



Review

Quantitative Structure–Toxicity Relationship in Bioactive Molecules from a Conceptual DFT Perspective

Ranita Pal ¹, Shanti Gopal Patra ² and Pratim Kumar Chattaraj ^{2,*}¹ Advanced Technology Development Centre, Indian Institute of Technology Kharagpur, Kharagpur 721302, India² Department of Chemistry, Indian Institute of Technology Kharagpur, Kharagpur 721302, India

* Correspondence: pkc@chem.iitkgp.ac.in

Abstract: The preclinical drug discovery stage often requires a large amount of costly and time-consuming experiments using huge sets of chemical compounds. In the last few decades, this process has undergone significant improvements by the introduction of quantitative structure-activity relationship (QSAR) modelling that uses a certain percentage of experimental data to predict the biological activity/property of compounds with similar structural skeleton and/or containing a particular functional group(s). The use of machine learning tools along with it has made life even easier for pharmaceutical researchers. Here, we discuss the toxicity of certain sets of bioactive compounds towards *Pimephales promelas* and *Tetrahymena pyriformis* in terms of the global conceptual density functional theory (CDFT)-based descriptor, electrophilicity index (ω). We have compared the results with those obtained by using the commonly used hydrophobicity parameter, $\log P$ (where P is the n -octanol/water partition coefficient), considering the greater ease of computing the ω descriptor. The Human African trypanosomiasis (HAT) curing activity of 32 pyridyl benzamide derivatives is also studied against *Trypanosoma brucei*. In this review article, we summarize these multiple linear regression (MLR)-based QSAR studies in terms of electrophilicity (ω , ω^2) and hydrophobicity ($\log P$, $(\log P)^2$) parameters.

Keywords: conceptual density functional theory; electrophilicity index; hydrophobicity; QSAR; multiple linear regression

**Citation:** Pal, R.; Patra, S.G.; Chattaraj,P.K. Quantitative Structure–Toxicity Relationship in Bioactive Molecules from a Conceptual DFT Perspective. *Pharmaceuticals* **2022**, *15*, 1383.<https://doi.org/10.3390/ph15111383>

Academic Editor: Osvaldo Andrade Santos-Filho

Received: 28 September 2022

Accepted: 7 November 2022

Published: 10 November 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the progress of modern science, especially biochemistry and synthetic organic chemistry, the field of drug discovery has witnessed huge advances in the use of statistical approaches. The process of clinical drug screening has always been an exhaustive and time-consuming process. Thus, the involvement of statistics and computational techniques in predicting the activity of potential drug molecules with structural similarities has greatly accelerated the process.

In this regard, quantitative structure-activity relationship (QSAR) modeling that uses a certain percentage of experimental data plays a crucial role to predict the biological activity/property of compounds with similar structural skeletons and/or containing a particular functional group(s) [1–6]. They are applicable in the fields of molecular modelling, drug discovery, eco-toxicology, antitumor treatment, etc. Hansch et al. [7] in 1962 first reported a QSAR-based study through the correlation between n -octanol/water partition coefficient and biological activity exhibited by those compounds. Many scientists have since followed in his footsteps and reported several such experimental and theoretical studies that have shaped the field of modern QSAR [8–36].

Toxicity of organic compounds is measured in terms of parameters like pLC_{50} , pIC_{50} , and $pIGC_{50}$ which essentially measure the negative logarithm of the concentration of the toxin needed to kill half of the target population. These measures of toxicity have been extensively used in QSAR studies [37–42]. In toxicology studies, the aforementioned

concentration depends on the toxicokinetic and toxicodynamic processes in a linear fashion. A library of descriptors, namely, quantum chemical and electronic energy descriptors, thermodynamic potentials, hydrophobicity, shape, topology, etc., have been used in QSAR-based studies [43]. Several global and local reactivity parameters are available within the purview of conceptual density functional theory (CDFT) that act as effective descriptors for predicting biological and toxicological activities [44–49]. Electrophilicity index (ω) is one such important descriptor that quantifies the electro-deficient nature of a molecule [50,51]. It has been previously utilized in predicting the biological activity (in terms of relative binding affinity values) of testosterone and estrogen derivatives [40], and the toxicity of benzidine [52] and polychlorinated biphenyls [38], to name a few. In recent times, machine learning is also becoming an integral part of drug design and QSAR studies in general [53–57].

This article provides a concise discussion on the toxicity of certain sets of bioactive compounds in terms of the global and local electrophilicity indices, and the simple yet effective methods utilized in developing the models.

2. Theoretical Background

According to Hansch [58,59], any correlation drawn between the biological activity of any system and its physicochemical properties ideally includes a steric part, a hydrophobic part, and an electronic part, to model a proper mathematical representation of the bioactivity. Now, the percentages of these three will vary depending on the mechanism, the receptor site, the mode of action, and several other factors. The hydrophobicity (or lipophilicity) parameter has gained a lot of importance in its usage as a descriptor owing to the fact that it can effectively describe the protoplasmic environment within a biosystem. However, using only the hydrophobic parameter is not enough in several types of mechanisms. Both receptor- and non-receptor-mediated toxicological reactions can occur either via covalent or noncovalent mechanisms. The latter form is especially important in aqueous toxicity, where the inclusion of the electronic parameter is required to produce a statistically relevant SAR model since hydrophobicity alone cannot describe the narcotic properties of polar chemicals. Thus, evaluating the electronic state of a compound is extremely useful in the prediction of its biological/toxicological properties, especially in reaction mechanisms driven by electrophile–nucleophile interactions. In this regard, we have resorted to CDFT-based chemical concepts.

