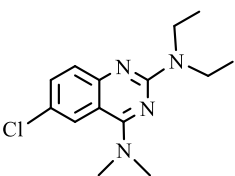
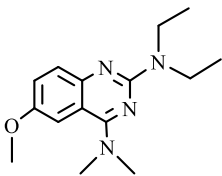
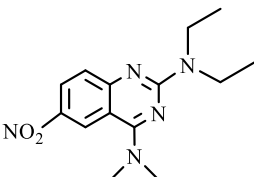
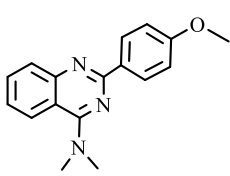
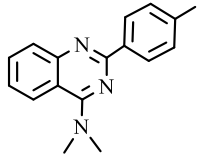
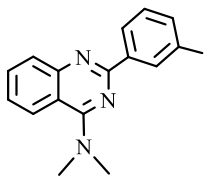
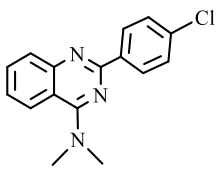
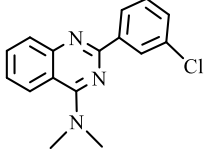
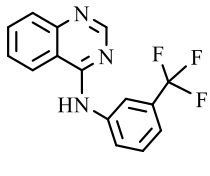
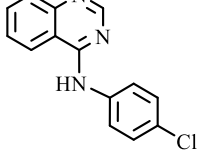
Experimental Database: A total of 46 quinazoline derivatives were selected from the literature with experimentally determined pK_a 's which range from 2.40 to 8.98 (**Table 1**). The training set composed of 30 molecules (**Table 2**) was used for obtaining linear relationships between the experimental pK_a 's and several DFT descriptors. Then, the applicability of the suggested protocol was tested on the test set composed of 16 molecules (**Table 3**).

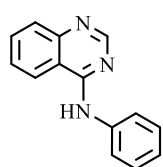
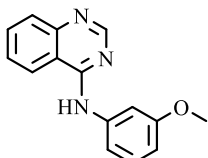
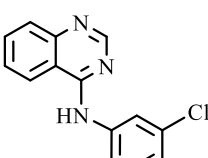
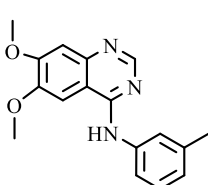
Table 1 Selected quinazoline and its derivatives in this study: Molecule number, IUPAC Name, 2D Representation and Experimental pK_a 's.

Molecule Number	IUPAC Name	2D Representation	Experimental pK_a
1	Quinazoline		3.31 [57]
2	2-Quinazolinamine		4.43 [58]
3	4-Quinazolinamine		5.73 [58]
4	6-Quinazolinamine		3.20 [58]
5	8-Quinazolinamine		2.40 [59]
6	4-Methoxyquinazoline		3.13 [59]
7	6-Methoxyquinazoline		2.85 [59]
8	8-Methoxyquinazoline		3.51 [59]
9	6-Hydroquinazoline		3.12 [60]
10	8-Hydroquinazoline		3.41 [59]
11	6-Chloroquinazoline		3.55 [59]
12	8-Chloroquinazoline		3.30 [59]
13	6-Nitroquinazoline		4.18 [59]

14	8-Nitroquinazoline		4.00 [59]
15	2-Phenyl-4-quinazolinamine		5.44 [61]
16	6-Methyl-2-phenyl-4-quinazolinamine		5.16 [61]
17	6-Methoxy-2-phenyl-4-quinazolinamine		5.33 [61]
18	7-Methoxy-2-phenyl-4-quinazolinamine		5.62 [61]
19	6-Nitro-2-phenyl-4-quinazolinamine		4.54 [61]
20	7-Nitro-2-phenyl-4-quinazolinamine		4.27 [61]
21	6-Bromo-2-phenyl-4-quinazolinamine		4.78 [61]
22	6-Chloro-2-phenyl-4-quinazolinamine		4.98 [61]
23	2-N,2-N-diethylquinazoline-2,4-diamine		7.79 [61]

24	2-N,2-N-diethyl quinazoline-6-methoxy- 2,4-diamine		7.82 [61]
25	2-N,2-N-diethyl quinazoline-7-methoxy- 2,4-diamine		8.31 [61]
26	2-N,2-N-diethyl quinazoline-6-chloro- 2,4-diamine		6.98 [61]
27	4-(N,N-dimethylamino)- 2-phenyl quinazoline		6.31 [61]
28	4-(N,N-dimethylamino)- 6-methoxy-2-phenyl quinazoline		6.61 [61]
29	4-(N,N-dimethylamino)- 7-methoxy-2-phenyl quinazoline		6.20 [61]
30	4-(N,N-dimethylamino)- 6-methyl-2-phenyl quinazoline		6.52 [61]
31	4-(N,N-dimethylamino)- 6-bromo-2-phenyl quinazoline		5.88 [61]
32	2-(diethylamino)-4- (N,N-dimethylamino) quinazoline		8.88 [61]

33	6-Chloro-2-(diethylamino)-4-(N,N-dimethylamino)quinazoline		7.91 [61]
34	6-Methoxy-2-(diethylamino)-4-(dimethylamino)quinazoline		8.98 [61]
35	6-Nitro-2-(diethylamino)-4-(dimethylamino)quinazoline		6.63 [61]
36	4-(N,N-dimethylamino)-6-methoxy-2-phenylquinazoline		6.62 [61]
37	4-(N,N-dimethylamino)-2-(p-methylphenyl)quinazoline		6.40 [61]
38	4-(N,N-dimethylamino)-2-(o-methylphenyl)quinazoline		6.28 [61]
39	4-(N,N-dimethylamino)-2-(p-chlorophenyl)quinazoline		5.88 [61]
40	4-(N,N-dimethylamino)-2-(o-chlorophenyl)quinazoline		5.54 [61]
41	N-[3-(trifluoromethyl)phenyl]-4-quinazolinamine		5.03 [61]
42	N-(4-chlorophenyl)-4-quinazolinamine		6.02 [62]

43	N-phenyl-4-quinazolinamine		6.08 [62]
44	N-(3-methoxyphenyl)-4-quinazolinamine		5.37 [62]
45	N-(3-chlorophenyl)-4-quinazolinamine		5.28 [62]
46	6,7-dimethoxy-N-(3-methylphenyl)-4-quinazolinamine		5.77 [62]

Quantum Mechanical Calculations: All of the quantum mechanical calculations were done by using the Gaussian16 program package [63]. A systematic conformational analysis was conducted for 46 molecules in their neutral forms using the semi-empirical PM6 method by using the conformer search module implemented in the Spartan program [64]. The geometries corresponding to local minima were further optimized by using five different density functionals (B3LYP [65], M06L [66], M06-2X [67], ω B97XD [68], MN12SX [69]). We considered these functionals based on their different families: B3LYP as a Generalized Gradient Approximation Exchange-Correlation (GGA XC) functional, M06L as a meta-GGA XC functional, M06-2X as hybrid meta-GGA XC functional, ω B97XD as hybrid-GGA XC functional and MN12SX as a meta-GGA hybrid screened XC functional. We did not use the Local Density Approximation (LDA) functionals, because previous studies showed that models based on meta-GGA and GGA functionals are substantially better than those including LDA functionals [70]. Two different basis sets (6-31+G(d) and 6-311++G(d,p)) and solvation models (SMD and CPCM) were tested. The partial atomic charges were derived from Natural Population Analysis

(NPA) [71]. Vibrational frequency analysis was performed to ensure that there is no imaginary frequency.

The chemical potential (μ) is the derivative of the energy as a function of the number of electrons at constant external potential and measures the tendency of an electron to escape from the system:

$$\mu = \left(\frac{\partial E}{\partial N} \right)_{v(r)} \cong -\frac{1}{2} (I + A) \cong \frac{1}{2} (\varepsilon_{LUMO} + \varepsilon_{HOMO}) \quad (1)$$

where I is the ionization potential, A is the electron affinity, ε_{LUMO} is the lowest unoccupied molecular orbital energy and ε_{HOMO} is the highest unoccupied molecular orbital energy.

The hardness (η) is the second derivative of the energy as a function of the number of electrons at constant external potential [72] and expresses the system's reaction to change in the number of electrons:

$$\eta = \left(\frac{\partial^2 E}{\partial N^2} \right)_{v(r)} \cong \frac{1}{2} (I - A) \cong \frac{1}{2} (\varepsilon_{LUMO} - \varepsilon_{HOMO}) \quad (2)$$

The electrophilicity index (ω) [73] measures the stabilization energy of the system when the electron density increases, and it can be expressed as:

$$\omega = \frac{\mu^2}{2\eta} \quad (3)$$

The Fukui function (f) [74] is the derivative of the electron density as a function of the number of electrons at constant external potential and describes how easily a molecular site can accept or donate electrons:

$$f = \left(\frac{\partial \rho}{\partial N} \right)_{v(r)} \quad (4)$$

The reactivity of an atom k with N electrons for a nucleophilic attack (equation 5) and electrophilic attack (equation 6) are given as:

$$f_k^+ = q_k(N + 1) - q_k(N) \quad (5)$$

$$f_k^- = q_k(N) - q_k(N - 1) \quad (6)$$

The condensed dual descriptor (Δf_k) characterizes if the process on atom k is driven by a nucleophilic attack ($\Delta f_k > 0$) or an electrophilic attack ($\Delta f_k < 0$) [75-80]:

$$\Delta f_k = f_k^+ - f_k^- \quad (7)$$

The Local Hyper Softness ($s_k^{(2)}$) has been defined to measure the local reactivity on atom k with respect to the molecular size [78,80]:

$$s_k^{(2)} = \frac{\Delta f_k}{\eta^2} \quad (8)$$

3. Results and Discussion

In this study, we present several QM-based protocols for fast and accurate pK_a prediction of quinazoline derivatives. In the first part, linear relationships between the experimental pK_a 's and calculated partial atomic charges on nitrogen atoms of quinazoline derivatives were constructed, the effect of DFT functionals, basis sets, the implicit solvent models and the atomic charge to be considered were evaluated. In the second part, the linear relationships between the experimental pK_a 's and eight different conceptual DFT descriptors were derived.