For a system containing N -electrons, chemical reactivity parameters such as electronegativity (χ) [60,61] and chemical hardness (η) [62] are obtained within the domain of CDFT [63]. The former describes the ability of an atom in a molecule to attract bonded electrons towards itself and is perhaps the most fundamental descriptor required for the analysis of chemical activity. In DFT, this parameter is defined as the first derivative of the total energy (E) with respect to the total number of electrons (N) while the external potential ($v(\vec{r})$) is kept constant. Parr [64] made a connection of this formulation with the negative of chemical potential (μ) which describes the escaping tendency of the electron cloud (see Equation (1)). Over the years many scales of electronegativity have been developed, such as those provided by Pauling [60,65], Mulliken [66], Allred-Rochow [67], etc. The second derivative of E with respect to N at constant $v(\vec{r})$, on the other hand, describes the chemical hardness (η) of a system. This, by extension to Equation (1), becomes the first derivative of μ with respect to N (see Equation (2)). These definitions of electronegativity, chemical potential, and hardness lead up to the description of the electrophilicity index (ω) [64,68,69]. It first originated in the field of organic chemistry by Ingold's [70,71] classification of organic chemical reactions in two groups: the electron-deficient species (electrophiles) characterized by their electrophilicity, and electron-rich species (nucleophiles) characterized by their nucleophilicity. These qualitative descriptions came long before any mathematical representation of electrophilicity was known. Finally, after Maynard's [72] qualitative description of ω as the ratio between the square of electronegativity and hardness, Parr

et al. [51] gave their quantitative description and defined it as the “electrophilic power” of a ligand (see Equation (3)), comparing it to the definition of power in classical electrostatics.

$$\chi = -\mu = \left(\frac{\partial E}{\partial N} \right)_{v(\vec{r})} \quad (1)$$

$$\eta = \left(\frac{\partial^2 E}{\partial N^2} \right)_{v(\vec{r})} = \left(\frac{\partial \mu}{\partial N} \right)_{v(\vec{r})} \quad (2)$$

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

During computation, we evoke the finite difference approximation [73] where the ∂N is approximated as the transfer of one electron to or from the neutral system and the ∂E becomes the corresponding energy change ($E_{N-1} - E_N$ or $E_N - E_{N+1}$). This way, the differential equation can be transformed into a set of algebraic equations. The self-consistent field (SCF) energy of the neutral system (with N electrons) and total single point energies of the cationic (E_{N-1}) and anionic (E_{N+1}) systems are calculated to obtain the ionization potential (I) and electron affinity (A) of the system as follows:

$$I = E_{N-1} - E_N \quad (4)$$

$$A = E_N - E_{N+1} \quad (5)$$

Thus, the hardness, chemical potential, and electronegativity are approximated as

$$\eta \approx I - A \quad (6)$$

$$\mu = -\chi \approx -\frac{1}{2}(I + A) \quad (7)$$

Alternatively, to avoid high computational cost, we sometimes employ Koopmans' theorem [74] to calculate I and A as follows (E_{HOMO} and E_{LUMO} are the energies of the highest occupied and lowest unoccupied orbitals, respectively)

$$I = -E_{\text{HOMO}} \quad (8)$$

$$A = -E_{\text{LUMO}} \quad (9)$$

Equations (6) and (7) then transform to the following

$$\eta \approx E_{\text{LUMO}} - E_{\text{HOMO}} \quad (10)$$

$$\mu = -\chi \approx \frac{1}{2}(E_{\text{HOMO}} + E_{\text{LUMO}}) \quad (11)$$

The electrophilicity index is then evaluated using the values of μ and η following Equation (3). Although the global reactivity indices are fundamentally related to the energies of HOMO and LUMO energies, they help bring out different aspects of the electronic structure of the systems [75,76].

3. Methodology

First and foremost, energy minimization by way of geometry optimization is performed on all the compounds present in the dataset, viz., a set of 15 benzene derivatives for studying the toxicity against *Pimephales promelas* (Figure 1) [56], datasets of polychlorinated dibenzofurans (PCDFs) (Table S1) and polychlorinated biphenyls (PCBs) (Table S2) against radio-labeled tetrachlorodibenzo-p-dioxin (TCDD) [77], a set of 252 aliphatic compounds (comprising alcohols, esters, acids, aldehydes, ketones, and amines; Table S3) against *Tetrahymena pyriformis* [78], and a set of 32 pyridyl benzamides (Table S4) against

Trypanosoma brucei [79]. The levels of theory used for computation of the above datasets are B3LYP/6-31G(d) (for the first two datasets), HF/6-311G**, and HF/6-31G(d) for the third and fourth datasets, respectively. The geometry optimization is followed by a frequency calculation on the compounds at the respective levels of theory to ensure the absence of any imaginary frequency. These computations are performed using the Gaussian 16 program package [80]. Relevant CDFT descriptors are then evaluated with the help of the equations provided in the “Theoretical background” section. We have then employed the multiple linear regression (MLR) and neural network (NN) methods to construct statistically relevant, robust QSAR models for predicting toxicological/biological activities of several datasets.

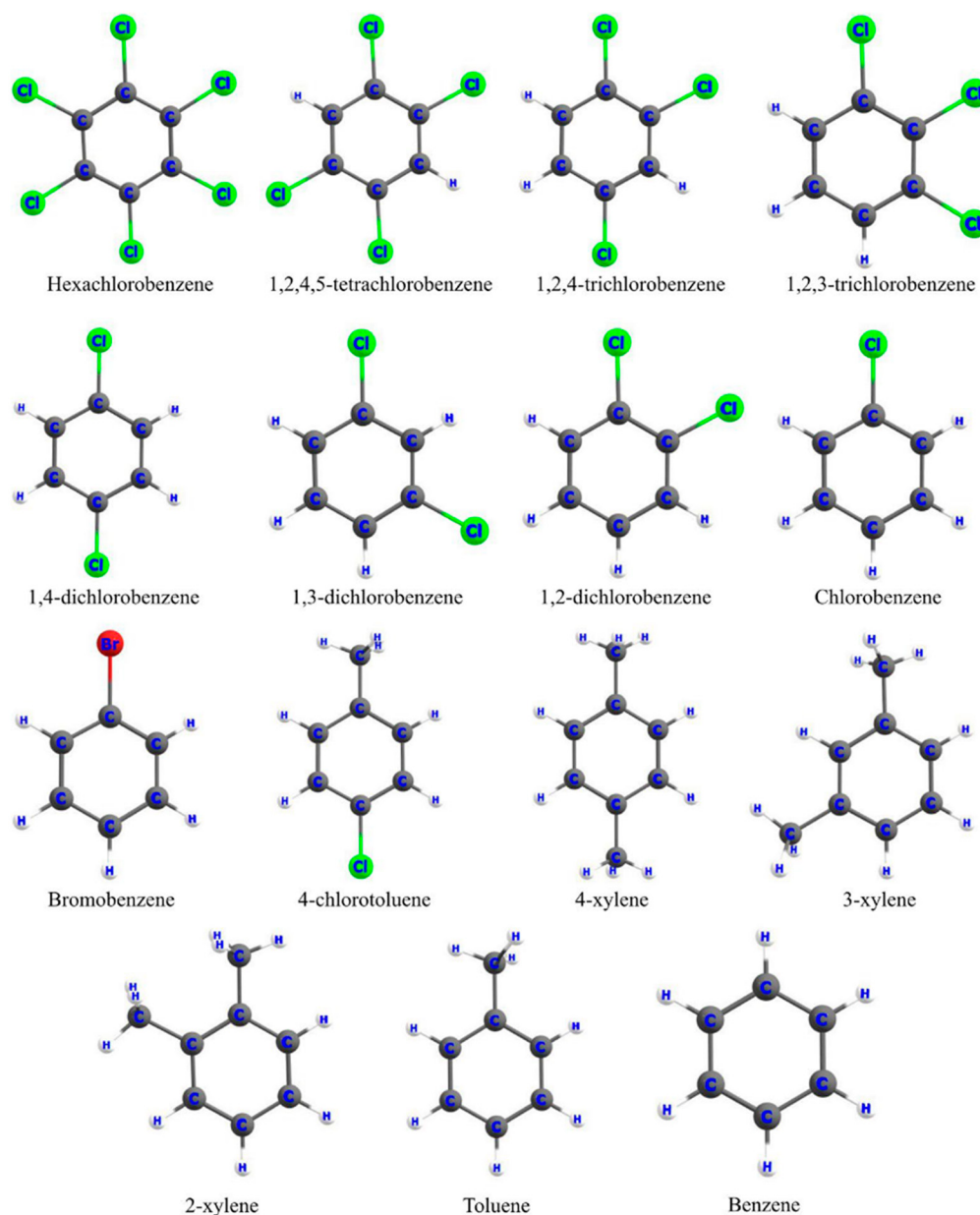


Figure 1. Geometries of the substituted benzene derivatives optimized at the B3LYP/6-31G(d) level of theory (reprinted from ref. [56] with permission from John Wiley & Sons. © 2018 John Wiley & Sons A/S).

The general formula of any QSAR model is as follows:

$$\text{Activity/Property/Toxicity} = f(\text{physicochemical properties})$$

$$= a_0 + a_1x_1 + a_2x_2 + a_3x_3 + \dots \quad (12)$$

where x_n are the descriptors and a_n are their corresponding coefficients.

In our cases, the mathematical descriptors for toxicity are pLC₅₀, pIC₅₀, and pIGC₅₀ acting as the dependent variables, whereas the global and local electrophilicity indices and hydrophobicity are the independent variables.

The efficacy of the constructed model is determined by some statistical parameters, viz., coefficient of determination (R^2), adjusted R^2 (R^2_{adj}), standard deviation (SD)

$$R^2 = 1 - \frac{\sum(Y_{obs} - Y_{calc})^2}{\sum(Y_{obs} - \bar{Y}_{obs})^2} \quad (13)$$

$$R^2_{adj} = \frac{(N - 1) * R^2 - p}{N - 1 - p} \quad (14)$$

$$SD = \sqrt{\frac{\sum(Y_{obs} - Y_{calc})^2}{N - 1 - p}} \quad (15)$$

where Y_{obs} and Y_{calc} are the experimental and predicted dependent variables, respectively (in our case, toxicities), N is the total number of observables, and p denotes the number of descriptors used in the model.

3.1. Multiple Linear Regression (MLR)

The simple yet most widely accepted method, multiple linear regression (MLR) [81–84], uses regression coefficient or the coefficient of determination (R^2) and standard deviation (SD) as its statistical metrics to judge the efficacy of the generated QSAR model. An initial descriptor selection based on the knowledge of the reaction mechanism is followed by a mathematical screening where possible combinations of the selected descriptors are tried on the whole dataset. The relevance of each descriptor to the target activity is revealed in terms of the magnitude and sign of the respective coefficients. Models with higher R^2 and lower SD values are selected for further use, while the rest are rejected. The selected combinations of descriptors are then employed on a training–test split of the dataset. The model is trained on the training set and is used to predict the activity of the compounds in the test set. Sometimes a validation set is also included. The regression model is generated on the training set with the experimental toxicity (pLC₅₀ or pIGC₅₀) or bioactivity as the dependent and the computed descriptors as the independent variables, followed by utilizing it to predict the activity of the test set compounds.

The above approach of splitting the data has the drawback that its SD may vary depending on which compounds are placed in the training and test sets. To remove any such bias, the model needs to be fitted several times with different combinations of training–test pairs each time, followed by evaluating the statistical metrics of the test set and averaging them out. This technique is known as cross-validation. Threefold cross-validation is employed by splitting the dataset into three equi-sized groups (sets A, B, and C), where two of them form the training set while the other forms the test set. Three such combinations are obtained. QSAR models are generated on all of them to obtain their respective R^2 and SD values. Another specific case of cross-validation, known as the leave-one-out (LOO) cross-validation, is an exhaustive method that includes all possible combinations within the dataset. In this method, the dataset (of, say, N number of compounds) is split into a training set with $N - 1$ number of compounds, leaving only one compound for testing the trained model. This is repeated N number of times and the desired metric is averaged. However, being a very exhaustive process, it tends to be rather time-consuming and computationally expensive.

3.2. Neural Networks (NNs)

Apart from MLR, we have also utilized a supervised machine learning technique [85–93], the multilayer perceptron (MLP) neural network [53,54,94–105] for toxicological predictions by generating QSAR models with known values of hydrophobic ($\log P$) and electronic (ω) parameters. An MLP comprises input, hidden, and output layers, where the units in the hidden layer are connected to those of the input layer by certain weights. The units in the output layer are connected in a similar fashion to those of the hidden layer. The hidden layer consists of a non-linear (most commonly, sigmoid) transfer function, $y = 1/(1 + e^{-x})$, where x is the total weighted input for the unit. Such transfer functions allow the hidden layer to mimic bio-neurons, hence earning it the name “artificial neurons”. The output layer units also consist of transfer functions which vary depending on the desired application of the MLP, e.g., while sigmoid functions are preferred for classification problems, linear functions are better suited for regression problems to predict real-valued quantities (like pIGC_{50}).