3.1. Linking Partial Atomic Charges to pK_a 's of Quinazoline Derivatives

One of the core components of determining a molecule's physical and chemical characteristics is its electronic distribution. Intuitively, partial atomic charges fit into the narrative of comprehending physical and chemical activity since they are a direct consequence of the electron distribution. The acidic or basic character of a ligand is closely related to the partial atomic charges as suggested before [54,55,81,82], therefore, in this part of the study the relationship between partial atomic charges and experimental pK_a 's of quinazoline derivatives is investigated. For the different classes of molecules, the natural tendency of the experimental pK_a values is to cluster in distinct ranges. Previous work of Haslak and Ugur showed that carboxylic acids obeyed $pK_a = -43.968Q - 32.411$ (where $Q = \max\{q(O_1), q(O_2)\}$, $R^2=0.955$) [55] whereas alcohols obeyed $pK_a = -38.847Q(O^-) - 17.647$ ($R^2=0.995$) and thiols obeyed $pK_a = -24.744Q(S^-) - 10.880$ ($R^2=0.986$) [54]. Different "best methodologies" are expected to emerge

since the reactive functional groups are attached to different molecule classes in each molecule type. For instance, the best linearities between pK_a 's of carboxylic acids and atomic charges were obtained when the NPA charges were derived at M06L/6-311G(d,p) level using the SMD implicit solvent model, whereas M06-2X/6-311G and B3LYP/3-21G in conjunction with CPCM were found to be the most reliable methods for thiols and alcohols, respectively. Therefore, to identify the best methodology for quinazoline class of molecules, many factors such as the choice of the DFT functional, the basis set, the implicit solvent model and the identity of the atomic charges to be considered are tested. The full detailed results are presented in the Supplementary Information (**Table S1-S5**). A deep analysis of the tested DFT functionals (B3LYP, M06L, M06-2X, ω B97XD, MN12SX), basis sets (6-31+G(d) and 6-311++G(d,p)), implicit solvent models (SMD and CPCM) and atomic charges ($q(N_1)$, $q(N_3)$, $\max\{q(N_1), q(N_3)\}$, $\min\{q(N_1), q(N_3)\}$, $\text{avg}\{q(N_1), q(N_3)\}$) was performed by constructing the linear equations by a least-square fit:

$$pK_a = a \cdot Q + b \quad (9)$$

where a and b are the fitted parameters and Q is the calculated NPA charge of the choice. Quinazoline and its derivatives have two N atoms in their cyclic structure at 1 and 3 positions and both can act as H-bond acceptors. The substituents on the ring may act as electron withdrawing or donating groups, increasing or decreasing the electron population on the N atoms. Thus, the protonation may take place on N atoms and the atomic charges on N atoms, Q , can be considered in different ways:

$$Q = q(N_1) \quad (10)$$

$$Q = q(N_3) \quad (11)$$

$$Q = \max\{q(N_1), q(N_3)\} \quad (12)$$

$$Q = \min\{q(N_1), q(N_3)\} \quad (13)$$

$$Q = \text{avg}\{q(N_1), q(N_3)\} \quad (14)$$

The charge on N_1 (equation 10), the charge on N_3 (equation 11), the highest value of charges on N atoms (equation 12), the lowest value of charges on N atoms (equation 13) or the average value of charges on N atoms (equation 14) have been extracted and correlated with the experimental pK_a 's. **Fig. 1** displays the distribution of Pearson correlation coefficients (R^2) obtained from each linear fit with respect to the DFT functionals and the atomic extraction scheme. Each plot was generated by holding a basis set and a solvent model fixed. According to the results, all the DFT methods and charge descriptor combinations give strong correlations between the experimental pK_a 's and calculated NPA atomic charges with $0.780 \leq R^2 \leq 0.930$. The relationship is observed to deviate from linearity when the charge on N_3 atom is considered ($0.773 \leq R^2 \leq 0.864$), becoming more reliable when N_1 is considered ($0.864 \leq R^2 \leq 0.930$). Among the DFT functionals, while the most accurate methods are found to be M06L, B3LYP and ω B97XD; MN12SX and M06-2X gave slightly weaker linear relationships with the experimental pK_a 's.

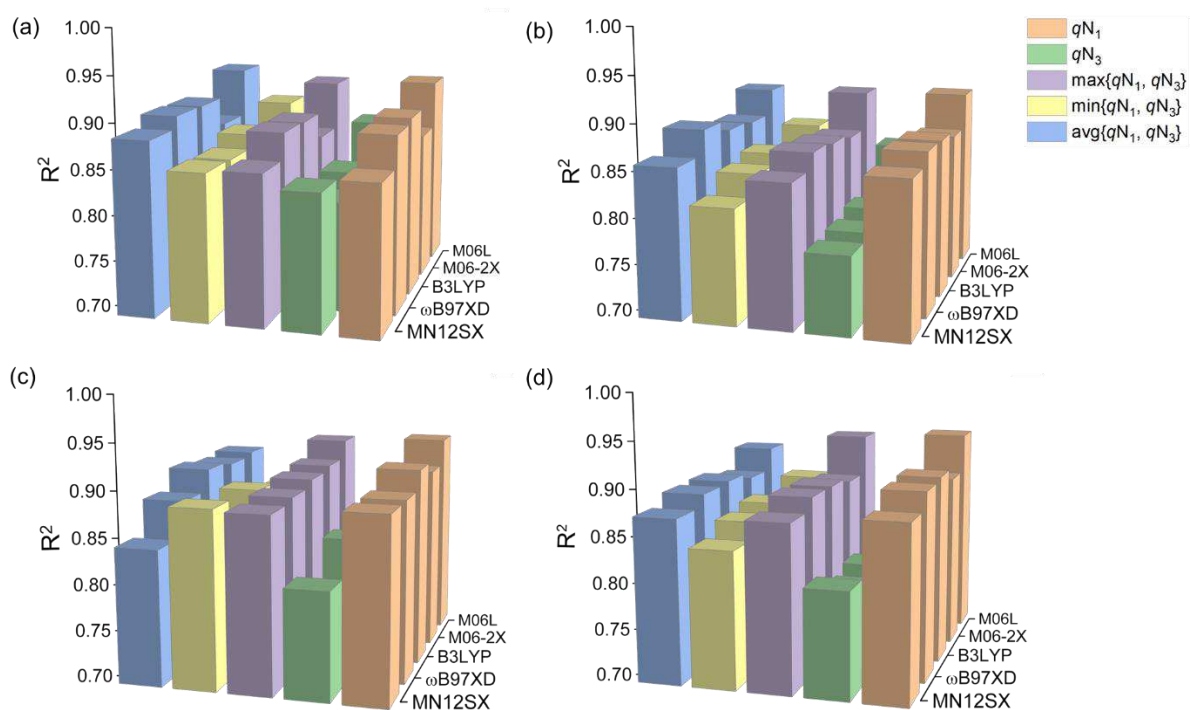


Fig. 1 R^2 distributions of the linear equations for five different DFT functionals and five different atomic charge descriptors considered in this work. Geometry optimizations and NPA

charge calculations were performed using (a) 6-31+G(d) basis set and CPCM solvent model, (b) 6-31+G(d) basis set and SMD solvent model, (c) 6-311+G(d,p) basis set and CPCM solvent model, (d) 6-311+G(d,p) basis set and SMD solvent model.

Out of the 125 combinations of DFT functionals, basis sets, solvation models and charge descriptors tested, the N_1 charge computed with NPA at the M06L/6-311++G(d,p) level of theory employing the SMD solvent model presented the best regression ($R^2=0.930$). Except for molecule **26**, all of the predicted pK_a 's are found to be within the range of ± 1 unit compared to experimental values, with mean absolute deviation (MAD) value of 0.40 and $MAX-\Delta pK_a$ value of 1.05. On the other hand, when the CPCM solvation model is used, R^2 is calculated as 0.927 with $MAD=0.40$ and $MAX-\Delta pK_a=0.96$. Due to lower $MAX-\Delta pK_a$ values obtained from CPCM model, the discussions will be based on the calculations performed at M06L/6-311++G(d,p) level of theory employing the CPCM solvent model, in the rest of the manuscript. **Fig. 2** displays the relationship between the calculated NPA atomic charges of atom N_1 and experimental pK_a 's for the training set. The predicted pK_a 's are calculated by using the parametrized equations in each fitting and the results are presented in **Table 2** for the training set and **Table 3** for the test set. These results indicate the presence of a strong correlation between experimental pK_a 's and the charges on N_1 atom.