In our studies, we have initialized all the weights with random values. The training data of descriptor-activity pairs are supplied as input, and the output error is calculated for each input as the squared difference of the calculated output from the real output. This error value is then back-propagated to update the weights between hidden-output and input-hidden layers. The process is repeated until convergence is reached, where the final set of weights is the trained set of parameters for the MLP which is then ready to predict the target activity of the unknown molecules in the test set. This step is done by supplying the descriptor values which get multiplied by the weights and sent to the hidden layer, where the sigmoid function acts on them to predict the target activity.

4. Case Studies

4.1. *Pimephales Promelas*

The toxicity of a set of 15 benzene derivatives (optimized structures are shown in Figure 1) towards *Pimephales promelas* (fathead minnow) is studied by developing SAR models with pLC_{50} as the dependent variable and the hydrophobicity and electrophilicity indices as the descriptors [56]. Initially, a training–test split of 10 and 5 is considered, followed by the threefold cross-validation where each subdivision of the dataset (i.e., A, B, and C) has five molecules each. The training was done taking either of the two sets, and then the test was performed with the third set.

The prediction ability of the regression models generated is judged in terms of their R^2 , adjusted- R^2 , and SD values. Several combinations of the electrophilicity (ω , ω^2 , and ω^3) and hydrophobicity ($\log P$ and $(\log P)^2$) are taken as the independent variables. The regression models obtained by applying the three-fold cross-validation technique reveal that ω^2 (with an average R^2 of 0.890) shows a higher correlation compared to that of ω (average $R^2 = 0.864$) or ω^3 (average $R^2 = 0.882$). The results obtained considering $\log P$ and $(\log P)^2$ are comparable to those obtained by using the electrophilicity indices. Considering the fact that ω can be obtained much more easily through computation than obtaining the experimental values of $\log P$, the former is more convenient, cheaper, and faster to use than the latter. The plot showing a good correlation between the experimental and predicted pLC_{50} values for different cases is provided in Figure 2. Analysis through the MLP neural network validates the results obtained from employing the simple MLR technique.

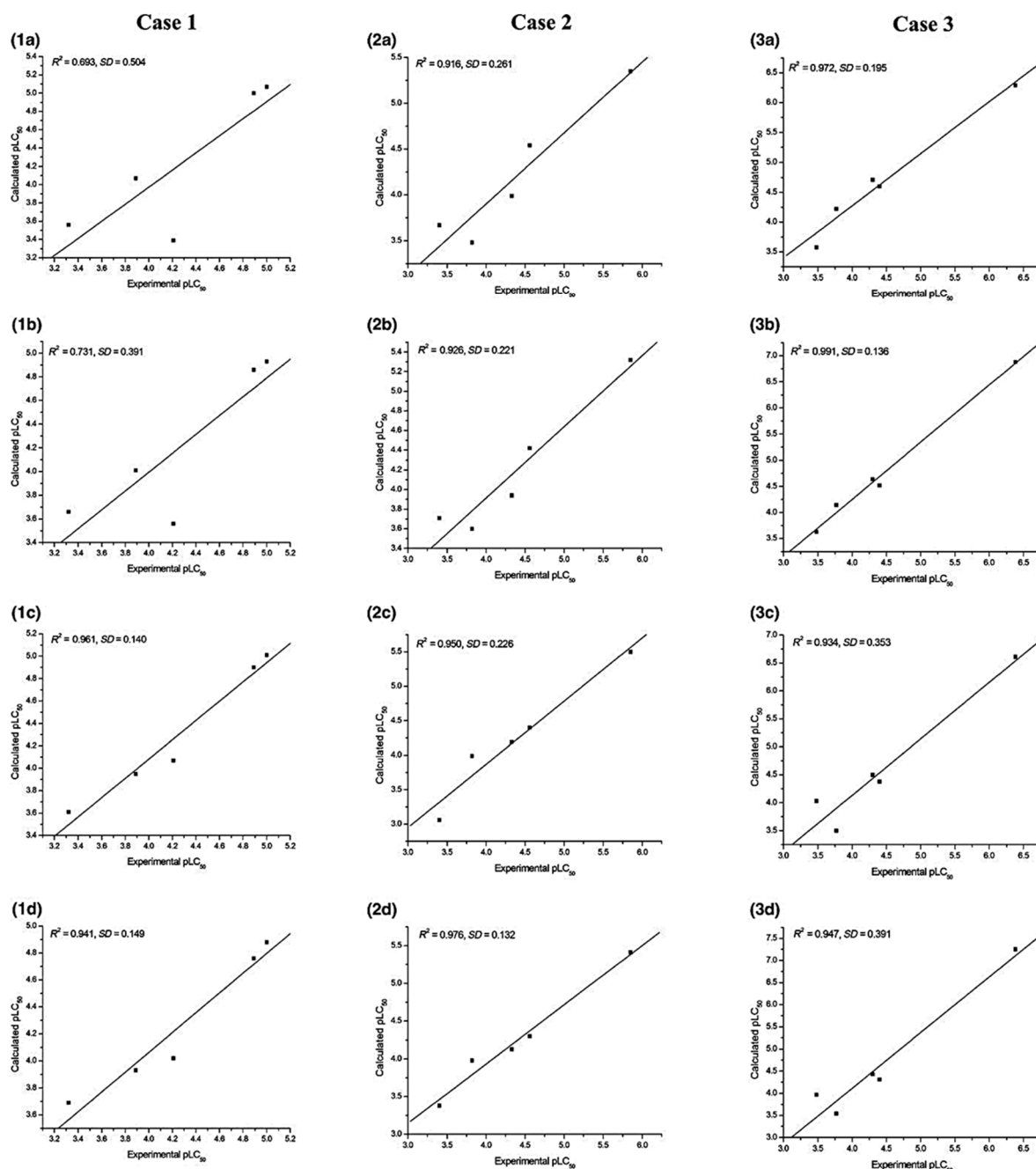


Figure 2. Experimental vs. calculated pLC_{50} plots with respect to (a) ω , (b) ω^2 , (c) $\log P$, and (d) $(\log P)^2$, for the test set with regression models developed using the training set in all three cases of the three-fold CV (reprinted from ref. [56] with permission from John Wiley & Sons. © 2018 John Wiley & Sons A/S).

4.2. *Tetrahymena Pyriformis*

A total of 252 aliphatic compounds grouped into six different categories based on their primary functional group, viz., alcohols, esters, acids, aldehydes, ketones, and amines, are selected for studying their toxicity ($\log(IGC_{50}^{-1})$) against the ciliate *Tetrahymena pyriformis* [78]. CDFT-based electronic descriptors, electrophilicity, and local philicity are considered as the independent variables for the QSTR modeling for toxicity prediction. The study is performed for each group separately, and for the whole dataset as well. Each of the six aforementioned groups (and their respective subgroups) is designated as electron accepting or donating by comparing their electronegativity values with those of the nucleic acid bases/selected DNA base pairs. Depending on whether the compound has an electron

accepting or donating tendency, the local philicity parameter, ω_m^+ or ω_m^- , respectively, is chosen as its descriptor of choice. To judge the model efficacy, the statistical parameter, viz., coefficient of determination (R^2), is calculated. Figure 3 depicts the correlation between the experimental and calculated toxicity values of all the six aforementioned groups of compounds, along with those for the electron accepting and donating groups. R^2 obtained in these cases are 0.831 (109 aliphatic alcohols), 0.787 (39 aliphatic acids), 0.766 (51 aliphatic esters), 0.803 (13 aliphatic aldehydes), 0.778 (15 ketones), and 0.791 (25 aliphatic amines). Similar calculations performed for all the electron acceptor compounds (171), irrespective of their functional group, delivers an R^2 of 0.801. The same for all the electron donor compounds (81) is 0.870. From the correlations obtained, it is very clear that the global electrophilicity along with local philicity makes up a good pair of descriptors in explaining the toxicity of these aliphatic compounds against *T. pyriformis*, especially when previous studies [106–109] have reported SAR models with a higher number of predictors and/or with poor correlations. The toxicity of the entire dataset is analyzed in terms of the number of carbon atoms present in the respective compounds [110].

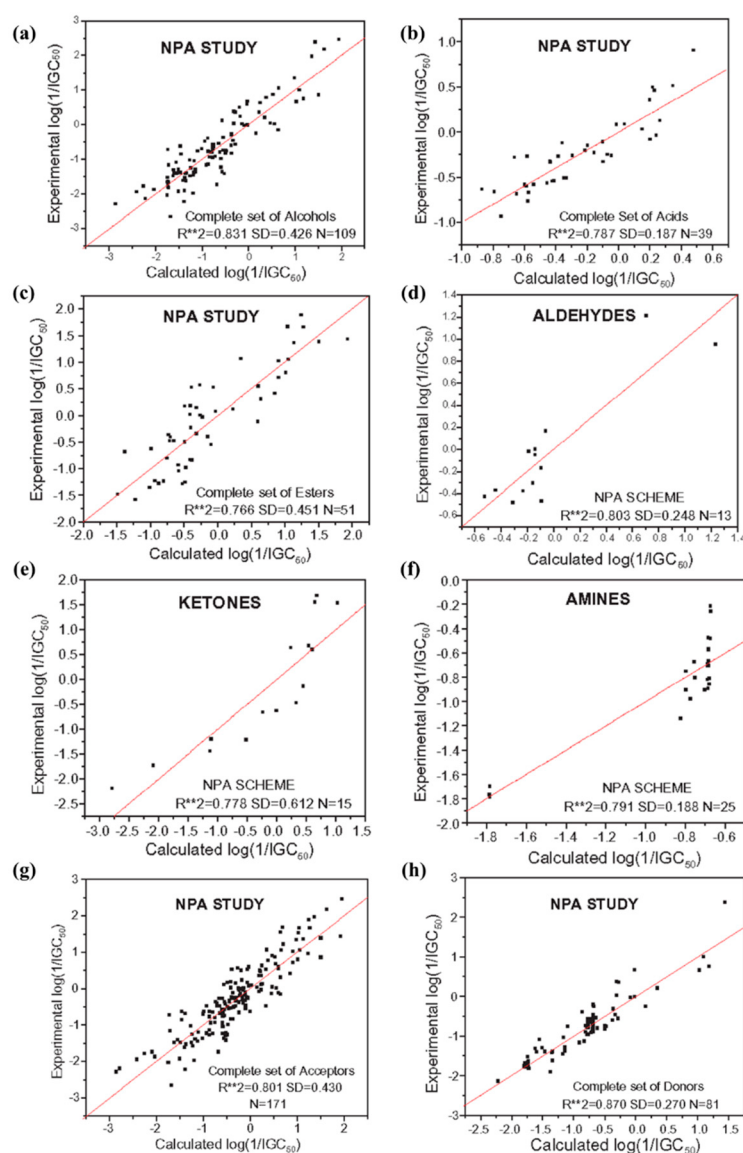


Figure 3. Correlation between the experimental and calculated $\log((IGC_{50})^{-1})$ of (a) alcohols, (b) esters, (c) acids, (d) aldehydes, (e) ketones, (f) amines, (g) all electron accepting compounds, and (h) all electron donating compounds (assorted from Ref. [78] Copyright © 2005 Elsevier Ltd.).

The global and local electrophilicity indices are also employed [77] on a dataset of polyaromatic hydrocarbons (PAH) to analyze the extent of their toxic effects towards displacing half of the radio-labeled tetrachlorodibenzo-*p*-dioxin (TCDD) from an arylhydrocarbon (Ah) receptor. For a dataset of electron acceptor [39,46] toxins, viz., polychlorinated dibenzofurans (PCDF), the model is trained using experimental pIC_{50} values as dependent and ω as the independent variables, respectively. The correlations obtained while using ω as the sole descriptor for the PCDFs, in the training–test split and the whole dataset, are $R = 0.891$ (for training; 0.834 for the test), and 0.891 (with an R^2 of 0.786), respectively. An F -ratio of 96.743 is obtained, which is a good number for the given number of data points and compared to other parameters. Ponc's [111] method is employed to calculate the probability of any random distribution producing the same R^2 , which turned out to be 10^{-6} , ruling out the possibility of any chance correlation. The dataset of PCBs is then used as an external validation set to test the efficacy of the predictive ability of the QSAR model constructed for the training set of the PCDFs. A correlation of 0.834 is obtained for the same, suggesting the relevance of ω as the sole descriptor in such electrophile–nucleophile mechanism-driven processes (see Figure 4).