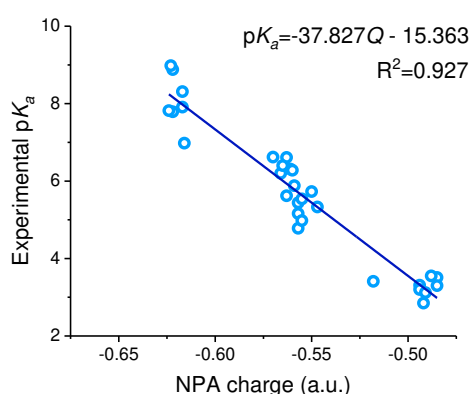


Fig. 2 Linear regression between experimental pK_a and the calculated NPA atomic charges on atom N_1 at M06L/6-311++G(d,p) level of theory with CPCM solvation model.

Table 2 Molecule number, experimental pK_a , predicted pK_a , difference between experimental and predicted pK_a values for the training set (M06L/6-311++G(d,p) // CPCM).

Molecule number	pK_a (exp.)	pK_a (pred.)	ΔpK_a
1	3.31	3.32	0.01
3	5.73	5.44	-0.29
4	3.20	3.32	0.12
7	2.85	3.25	0.40
8	3.51	2.98	-0.53
9	3.12	3.21	0.09
10	3.41	4.23	0.82
11	3.55	3.10	-0.45
12	3.30	2.98	-0.32
15	5.44	5.71	0.27
16	5.16	5.71	0.55
17	5.33	5.33	0.00
18	5.62	5.93	0.31
21	4.78	5.71	0.93
22	4.98	5.63	0.65
23	7.79	8.16	0.38
24	7.82	8.24	0.42
25	8.31	7.98	-0.33
26	6.98	7.94	0.96
27	6.31	5.86	-0.45
28	6.61	5.93	-0.68
29	6.20	6.05	-0.15
32	8.88	8.16	-0.71
33	7.91	7.98	0.07
34	8.98	8.20	-0.78
36	6.62	6.20	-0.42
37	6.40	6.01	-0.39
38	6.28	5.82	-0.46
39	5.88	5.78	-0.10
40	5.54	5.63	0.09

Table 3 Molecule number, experimental pK_a , predicted pK_a , difference between experimental and predicted pK_a values for the test set (M06L/6-311++G(d,p) // CPCM).

Molecule number	pK_a (exp.)	pK_a (pred.)	ΔpK_a
2	4.43	6.16	1.73
5	2.40	3.78	1.38
6	3.13	4.42	1.29
13	4.18	3.02	-1.16
14	4.00	2.76	-1.24
19	4.54	5.59	1.05

20	4.27	5.14	0.87
30	6.52	5.71	-0.81
31	5.88	5.67	-0.21
35	6.63	7.33	0.70
41	5.03	4.84	-0.19
42	6.02	4.99	-1.03
43	6.08	5.06	-1.02
44	5.37	5.06	-0.31
45	5.28	4.87	-0.41
46	5.77	5.21	-0.56

The strongest outliers in the test set are two amino substituted compounds **2** and **5**, a methoxy substituted compound **6** and two nitro substituted compounds **13** and **14**. The two nitrogen atoms of quinazoline ring are potential protonation sites and the protonation of N atoms is a dynamic process; with the proton exchange between two N atoms, equilibrium is achieved. While the proton transfer from N₁ to N₃ atom takes place, -NH₂ substituent on position 2 (molecule **2**) might be involved in equilibrium, this fact results in deviations in pK_a measurements. On the other hand, when the amino moieties are away from N₁, in the case of molecules **3** and **4** which have amino substituents at 4 and 6 positions of quinazoline ring, our model is found to be successful in predicting the pK_a 's of these molecules. Compound **6** has a bulky methoxy substituent attached to position 4, in which -CH₃ interacts with N₃ atom in its ground state to make favorable interactions. However, as the proton exchange occurs between N₁ and N₃, -CH₃ of methoxy moiety is assumed to induce extra electrostatic repulsion which results in a more destabilized conjugate base compared to other methoxy substituted quinazolines. In this respect, we would expect more basic character for molecule **4**, as our model predicted ($pK_a^{\text{calc}} = 4.42$), as opposed to molecules **7** ($pK_a^{\text{exp}} = 2.85$) and **8** ($pK_a^{\text{exp}} = 3.51$). For nitro substituted compounds **13** and **14**, our model suggests more acidic characters, i.e. lower pK_a 's, compared to experimental values. The strong electronegativity of -NO₂ group induces strong electron-withdrawing effect and thus alters the π -electron delocalization of the ring it is attached to, this could lead to shifts in the equilibrium between N atoms during the proton

exchange. These compounds have only nitro substituents on the quinazoline ring, whereas compounds **19**, **20** and **35** have a phenyl or diethylamino at position 2 and amino substituents on position 4 in addition to -NO₂, for which the model predicts less acidic characters ($\Delta pK_a = 1.05, 0.87, 0.70$, respectively). Therefore, the electron donating substituents on quinazoline ring in compounds **19**, **20** and **35** stabilize the nature of the electron delocalization in these molecules and thus enables us to predict more accurate pK_a 's.

Fig. 3 displays the relationships between the experimental pK_a 's of quinazoline derivatives, and the calculated atomic charges as expressed in equations 10-14 employing M06L/6-311++G(d,p) level of theory with CPCM solvation model. When the charge on N₃ atom is considered, the relationship is the least trustworthy (**Fig. 3a**) among all the charge descriptors considered, while the highest N atomic charge (least negative) scheme (**Fig. 3b**) gives closer relationship to the N₁ atomic charge (**Fig. 3**, $R^2=0.927$). When the lowest N charge (most negative) is used to relate with experimental pK_a 's (**Fig. 3c**), the relationship is slightly more accurate than N₃ scheme and less accurate than the average sum of N₁ and N₃ charges (**Fig. 3d**). As stated before, both N atoms are prone to be protonated: the most negative N (N₃) is supposed to attack the proton more easily and the equilibrium will quickly shift to the left towards the protonated species. However, the equilibrium constant between the proton and the base -and thus the pK_a - may be better correlated with the rate determining step which is expressed by the charge of the least negative N (N₁) which takes part in the rate determining step. As indicated in our previous study the least negative center (N₁) indicates its relative stability as compared to N₃, thus its dominance as an acid [55].

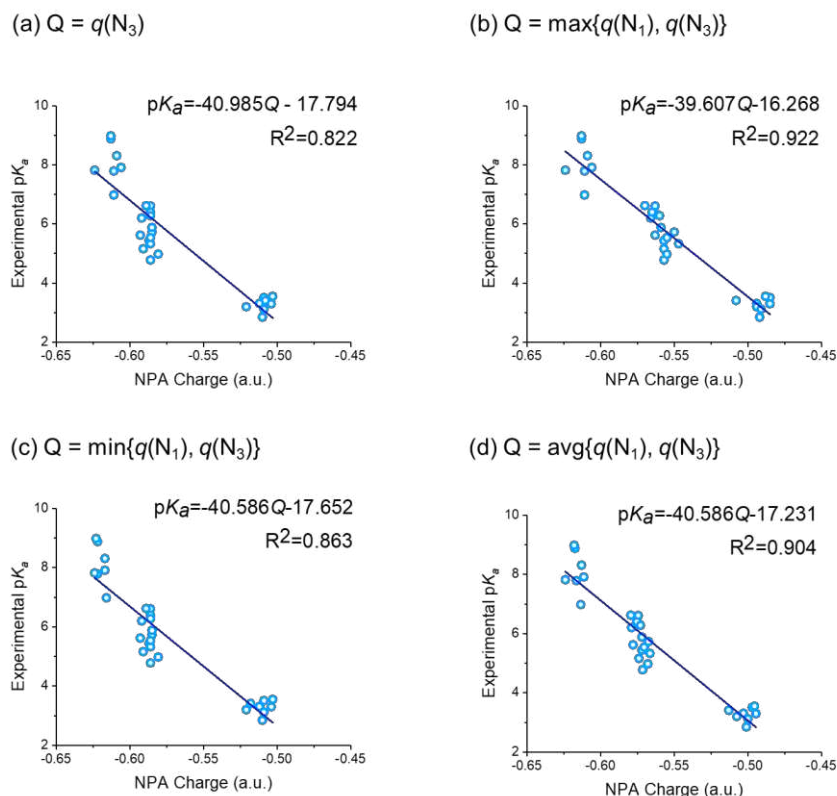


Fig. 3 Effect of choice of the charge descriptor on the linear regression between experimental pK_a 's and (a) charge on N_3 atom, (b) highest value of charges on N_1 and N_3 atoms, (c) lowest value of charges on N_1 and N_3 atoms, (d) average value of charges on N_1 and N_3 atoms (M06L/6-311++G(d,p) // CPCM).

Since the correct description of the solute and solvent interactions are vital in charge derivation, the aqueous environment was mimicked by two different implicit solvent models: SMD and CPCM. Therefore, the relationships between the atomic charges calculated in the presence of implicit water and the experimental pK_a 's of the training set molecules were examined. **Fig. 4** displays the relationship between the experimental pK_a 's and the calculated N_1 atomic charges employing M06L/6-311++G(d,p) level of theory with the SMD solvent model. According to the data obtained, SMD can also be used with confidence ($R^2=0.930$). For all the combinations of DFT methods, basis sets and charge descriptors, the linear regression fits of SMD and CPCM are very similar. Thus, both models are good in describing the interactions between the quinazoline-like molecules and water.