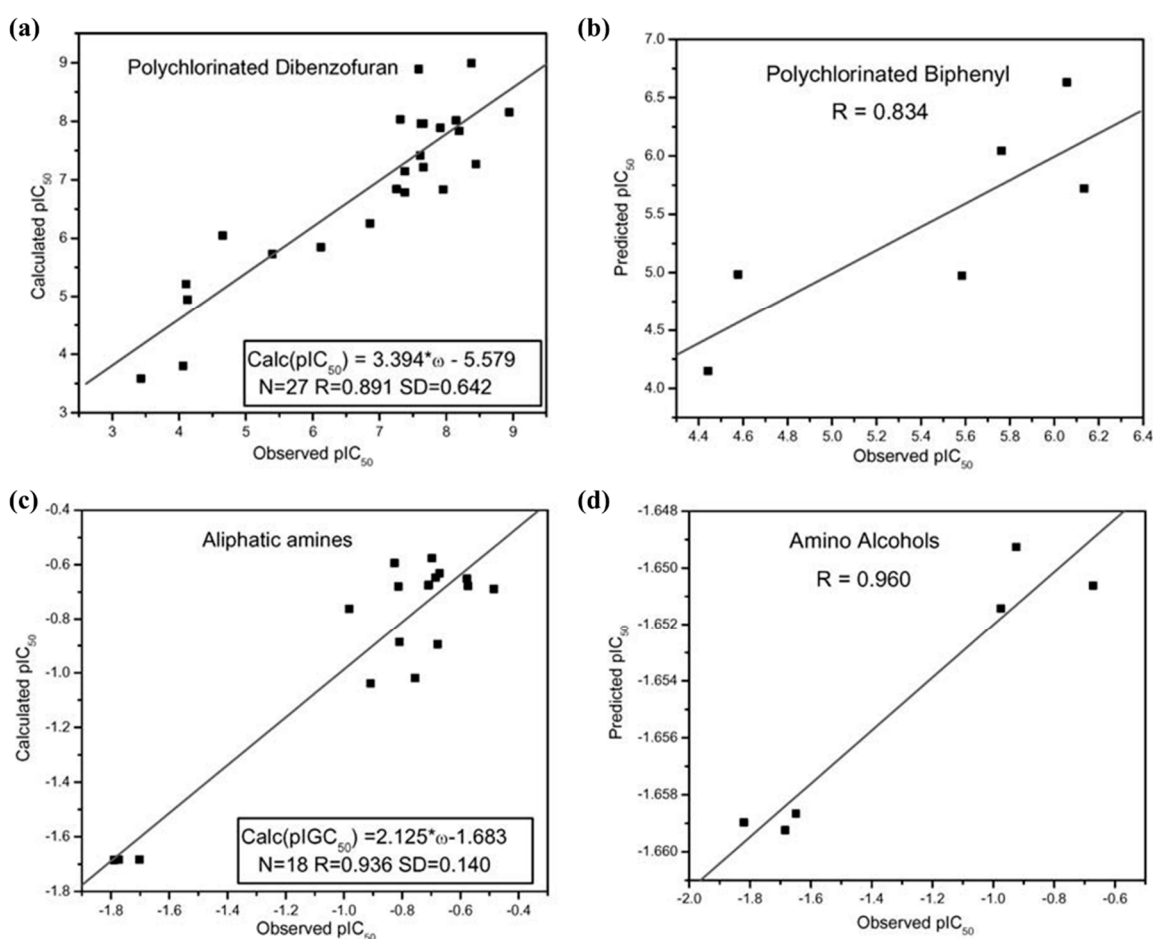


Figure 4. Plots of calculated vs. observed pIC_{50} for (a) training set of PCDFs, (b) test set of PCBs, (c) training set of aliphatic amines, and (d) test set of amino alcohols (adapted from Ref. [77] with permission from Springer Nature. © 2006, Springer Science Business Media, Inc.).

In the presence of biomolecules, while PCDFs and PCBs exhibit an electron-accepting nature [39,52,112,113], aliphatic amines behave as electron donors. The electron accepting/donating nature of these compounds is determined by analyzing the transfer of a fractional number of electrons (ΔN) from these compounds to the DNA base pairs/nucleic acid bases. A positive and negative ΔN represents the compounds as electron acceptors and

donors, respectively. The toxicity of these amines towards *T. pyriformis* is also studied. They interact with the biomolecules through the donation of electrons [78]. Since there exists no reliable measure of global nucleophilicity, the maximum value of the local variant ($\omega_{\max}^-$) at the N atom of the amines is considered as the independent variable of the regression model. A correlation (R) of 0.936 was obtained with the NPA-derived $\omega_{\max}^-$ values for the QSAR equation:

$$\text{pIC}_{50} = 2.2137 (0.2076) * \omega_{N_{\max}}^- - 1.6895 (+0.0822) \quad (16)$$

$$N = 18, SD = 0.1493, R = 0.9363, R_{adj}^2 = 0.8689$$

A high F -ratio of 113.7066 is obtained, and a high Q^2 value (leaving five points) of 0.8243 is obtained. The possibility of a chance correlation is ruled out by a probability value of 0.18×10^{-5} . Further, models are also generated for the combined datasets of amines and amino alcohols, followed by a training–test split. The predictive ability of global and local electrophilicity indices is demonstrated for both electron acceptors and donors in both the gas and solution phases.

Another study [114] on *T. pyriformis* includes the performance of an extensive MLP neural network and an MLR study using electrophilicity indices in conjunction with the hydrophobicity parameters of 169 aliphatic compounds. The correlation coefficient of the models based on the hydrophobicity and electrophilicity parameters lies within 0.790–0.983 and 0.703–0.779. Among all possible descriptor combinations, the best QSTR models turned out to be the ones with $\{\omega, \log P\}$, $\{\omega^2, \log P\}$ and $\{\log P, (\log P)^2\}$ pairs of descriptors.