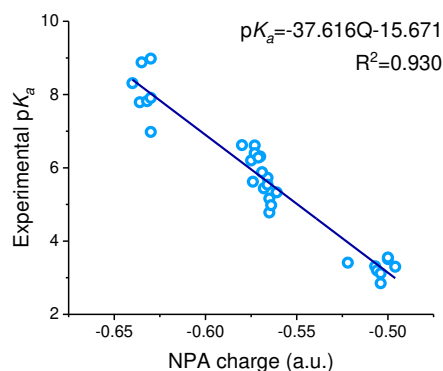


Fig. 4 Effect of choice of the solvent model on the linear regression between experimental pK_a 's and the charge on N_1 atom (M06L/6-311++G(d,p) // SMD).

Lastly, the effect of DFT functionals and basis sets on pK_a predictions of quinazoline derivatives was analyzed by following the same protocol outlined above. The NPA charges on N_1 atom computed at various levels of theories with the CPCM model were related with the experimental pK_a 's of the training set molecules. According to **Fig. 5**, all the DFT functional and basis set combinations give very accurate correlations between the NPA atomic charges on N_1 atom calculated with CPCM solvation model and experimental pK_a 's with R^2 range of $0.864 \leq R^2 \leq 0.927$. Additional diffuse and polarization functions are observed to increase the power of predictivity, i.e., 6-311++G(d,p) basis set gives more accurate correlations compared to 6-31+G(d), independent of the DFT functional. Among five DFT functionals, M06-2X and MN12SX give the poorest relationships ($0.874 \leq R^2 \leq 0.900$, $0.45 \leq \text{MAD} \leq 0.50$ and $0.864 \leq R^2 \leq 0.896$, $0.47 \leq \text{MAD} \leq 0.53$), M06L has the highest R^2 's ($0.915 \leq R^2 \leq 0.927$), smallest MADs ($0.40 \leq \text{MAD} \leq 0.41$) and ΔpK_a 's ($0.00 \leq |\Delta pK_a| \leq 1.24$), and overall, M06L/6-311+G(d,p) level of theory gives the most accurate pK_a 's.

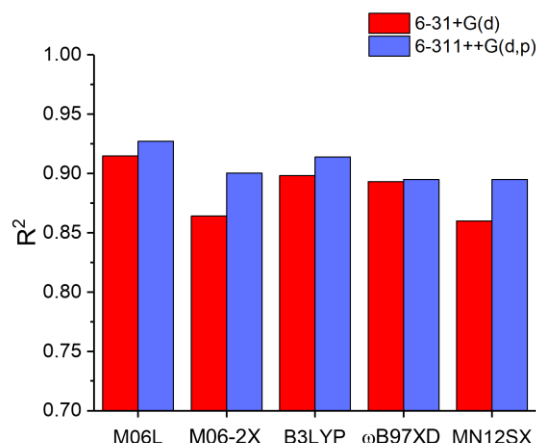


Fig. 5 R^2 distributions of the constructed linear equations for five different DFT functionals and two different basis sets considered in this work. Geometry optimizations and NPA charge calculations on N_1 atom were performed using CPCM solvent model.

We have also considered a Boltzmann distribution (BD) taking into account the Gibbs free energies for all the conformations with the specified methodology, in addition to the ground state (GS) conformation. The results in **Table 4** indicate that the regression coefficient is almost similar in both cases.

Table 4 R^2 distributions of the linear equations for five different atomic charge descriptors considered in this work (M06L/6-311++G(d,p) // CPCM) (GS = ground state, BD = Boltzmann distribution).

Atomic Extraction Scheme	R^2 (GS)	R^2 (BD)
$q(N_1)$	0.927	0.927
$q(N_3)$	0.822	0.835
$\max\{q(N_1), q(N_3)\}$	0.922	0.933
$\min\{q(N_1), q(N_3)\}$	0.863	0.874
$\text{avg}\{q(N_1), q(N_3)\}$	0.904	0.915

Additionally, we attempted to use the mixed solvation model, however in our calculations with explicit water molecules (one and three) we have found that the results were not consistent with each other, the position of the explicit water molecules have changed the results tremendously. Indeed, the results with mixed solvation model should be treated with care since

they can be very much dependent on the position of the explicit solvent molecules included in the calculations [83].

3.2. Linking Conceptual DFT Descriptors to pK_a 's of Quinazoline Derivatives

The chemical information about the intrinsic properties of the molecules is encoded by the quantum chemical descriptors and they can be associated with reactivity and activity behaviors of the molecules. Due to the inexpensive and fast calculations of the QM descriptors, their applications for various reactivity measurements of chemical structures are very attractive. In this part of the study, the pK_a 's of the quinazoline derivatives are intended to be predicted by using the global and local conceptual DFT descriptors. Ionization energy (I), electron affinity (A), chemical potential (μ), hardness (η), electrophilicity index (ω), Fukui functions (f^+ , f^-), condensed dual descriptor (Δf) and local hypersoftness ($s_k^{(2)}$) were calculated, and their linear relationships with the experimental pK_a 's of quinazoline-derived compounds were investigated.

In the previous part, we concluded that M06L and B3LYP functionals give very reasonable correlations between the NPA atomic charges on N₁ atom of quinazoline-derived compounds and their experimental pK_a 's, which points out that these functionals describe the electronic nature of quinazoline derivatives successfully. Thus, the QM descriptors are calculated by M06L and B3LYP functionals in combination with 6-311++G(d,p) basis set employing SMD and CPCM solvation models. As in the previous section, the atomic extraction scheme was taken into account for calculation of the local descriptors ($\dot{X}(N_1)$, $\dot{X}(N_3)$, $\max\{\dot{X}(N_1), \dot{X}(N_3)\}$, $\min\{\dot{X}(N_1), \dot{X}(N_3)\}$, $\text{avg}\{\dot{X}(N_1), \dot{X}(N_3)\}$, where \dot{X} represents f^+ , f^- , Δf or $s^{(2)}$. The analysis of the tested DFT functionals, implicit solvent models, global and local descriptors were performed by constructing the linear equations by a least-square fit:

$$pK_a = a \cdot \dot{X} + b \quad (15)$$

where a and b are the fitted parameters and \dot{X} is the calculated global or local descriptor of the choice. The full detailed results are presented in the Supplementary Information (**Table S6-S30**). **Fig. 6** displays the distribution of Pearson correlation coefficients (R^2) obtained from each

linear fit with respect to the methodology applied by using the 6-311++G(d,p) basis set and global descriptors calculated. According to the results, the strength of the relationship between the global descriptors and the experimental pK_a 's is observed to be slightly dependent on the choice of the DFT functional employed. Overall, M06L/6-311++G(d,p) level of theory in conjunction with CPCM solvation model gives the strongest correlations between the global descriptors calculated and the experimental pK_a 's of the quinazoline derivatives.

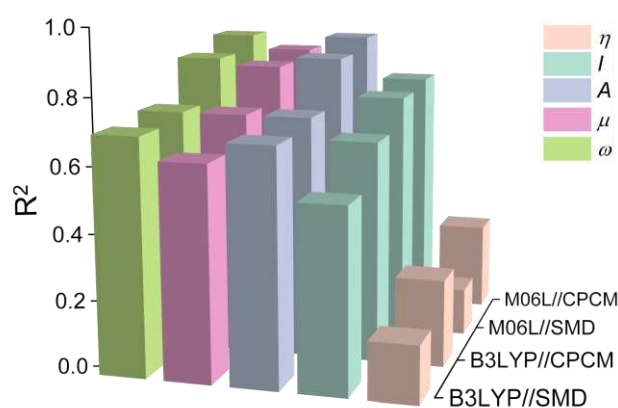


Fig. 6 R^2 distributions of the linear equations for two different DFT functionals, solvation models and five different global QM descriptors considered in this work.

Among the tested global descriptors (A , I , ω , η , μ), very poor relationship between the calculated chemical hardness (η) and the experimental pK_a 's of the quinazoline derivatives was found with a R^2 range of $0.142 \leq R^2 \leq 0.266$. On the other hand, the electron affinity (A), electrophilicity index (ω) and chemical potential (μ) calculated at the M06L/6-311++G(d,p) level of theory employing the CPCM solvent model presented the best regressions ($R^2=0.893$, 0.892 , 0.847 , respectively). These descriptors are related with the molecule's attraction towards the electron pair or bond, consequently, they govern the charge rearrangement during the protonation/deprotonation process of a molecule [84]. They have been used before as a measure for Lewis acidity and their close relation with pK_a has been shown [85-88]. A is the negative of LUMO energy of a molecule, and the lower LUMO energy is associated with larger ability to gain electron, and thus, larger A boosts higher acidity (lower pK_a) [89]. **Fig. 7** displays the

relationship between the calculated A and experimental pK_a 's for the training set. The predicted pK_a 's are calculated by using the parametrized equation in fitting and the results are presented in **Table S31** for the training set and **Table S32** for the test set. The predicted pK_a 's are within the range of ± 1 unit compared to experimental values for 28 molecules in the training set, and the maximum pK_a deviation is 1.47 unit, with mean absolute deviation (MAD) value of 0.44. These results indicate the presence of a strong correlation between experimental pK_a 's and the electron affinity of the quinazoline derived molecules. However, the predicted pK_a 's of nitro substituted quinazolines deviate more than 5 units, as presented in **Table S32**, probably due to the inefficiency in the calculated electron affinities because of their strong electronegativities. The electrophilicity index (ω) which is the extent of the electron deficiency of a molecule, the ability of a molecule to accept the electrons, is inversely proportional to the acidity degree in the same manner as A . **Fig. S1a** represents the relationship between the calculated ω and the experimental pK_a 's which shows that the predictivity power of ω is as accurate as the predictivity power of A with MAD value of 0.48 and MAX- ΔpK_a value of 1.13. Since μ is the negative of electronegativity (χ) [90], as μ value is more negative the molecule tends to gain electrons more easily and as a result, smaller μ promotes the acidity. As can be observed from **Fig. S1b**, which displays the relationship between the calculated μ and the experimental pK_a 's, there is a strong correlation with a MAD value of 0.53. However, MAX- ΔpK_a is calculated to be 2.31 for molecule **4** which is observed to be the strongest outlier in all calculations when the global descriptors are used.

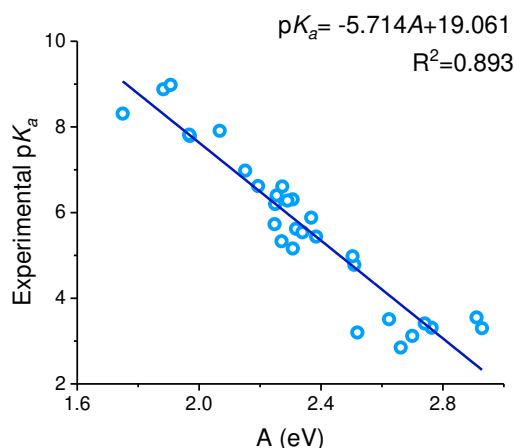


Fig. 7 Linear regression between experimental pK_a and calculated electron affinity (A) at M06L/6-311++G(d,p) level of theory with CPCM solvation model.

Following the successful applications of global descriptors on the pK_a estimation of quinazoline derivatives, the predictive power of local descriptors is examined. The distributions of Pearson correlation coefficients (R^2) obtained from each linear fit with respect to the methodology applied and local descriptors computed by employing five different atomic descriptor extraction schemes are presented in **Fig. 8 (a-d)**. For the local descriptors considered (f_k^+ , f_k^- , Δf_k , $s_k^{(2)}$), both M06L and B3LYP functionals in conjunction with SMD or CPCM solvation models give similar R^2 's for each atomic extraction scheme. As a general trend, no relationship between any of the descriptors calculated for N_3 atom and the experimental pK_a 's of the quinazoline derivatives was found. On the other hand, all the descriptors calculated for N_1 atom give strong correlations with the experimental pK_a 's of the molecules in the training set with a R^2 range of $0.766 \leq R^2 \leq 0.853$. These results emphasize the relative stability of N_1 compared to N_3 as an acid in accordance with the results obtained in the previous section. When the nucleophilic (f_k^+) and electrophilic (f_k^-) Fukui functions are considered (**Fig. 8a-b**), for the two of the functionals and the solvent models employed, the linearities obtained with $f^+(N_1)$ and $f^-(N_1)$ are observed to be the most accurate ($0.786 \leq R^2 \leq 0.833$ and $0.766 \leq R^2 \leq 0.812$, respectively), and no relationships were found when $\max\{f^+(N_1), f^+(N_3)\}$, $\max\{f^-(N_1), f^-(N_3)\}$,

$f^+(\text{N}_3)$ and $f^-(\text{N}_3)$ are considered. Based on the frontier-electron theory of chemical reactivity [74], f^+ gives information about the sites in a molecule which are prone to nucleophilic attack and f^- identifies the sites which are more susceptible to electrophilic attacks, which can be interpreted as large f^+ values indicate the electron acceptance and large f^- values indicate the electron donation. Fukui functions can provide useful information about the acid-base reactions, since proton acts as an electrophile which attaches itself to the lone pairs located on the N atoms of quinazoline derivatives (the nucleophilic site). The relationship between the calculated f^+ values on N_1 atom and the experimental $\text{p}K_a$'s of quinazoline molecules are observed to be directly proportional, whereas the relationship between the calculated f^- values on N_1 atom and the experimental $\text{p}K_a$'s are inversely proportional to each other, as displayed in **Fig. S2(a-b)**. According to the data obtained, the calculated fukui functions for N_1 and N_3 atoms are all negative irrespective of the functionals and solvation models employed and sometimes there are overlapping results. As stated by Martinez-Araya [91], condensed dual descriptor (Δf_k) and its analogue local hypersoftness ($s_k^{(2)}$) can provide more accurate conclusions about the preferred sites for nucleophilic and electrophilic attacks. As can be seen from **Fig. 8(c-d)**, the relationships obtained with $\max\{\Delta f(\text{N}_1), \Delta f(\text{N}_3)\}$, $\max\{s^{(2)}(\text{N}_1), s^{(2)}(\text{N}_3)\}$, $\Delta f(\text{N}_1)$ and $s^{(2)}(\text{N}_1)$ are linear ($0.840 \leq R^2 \leq 0.863$, $0.847 \leq R^2 \leq 0.883$, $0.795 \leq R^2 \leq 0.840$, and $0.805 \leq R^2 \leq 0.853$, respectively), as opposed to the relationships with $\min\{\Delta f(\text{N}_1), \Delta f(\text{N}_3)\}$, $\min\{s^{(2)}(\text{N}_1), s^{(2)}(\text{N}_3)\}$, $\Delta f(\text{N}_3)$ and $s^{(2)}(\text{N}_3)$.

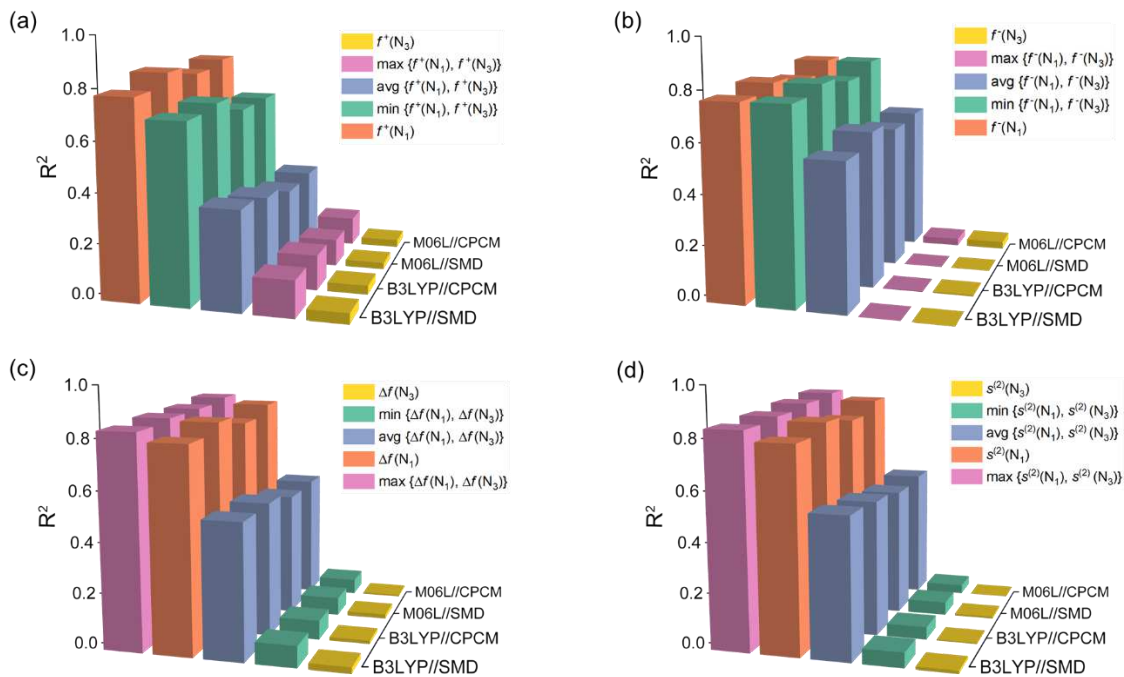


Fig. 8 R^2 distributions of the linear equations for two different DFT functionals, solvation models and four different atomic descriptor extraction schemes for (a) f_k^+ , (b) f_k^- , (c) Δf_k and (d) $s_k^{(2)}$.

Overall, the best linearity was obtained when the highest local hyper softness ($s_k^{(2)}$) calculated for N_1 and N_3 atoms with M06L/6-311++G(d,p) level using the CPCM implicit solvent model is considered, with a R^2 of 0.883, as illustrated in **Fig. 9**. The predicted pK_a 's are within the range of ± 1 unit compared to experimental values for 27 molecules in the training set, and the maximum pK_a deviation is 1.58 unit, with mean absolute deviation (MAD) value of 0.47 which indicates the presence of a strong correlation between experimental pK_a 's and the maximum value of the calculated local hyper softness on the N atoms of the quinazoline derived molecules. Since $s_k^{(2)}$ strongly resembles the Δf_k , the linear relationship obtained from calculated Δf values is as accurate as $s_k^{(2)}$ when the highest Δf calculated for N_1 and N_3 atoms at the same level of theory is considered, with $R^2=0.863$, $MAD=0.53$ and $MAX-\Delta pK_a=1.70$ (**Fig. S3**). The stronger relationship of $s_k^{(2)}$ is accounted on the fact that local hypersoftness

takes into account the size differences. For both cases, the predicted pK_a 's of molecules **25** and **34**, which carry methoxy substituents on 7 and 6 positions of quinazoline ring, deviate the most from experimental pK_a 's. On the other hand, none of the derived models, concerning the conceptual DFT descriptors, work well for nitro substituted quinazoline molecules.