4.3. *Trypanosoma Brucei*

A total of 32 numbers of pyridyl benzamide derivatives are considered for the study against the sleeping sickness, Human African trypanosomiasis (HAT) causing parasite *Trypanosoma brucei* [79]. All possible combinations of ω , ω^2 and other descriptors utilized by Masand and coworkers [115] (provided in Table 1) are utilized to generate relevant regression models through MLR. The R^2 and SD values obtained for the undivided dataset and the three combinations of the training–test split in the three-fold CV are provided in Table 2, which indicates that the 1st, 3rd, 4th, 6th, 7th, and 10th models show good correlation, while for the rest of the models the R^2 does not cross the threshold value of 0.6 and is thus statistically irrelevant. An important observation is noted that on the removal of the RDF55s descriptor, there is a drastic reduction in the model efficiency, which clearly suggests its importance in this study.

Table 1. List of descriptors obtained from Ref. [115] used along with the electrophilicity index to develop QSTR models describing the HAT activity of pyridyl benzamide derivatives.

GATS8c	Geary autocorrelation of lag-8/weighted by atomic charges
RDF40p	Radial distribution function-040/weighted by relative polarizabilities
RDF55s	Radial distribution function-055/weighted by relative I-state
E1	1st component accessibility directional WHIM index/weighted by relative I-state
RDF40m	Radial distribution function-040/weighted by relative mass

Table 2. R^2 and SD values obtained for the undivided dataset and the three combinations of training–test split in the three-fold cross-validation (reprinted from Ref. [79]. © 2019, IGI Global).

Model No.	Generalized Regression Equations	Undivided		Case 1		Case 2		Case 3	
		R^2	SD	R^2	SD	R^2	SD	R^2	SD
1	$pIC_{50} = a + b \cdot GATS8c + c \cdot RDF40p + d \cdot RDF55s$	0.8284	0.1960	0.9182	0.1320	0.8216	0.1877	0.6971	0.2213
2	$pIC_{50} = a + b \cdot GATS8c + c \cdot RDF40p + d \cdot \omega$	0.3743	0.3742	0.4839	0.1872	0.3764	0.1546	0.2351	0.4378
3	$pIC_{50} = a + b \cdot GATS8c + c \cdot \omega + d \cdot RDF55s$	0.7599	0.2318	0.8325	0.1768	0.7829	0.2063	0.5114	0.2415
4	$pIC_{50} = a + b \cdot \omega + c \cdot RDF40p + d \cdot RDF55s$	0.7113	0.2542	0.7644	0.1983	0.7098	0.2283	0.4961	0.2529
5	$pIC_{50} = a + b \cdot GATS8c + c \cdot RDF40p + d \cdot \omega^2$	0.3650	0.3770	0.4825	0.1827	0.3901	0.1447	0.2211	0.4098
6	$pIC_{50} = a + b \cdot GATS8c + c \cdot \omega^2 + d \cdot RDF55s$	0.7592	0.2322	0.8323	0.1762	0.7853	0.2058	0.5817	0.2299
7	$pIC_{50} = a + b \cdot \omega^2 + c \cdot RDF40p + d \cdot RDF55s$	0.7068	0.2562	0.7620	0.1977	0.7101	0.2301	0.4645	0.2575
8	$pIC_{50} = a + b \cdot GATS8c + c \cdot \omega + d \cdot \omega^2$	0.3285	0.3877	0.4334	0.1888	0.2506	0.1882	0.1725	0.4421
9	$pIC_{50} = a + b \cdot \omega + c \cdot RDF40p + d \cdot \omega^2$	0.3660	0.3767	0.4746	0.1781	0.2911	0.1800	0.1810	0.4594
10	$pIC_{50} = a + b \cdot \omega + c \cdot \omega^2 + d \cdot RDF55s$	0.6836	0.2661	0.7637	0.2014	0.6962	0.2212	0.5163	0.2364
11	$pIC_{50} = a + b \cdot E1s + c \cdot RDF40m + d \cdot GATS6m$	0.3056	0.3942	0.3991	0.1983	0.3199	0.2404	0.1159	0.2055
12	$pIC_{50} = a + b \cdot E1s + c \cdot RDF40m + d \cdot \omega$	0.3647	0.3771	0.4949	0.1839	0.3196	0.1784	0.2793	0.3939
13	$pIC_{50} = a + b \cdot E1s + c \cdot \omega + d \cdot GATS6m$	0.4847	0.3396	0.5936	0.2275	0.5415	0.2138	0.3369	0.3171
14	$pIC_{50} = a + b \cdot \omega + c \cdot RDF40m + d \cdot GATS6m$	0.4758	0.3425	0.5824	0.2177	0.5019	0.2127	0.3263	0.3134
15	$pIC_{50} = a + b \cdot E1s + c \cdot RDF40m + d \cdot \omega^2$	0.3571	0.3793	0.4997	0.1769	0.3241	0.1706	0.2135	0.4273
16	$pIC_{50} = a + b \cdot E1s + c \cdot \omega^2 + d \cdot GATS6m$	0.4763	0.3424	0.5839	0.2266	0.5473	0.2083	0.2648	0.3481
17	$pIC_{50} = a + b \cdot \omega^2 + c \cdot RDF40m + d \cdot GATS6m$	0.4666	0.3455	0.5708	0.2181	0.5105	0.2030	0.2518	0.3441
18	$pIC_{50} = a + b \cdot E1s + c \cdot \omega + d \cdot \omega^2$	0.3421	0.3421	0.4673	0.1865	0.2898	0.1777	0.1339	0.4681
19	$pIC_{50} = a + b \cdot \omega + c \cdot \omega^2 + d \cdot GATS6m$	0.4784	0.3417	0.5861	0.2301	0.4044	0.2680	0.1784	0.3648
20	$pIC_{50} = a + b \cdot \omega + c \cdot RDF40m + d \cdot \omega^2$	0.3583	0.3790	0.4822	0.1844	0.2185	0.2091	0.1454	0.4405
21	$pIC_{50} = a + b \cdot \omega + c \cdot \omega^2$	0.3272	0.3813	0.4540	0.1813	0.2922	0.1765	0.1288	0.4612

A technique for comparing the developed regression models, known as the sum of ranking differences (SRDs) [116,117], is carried out where the models are ranked in order of their efficacy. It requires the data to be arranged in a matrix structure with the rows representing the statistical metrics (R^2 and SD), and the columns are the models to be ranked. An ideal or golden standard is chosen (here they are the highest R^2 and least SD values) whose difference from the R^2 and SD values of each of the models generates the SRD values. These SRD values are then compared to judge the models' efficacy. In this ranking method, the lower the SRD value, the better is the model. Figure 5 represents the ranking of the models, showcasing their relative position and extent of the similarity.