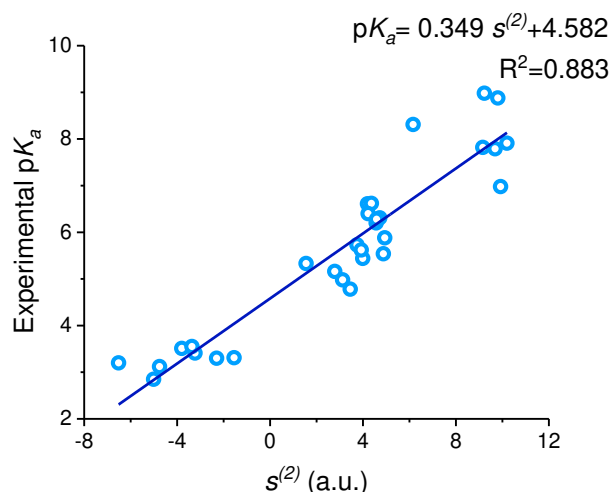


Fig. 9 Linear regression between experimental pK_a and calculated local hyper-softness ($s^{(2)} = \max\{s^{(2)}(N_1), s^{(2)}(N_3)\}$) at M06L/6-311++G(d,p) level of theory with CPCM solvation model.

The prediction power of our methods developed in sections 3.1 and 3.2 is additionally assessed by comparing the predicted pK_a 's of quinazoline derivatives with ChemAxon [92], which is a commercial pK_a prediction tool widely used in drug development processes. With ChemAxon, the micro pK_a 's are identified by the usage of the empirically calculated partial atomic charges in molecules as parameters. Although the predicted pK_a 's of the molecules in the training set by ChemAxon listed in **Table 5** are mostly found to be in good agreement with the experimental values, 7 molecules out of 30 deviate more than 1 unit. On the other hand, the correlation between the predicted and experimental pK_a 's are observed to be weaker ($R^2=0.820$) with higher MAD value (0.75) compared to the methodologies developed in this study.

Table 5 Molecule number, experimental pK_a , predicted pK_a by using the ChemAxon's pK_a Prediction Tool, difference between experimental and predicted pK_a values for the training set (M06L/6-311++G(d,p)//CPCM).

Molecule number	pK_a (exp.)	pK_a (pred.)	ΔpK_a
1	3.31	3.28	-0.03
3	5.74	5.54	-0.20
4	3.21	3.95	0.74
7	2.86	3.42	0.56
8	3.52	2.6	-0.92
9	3.13	3.44	0.31
10	3.42	2.65	-0.77
11	3.56	2.35	-1.21
12	3.31	1.72	-1.59
15	5.45	5.73	0.28
16	5.17	5.83	0.66
17	5.34	5.77	0.43
18	5.63	6.3	0.67
21	4.79	4.84	0.05
22	4.99	4.82	-0.17
23	7.80	7.21	-0.59
24	7.83	7.26	-0.57
25	8.32	7.78	-0.54
26	6.99	6.29	-0.70
27	6.32	5.52	-0.80
28	6.62	5.55	-1.07
29	6.21	6.05	-0.16
32	8.89	6.99	-1.90
33	7.92	6.07	-1.85
34	8.99	7.04	-1.95
36	6.63	5.62	-1.01
37	6.41	5.59	-0.82
38	6.29	5.52	-0.77
39	5.89	5.51	-0.38
40	5.55	4.65	-0.90

4. Conclusions

In this study we have undertaken the task of testing five different DFT functionals (B3LYP, M06L, M06-2X, ω B97XD, MN12SX), two basis sets (6-31+G(d) and 6-311++G(d,p)) and two solvation models (SMD and CPCM) to construct relationships between quantum mechanical descriptors and experimental pK_a 's of quinazoline derivatives. We have demonstrated that the calculated atomic charge on N₁ atom ($Q(N_1)$), electron affinity (A),

electrophilicity index (ω), chemical potential (μ), the highest local hyper softness calculated for N_1 and N_3 atoms ($\max\{s^{(2)}(N_1), s^{(2)}(N_3)\}$) and the highest condensed dual descriptor calculated for N_1 and N_3 atoms ($\max\{\Delta f(N_1), \Delta f(N_3)\}$) have strong correlations with the experimental pK_a 's. The most accurate pK_a 's are obtained via correlations with calculated atomic charge on N_1 atom (NPA/ M06L/6-311++G(d,p) // CPCM) with $R^2=0.927$. Furthermore, the Boltzmann distribution of the ground state stationary structures have been taken into account by considering the Gibbs free energies for all the conformations with the specified methodology, this study has yielded similar results as the ones obtained with GS conformation via correlation with atomic charges on N_1 . Correlation with the electronic charge on N_1 rather than N_3 has been attributed to the protonation of the former being pK_a determining. The results of Conceptual DFT have indicated the presence of a strong correlation between experimental pK_a 's and the electron affinity of the quinazoline derived molecules. Nevertheless, the constructed models, especially conceptual DFT descriptors, fail to estimate the pK_a 's of nitro substituted quinazoline molecules (M06L/6-311++G(d,p) // CPCM). Overall, we believe that generating the data from multiple linear regressions using the descriptors suggested in this study can be useful in data production for successful machine learning models for pK_a prediction.

AUTHOR INFORMATION

Corresponding Authors

Viktorya Aviyente — *Bogazici University, Department of Chemistry, 34342, Bebek, Istanbul, Turkey*; orcid.org/0000-0001-9430-4096

E-mail: aviye@boun.edu.tr

Fatma Ahu Akin — *Bogazici University, Department of Chemistry, 34342, Bebek, Istanbul, Turkey*; orcid.org/0000-0001-9391-7026

E-mail: ahu.akin@boun.edu.tr

Authors

Melisa Kiran — Bogazici University, Department of Chemistry, 34342, Bebek, Istanbul, Turkey

Zeynep Pinar Haslak — Bogazici University, Department of Chemistry, 34342, Bebek, Istanbul, Turkey; Université de Reims Champagne Ardenne, 51687, Reims, France;
orcid.org/0000-0002-2850-9816

Halit Ates — *Bogazici University, Department of Chemistry, 34342, Bebek, Istanbul, Turkey;*
orcid.org/0009-0006-7448-4556

Author Contributions

MK: Investigation, Methodology, Software, Writing – original draft, Writing – review & editing.

ZPH: Investigation, Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing.

HA: Investigation, Methodology, Software, Writing – original draft, Writing – review & editing.

FAA: Conceptualization, Supervision, Methodology, Writing – original draft, Writing – review & editing.

VA: Conceptualization, Supervision, Methodology, Writing – original draft, Writing – review & editing.

MK and ZPH contributed equally.

Conflicts of interest

The authors declare no competing financial interest.

Acknowledgments

The research was funded by Bogazici University Research Fund through Project 19143. Computational resources were partially funded by the National Center for High Performance Computing of Turkey (UHEM) with grant number 5010142021.

Data availability

† Supplementary Information (SI) available: Linear regressions between experimental pK_a and calculated electrophilicity index (ω), chemical potential (μ), nucleophilic fukui function (f^+) and electrophilic fukui function (f^-) on N_1 atom, and condensed dual function ($\Delta f = \max\{\Delta f(N_1), \Delta f(N_3)\}$); R^2 , MAD, and MAX- ΔpK_a results for the training set with five different atomic charge descriptors for five different DFT functionals, two basis sets and two solvation models, with five different global descriptors and twenty different local descriptors for two different DFT functionals and solvation models; predicted pK_a 's by using the linear equation obtained from the linear fit between experimental pK_a 's and A for the training and test sets.

The cartesian coordinates of the test set and the training set can be found on <https://doi.org/10.5281/zenodo.7870894>.

References

1. Wakeling AE, Guy SP, Woodburn JR, Ashton SE, Curry BJ, Barker AJ, Gibson KH (2002) *Cancer Res* 62(20):5749–5754.
2. Vasdev N, Dorff PN, Gibbs AR, Nandanan E, Reid LM, O'Neil JP, VanBrocklin HF (2005) *J Labelled Comp Radiopharm* 48(2):109–115. <https://doi.org/10.1002/jlcr.903>.
3. Al-Rashood ST, Aboldahab IA, Nagi MN, Abouzeid LA, Abdel-Aziz AAM, Abdel-hamide SG, Youssef KM, Al-Obaid AM, El-Subbagh HI (2006) *Bioorg Med Chem* 14(24):8608–8621. <https://doi.org/10.1016/j.bmc.2006.08.030>.
4. Chandregowda V, Kush AK, Chandrasekara Reddy G (2009) *Eur J Med Chem* 44(7):3046–3055. <https://doi.org/10.1016/j.ejmech.2008.07.023>.
5. Gupta V, Kashaw SK, Jatav V, Mishra P (2008) *Med Chem Res* 17(2–7):205–211. <https://doi.org/10.1007/s00044-007-9054-3>.

6. Antipenko L, Karpenko A, Kovalenko S, Katsev A, Komarovska-Porokhnyavets E, Novikov V, Chekotilo A (2009) *Chem Pharm Bull (Tokyo)* 57(6):580–585. <https://doi.org/10.1248/cpb.57.580>.
7. Rohini R, Muralidhar Reddy P, Shanker K, Hu A, Ravinder V (2010) *Eur J Med Chem* 45(3):1200–1205. <https://doi.org/10.1016/j.ejmech.2009.11.038>.
8. Xie YQ, Huang ZL, Yan HD, Li J, Ye LY, Che LM, Tu S (2015) *Chem Biol Drug Des* 85(6):743–755. <https://doi.org/10.1111/cbdd.12460>.
9. Lv X, Yang L, Fan Z, Bao X (2018) *J Saudi Chem Soc* 22(1):101–109. <https://doi.org/10.1016/j.jscs.2017.07.008>.
10. Giri RS, Thaker HM, Giordano T, Williams J, Rogers D, Sudersanam V, Vasu KK (2009) *Eur J Med Chem* 44(5):2184–2189. <https://doi.org/10.1016/j.ejmech.2008.10.031>.
11. Alagarsamy V, Raja Solomon V, Dhanabal K (2007) *Bioorg Med Chem* 15(1):235–241. <https://doi.org/10.1016/j.bmc.2006.09.065>.
12. Chandrika PM, Yakaiah T, Narsaiah B, Sridhar V, Venugopal G, Rao V, Kumar P, Murthy SN, Raghu A, Rao (2009) *Ind J Chem* 48:840–847.
13. Rather B, Paneersalvam P, Raj T, Ishar MPS, Singh B, Sharma V (2010) *Indian J Pharm Sci* 72(3):375. <https://doi.org/10.4103/0250-474X.70488>.
14. Khosropour AR, Mohammadpoor-Baltork I, Ghorbankhani H (2006) *Tetrahedron Lett* 47(21):3561–3564. <https://doi.org/10.1016/j.tetlet.2006.03.079>.
15. Nandy P, Vishalakshi MT, Bhat AR (2006) *Indian J of Heterocycl Chem* 15:293–294.
16. Alvarado M, Barceló M, Carro L, Masaguer CF, Raviña E (2006) *Chem Biodivers* 3(1):106–117. <https://doi.org/10.1002/cbdv.200690001>.
17. Jiang S, Zeng Q, Gettayacamin M, Tungtaeng A, Wannaying S, Lim A, Hansukjariya P, Okunji CO, Zhu S, Fang D (2005) *Antimicrob Agents Chemother* 49(3):1169–1176. <https://doi.org/10.1128/AAC.49.3.1169-1176.2005>.
18. Sen D, Banerjee A, Ghosh A, Chatterjee T (2010) *J Adv Pharm Technol Res* 1(4):401. <https://doi.org/10.4103/0110-5558.76439>.
19. Sasmal S, Balasubrahmanyam D, Kanna Reddy H. R, Balaji G, Srinivas G, Cheera S, Abbineni C, Sasmal PK, Khanna I, Sebastian VJ, Jadhav VP, Singh MP, Talwar R, Suresh J, Shashikumar D, Harinder Reddy K, Sihorkar V, Frimurer TM, Rist Ø, Elster L, Högberg T (2012) *Bioorg Med Chem Lett* 22(9):3163–3167. <https://doi.org/10.1016/j.bmcl.2012.03.049>.
20. Albert A, Armarego WLF, Spinner E (1961) *J Chem Soc* 2689–2696. <https://doi.org/10.1039/JR9610002689>.
21. Armarego, WLF (1963) *J Chem Soc* 253–309. [https://doi.org/10.1016/S0065-2725\(08\)60527-9](https://doi.org/10.1016/S0065-2725(08)60527-9).
22. Goodman, LS (1996) Goodman and Gilman's The Pharmacological Basis of Therapeutics (11th ed) Brunton LL, Lazo JS, Parker KL, New York.

23. Reijenga J, van Hoof A, van Loon A, Teunissen B (2013) *Anal Chem Insights* 8. <https://doi.org/10.4137/ACI.S12304>.
24. Thapa B, Schlegel HB (2016) *J Phys Chem A* 120(28):5726–5735. <https://doi.org/10.1021/acs.jpca.6b05040>.
25. Casasnovas R, Ortega-Castro J, Frau J, Donoso J, Muñoz F (2014) *Int J Quantum Chem* 114(20):1350–1363. <https://doi.org/10.1002/qua.24699>.
26. Pliego JR, Riveros JM (2002) *J Phys Chem A* 106(32):7434–7439. <https://doi.org/10.1021/jp025928n>.
27. Sastre S, Casasnovas R, Muñoz F, Frau J (2016) *Phys Chem Chem Phys* 18(16):11202–11212. <https://doi.org/10.1039/c5cp07053h>.
28. Pliego JR, Riveros JM (2002) *J Phys Chem A* 106(32):7434–7439. <https://doi.org/10.1021/jp025928n>.
29. Sandoval-Lira J, Mondragón-Solórzano G, Lugo-Fuentes LI, Barroso-Flores J (2020) *J Chem Inf Model* 60(3):1445–1452. <https://doi.org/10.1021/acs.jcim.9b01173>.
30. Chipman DM (2006) *J Chem Phys* 124(22):224111. <https://doi.org/10.1063/1.2203068>.
31. Liptak MD, Gross KC, Seybold PG, Feldgus S, Shields GC (2002) *J Am Chem Soc* 124(22):6421–6427. <https://doi.org/10.1021/ja012474j>.
32. Toth, AM, Liptak MD, Phillips DL, Shields GC (2001) *J Chem Phys* 114(10):4595. <https://doi.org/10.1063/1.1337862>.
33. Liptak MD, Shields GC (2001) *J Am Chem Soc* 123(30):7314–7319. <https://doi.org/10.1021/ja010534f>.
34. Liptak MD, Shields GC (2001) *Int J Quantum Chem* 85(6):727–741.
35. Klamt A, Schüürmann G (1993) *J Chem Soc Perkin Trans* 2:799–805. <https://doi.org/10.1039/P29930000799>.
36. Sastre S, Frau J, Glossman-Mitnik D (2017) *Molecules* 22(3):458. <https://doi.org/10.3390/molecules22030458>.
37. Silva CO, da Silva EC, Nascimento MAC (2000) *J Phys Chem A* 104(11):2402–2409. <https://doi.org/10.1021/jp992103d>.
38. Adam KR (2002) *J Phys Chem A* 106(49):11963–11972. <https://doi.org/10.1021/jp026577f>.
39. Saracino GAA, Improta R, Barone V (2003) *Chem Phys Lett* 373(3–4):411–415. [https://doi.org/10.1016/S0009-2614\(03\)00607-9](https://doi.org/10.1016/S0009-2614(03)00607-9).
40. Almerindo GI, Tondo DW, Pliego JR (2004) *J Phys Chem A* 108(1):166–171. <https://doi.org/10.1021/jp0361071>.
41. Klamt A, Eckert F, Diedenhofen M, Beck ME (2003) *J Phys Chem A* 107(44):9380–9386. <https://doi.org/10.1021/jp034688o>.

42. Magill AM, Cavell KJ, Yates BF (2004) *J Am Chem Soc* 126(28):8717–8724. <https://doi.org/10.1021/ja038973x>.
43. Vianello R, Maksić ZB (2005) *Eur J Org Chem* 2005(16):3571–3580. <https://doi.org/10.1002/ejoc.200500153>.
44. Caballero-García G, Mondragón-Solórzano G, Torres-Cadena R, Díaz-García M, Sandoval-Lira J, Barroso-Flores J (2018) *Molecules* 24(1):79. <https://doi.org/10.3390/molecules24010079>.
45. Pracht P, Wilcken R, Udvarhelyi A, Rodde S, Grimme S (2018) *J Comput Aided Mol Des* 32(10):1139–1149. <https://doi.org/10.1007/s10822-018-0145-7>.
46. Rayne S, Forest K (2010) *J Mol Struct THEOCHEM* 949(1–3):60–69. <https://doi.org/10.1016/j.theochem.2010.03.003>.
47. Uudsemaa M, Kanger T, Lopp M, Tamm T (2010) *Chem Phys Lett* 485(1–3):83–86. <https://doi.org/10.1016/j.cplett.2009.12.006>.
48. Dissanayake DP, Senthilnithy R (2009) *J Mol Struct THEOCHEM* 910(1–3):93–98. <https://doi.org/10.1016/j.theochem.2009.06.021>.
49. Delgado EJ (2009) *Chem Phys Lett* 471(1–3):133–135. <https://doi.org/10.1016/j.cplett.2009.02.030>.
50. Namazian M, Zakery M, Noorbala MR, Coote ML (2008) *Chem Phys Lett* 451(1–3):163–168. <https://doi.org/10.1016/j.cplett.2007.11.088>.
51. Sastre S, Casasnovas R, Muñoz F, Frau J (2013) *Theor Chem Acc* 132(2):1310. <https://doi.org/10.1007/s00214-012-1310-z>.
52. Biswas A, Lo R, Ganguly B (2013) *Synlett* 24(19):2519–2524. <https://doi.org/10.1055/s-0033-1339545>.
53. Fındık BK, Haslak ZP, Arslan E, Aviyente V (2021) *J Comput Aided Mol Des* 35(7):841–851. <https://doi.org/10.1007/s10822-021-00402-9>.
54. Ugur I, Marion A, Parant S, Jensen JH, Monard G (2014) *J Chem Inf Model* 54(8):2200–2213. <https://doi.org/10.1021/ci500079w>.
55. Haslak ZP, Zareb S, Dogan I, Aviyente V, Monard G (2021) *J Chem Inf Model* 61(6):2733–2743. <https://doi.org/10.1021/acs.jcim.1c00059>.
56. Frau J, Muñoz F, Glossman-Mitnik DA (2016) *Molecules* 21(12):1650. <https://doi.org/10.3390/molecules21121650>.
57. Patai S (1964) *Isr J Chem* 2(1):35–35. <https://doi.org/10.1002/ijch.196400009>.
58. Albert A, Goldacre R, Phillips J (1948) *J Chem Soc*:2240–2249. <https://doi.org/10.1039/jr9480002240>.
59. Armarego, WLF (1963) *J Chem Soc*:61–572. <https://doi.org/10.1039/jr9620000561>.
60. Albert A, Phillips JN (1956) *J Chem Soc*:1294–1304. <https://doi.org/10.1039/jr9560001294>.

61. Zielinski W, Kudelko A (2004) *Arkivoc* 2005(5):66–82. <https://doi.org/10.3998/ark.5550190.0006.507>.
62. Işık M, Rustenburg AS, Rizzi A, Gunner MR, Mobley DL, Chodera JD (2021) *J Comput Aided Mol Des* 35(2):131–166. <https://doi.org/10.1007/s10822-020-00362-6>.
63. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Scalmani G, Barone V, Petersson GA, Nakatsuji H, Li X, Caricato M, Marenich AV, Bloino J, Janesko BG, Gomperts R, Mennucci B, Hratchian HP, Ortiz JV, Izmaylov AF, Sonnenberg JL, Williams-Young D, Ding F, Lipparini F, Egidi F, Goings J, Peng B, Petrone A, Henderson T, Ranasinghe D, Zakrzewski VG, Gao J, Rega N, Zheng G, Liang W, Hada M, Ehara M, Toyota K, Fukuda R, Hasegawa J, Ishida M, Nakajima T, Honda Y, Kitao O, Nakai H, Vreven T, Throssell K, Montgomery Jr. JA, Peralta JE, Ogliaro F, Bearpark MJ, Heyd JJ, Brothers EN, Kudin KN, Staroverov VN, Keith TA, Kobayashi R, Normand J, Raghavachari K, Rendell AP, Burant J. C, Iyengar SS, Tomasi J, Cossi M, Millam JM, Klene M, Adamo C, Cammi R, Ochterski JW, Martin RL, Morokuma K, Farkas O, Foresman JB, Fox DJ (2016) *Gaussian16 Revision A.03*.
64. Shao Y, Molnar LF, Jung Y, Kussmann J, Ochsenfeld C, Brown ST, Gilbert ATB, Slipchenko LV, Levchenko SV, O'Neill DP, DiStasio Jr RA, Lochan RC, Wang T, Beran GJO, Besley NA, Herbert JM, Yeh Lin C, Van Voorhis T, Hung Chien S, Sodt A, Steele RP, Rassolov VA, Maslen PE, Korambath PP, Adamson RD, Austin B, Baker J, Byrd EFC, Dachsel H, Doerksen RJ, Dreuw A, Dunietz BD, Dutoi AD, Furlani TR, Gwaltney SR, Heyden A, Hirata S, Hsu CP, Kedziora G, Khalliulin RZ, Klunzinger P, Lee AM, Lee MS, Liang W, Lotan I, Nair N, Peters B, Proynov EI, Pieniazek PA, Min Rhee Y, Ritchie J, Rosta E, David Sherrill C, Simmonett AC, Subotnik JE, Lee Woodcock III H, Zhang W, Bell AT, Chakraborty AK, Chipman DM, Keil FJ, Warshel A, Hehre WJ, Schaefer III HF, Kong J, Krylov AI, Gill PMW, Head-Gordon M (2006) *Phys Chem Chem Phys* 8(27):3172–3191. <https://doi.org/10.1039/B517914A>.
65. Lee C, Yang W, Parr RG (1988) *Phys Rev B* 37(2):785–789. <https://doi.org/10.1103/PhysRevB.37.785>.
66. Zhao Y, Truhlar DG (2006) *J Chem Phys* 125(19):194101. <https://doi.org/10.1063/1.2370993>.
67. Zhao Y, Truhlar DG (2008) *Theor Chem Acc* 120(1–3):215–241. <https://doi.org/10.1007/s00214-007-0310-x>.
68. Chai JD, Head-Gordon M (2008) *Phys Chem Chem Phys* 10(44):6615–6620. <https://doi.org/10.1039/b810189b>.
69. Peverati R, Truhlar DG (2012) *Phys Chem Chem Phys* 14(47):16187–16191. <https://doi.org/10.1039/c2cp42576a>.
70. Ziesche P, Kurth S, Perdew JP (1998) *Comput Mater Sci* 11(2):122–127. [https://doi.org/10.1016/S0927-0256\(97\)00206-1](https://doi.org/10.1016/S0927-0256(97)00206-1).
71. Reed AE, Weinstock RB, Weinhold F (1985) *J Chem Phys* 83(2):735–746. <https://doi.org/10.1063/1.449486>.
72. Parr RG, Pearson RG (1983) *J Am Chem Soc* 105(26):7512–7516. <https://doi.org/10.1021/ja00364a005>.

73. Parr RG, Szentpály L, Liu S (1999) *J Am Chem Soc* 121(9):1922–1924. <https://doi.org/10.1021/ja983494x>.
74. Parr RG, Yang W (1984) *J Am Chem Soc* 106(14):4049–4050. <https://doi.org/10.1021/ja00326a036>.
75. Toro-Labbé A (2007) *Theoretical Aspects of Chemical Reactivity*, 1st ed, Amsterdam, Vol. 19.
76. Morell C, Grand A, Toro-Labbé A (2005) *J Phys Chem A* 109(1):205–212. <https://doi.org/10.1021/jp046577a>.
77. Morell C, Grand A, Toro-Labbé A (2006) *Chem Phys Lett* 425(4–6):342–346. <https://doi.org/10.1016/j.cplett.2006.05.003>.
78. Cárdenas C, Rabi N, Ayers PW, Morell C, Jaramillo P, Fuentealba P (2009) *J Phys Chem A* 113(30):8660–8667. <https://doi.org/10.1021/jp902792n>.
79. Morell C, Ayers PW, Grand A, Gutiérrez-Oliva S, Toro-Labbé A (2008) *Phys Chem Chem Phys* 10(48):7239–7246. <https://doi.org/10.1039/b810343g>.
80. Ayers PW, Morell C, De Proft F, Geerlings P (2007) *Chem Eur J* 13(29):8240–8247. <https://doi.org/10.1002/chem.200700365>.
81. Gross KC, Seybold PG, Hadad CM (2002) *Int J Quantum Chem* 90(1):445–458. <https://doi.org/10.1002/qua.10108>.
82. Seybold PG, Shields GC (2015) *Wiley Interdiscip Rev Comput Mol Sci* 5(3):290–297. <https://doi.org/10.1002/wcms.1218>.
83. Sterck BD, Vaneerdeweg R, Prez FD, Waroquier M, Speybroeck VV (2010) *Macromolecules* 43(2):827–836. <https://doi.org/10.1021/ma9018747>.
84. Lohr LL (1984) *J Phys Chem* 88(16):3607–3611. <https://doi.org/10.1021/j150660a046>.
85. Gupta K, Roy DR, Subramanian V, Chattaraj PK (2007) *J Mol Struct THEOCHEM* 812(1–3):13–24. <https://doi.org/10.1016/j.theochem.2007.02.013>.
86. Frau J, Flores-Holguín N, Glossman-Mitnik D (2018) *Mar Drugs* 16(9):302. <https://doi.org/10.3390/md16090302>.
87. Rajak SK, Islam N, Ghosh DC (2011) *J Quantum Inf Sci* 01(02):87–95. <https://doi.org/10.4236/jqis.2011.12012>.
88. Gupta K, Giri S, Chattaraj PK (2008) *New J Chem* 32(11):1945. <https://doi.org/10.1039/b803655a>.
89. Sastre G, Corma A (1999) *Chem Phys Lett* 302(5–6):447–453. [https://doi.org/10.1016/S0009-2614\(99\)00177-3](https://doi.org/10.1016/S0009-2614(99)00177-3).
90. Parr RG, Donnelly RA, Levy M, Palke WE (1978) *J Chem Phys* 68(8):3801–3807. <https://doi.org/10.1063/1.436185>.
91. Martínez-Araya JI (2015) *J Math Chem* 53(2):451–465. <https://doi.org/10.1007/s10910-014-0437-7>.

92. Marvin (2023): “Marvin was used for drawing and characterizing chemical structures, Marvin 23.2.0, ChemAxon (<http://www.chemaxon.com>)”.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryInformation1.pdf](#)