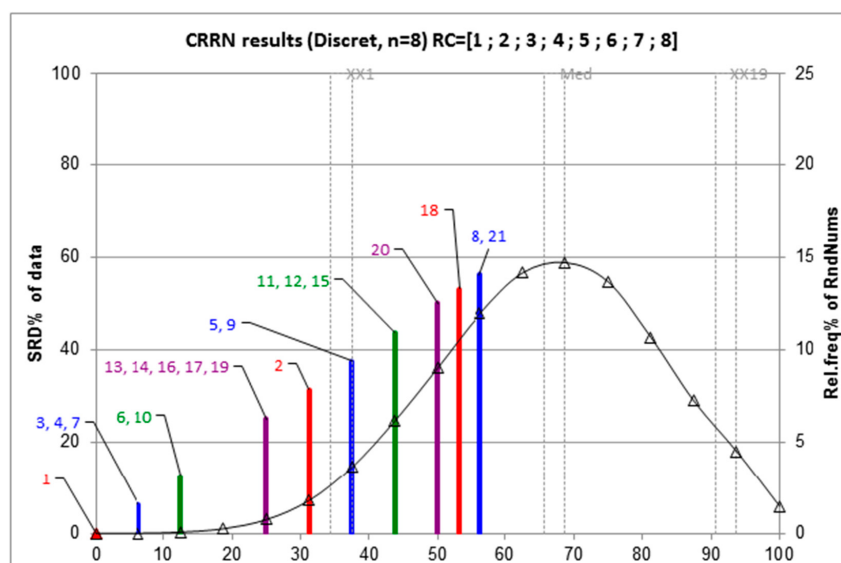


Figure 5. Descriptors choice evaluation using SRD with ties and a scaling between 0–100. The different colored lines represent ranking of different regression models, showcasing their relative position and extent of the similarity. The respective model numbers are provided as per Table 2. (reprinted from Ref. [79]. © 2019, IGI Global).

Several other studies on QSAR modeling are reported on the effective use of ω as the electronic factor. The toxicity of arsenic ions is predicted using regression models based on atomic charge along with global and local electrophilicity indices. The model was trained on datasets of alkali and transition-metal ions [118]. The bio-activity of the derivatives of the sex hormones, testosterone and estrogen, are also successfully reported in terms of ω [40]. Several other quantum chemical descriptors, viz., IP, EA, η , softness (S), χ , along with ω are employed for studying the correlations in an alkane series, where it is noted that the IP performs best in describing various macroscopic properties [119].

5. Conclusions

The present article focuses on an analysis relating to the simplest yet effective regression techniques in the field of QSAR. Specifically, it delivers strong evidence for the effectiveness of both global and local electrophilicities in predicting toxicity and bioactivity. The ease of their computation compared to other common descriptors comes as an added bonus for their extensive usage. Employing them, either as solo descriptors in single-parameter models, or with other descriptors like the number of carbon atoms, charge transfer, hydrophobicity, etc., in multi-parameter regression models, has resulted in high coefficients of determination and low standard deviation. Needless to say, however, prior knowledge of the reaction mechanism is of the utmost importance in order to select appropriate descriptors for the target activity.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ph15111383/s1>, Table S1: Polychlorinated dibenzofurans with identity number (ID) representing the substitution pattern. (reprinted from ref. [77] with permission from Springer Nature. Copyright © 2006, Springer Science Business Media, Inc.); Table S2: Polychlorinated biphenyls with identity number (ID) representing the substitution pattern. (reprinted from ref. [77] with permission from Springer Nature. Copyright © 2006, Springer Science Business Media, Inc.). Table S3: Dataset of 252 aliphatic compounds considered against *Tetrahymena pyriformis* [78]. Table S4: The dataset of 32 pyridyl benzamides considered against *Trypanosoma brucei* (reprinted from Ref. [79]. © 2019, IGI Global).

Author Contributions: Conceptualization, P.K.C., R.P. and S.G.P.; methodology, R.P. and S.G.P.; software, R.P.; validation, P.K.C., R.P. and S.G.P.; formal analysis, R.P. and S.G.P.; investigation, R.P. and S.G.P.; resources, P.K.C.; data curation, R.P. and S.G.P.; writing—original draft preparation, R.P. and S.G.P.; writing—review and editing, P.K.C., R.P. and S.G.P.; visualization, P.K.C.; supervision, P.K.C.; project administration, P.K.C.; funding acquisition, P.K.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Department of Science and Technology (DST), New Delhi, grant number SR/S2/JCB-09/2009.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article and supplementary material.

Acknowledgments: P.K.C. thanks Chia Ming Chang for kindly inviting him to contribute a review article to the Special Issue of Pharmaceuticals on “Applications of Conceptual Density Functional Theory to the Chemistry and Discovery of Bioactive Compounds”. R.P. and S.G.P. thank CSIR and IIT Kharagpur, respectively, for their fellowships.

Conflicts of Interest: The authors declare that they have no conflict of interest regarding the publication of this article, financial, and/or otherwise.

References

1. Karcher, W.; Devillers, J. SAR and QSAR in environmental chemistry and toxicology: Scientific tool or wishful thinking? In *Practical Applications of Quantitative Structure–Activity Relationships (QSAR) in Environmental Chemistry and Toxicology*; Karcher, W., Devillers, J., Eds.; Kluwer Academic: Dordrecht, The Netherlands, 1990; pp. 1–12.
2. Selassie, C.; Mekapati, S.; Verma, R. QSAR: Then and Now. *Curr. Top. Med. Chem.* **2002**, *2*, 1357–1379. [[CrossRef](#)]

3. Roy, K.; Mitra, I. Advances in quantitative structure–activity relationship models of antioxidants. *Expert Opin. Drug Discov.* **2009**, *4*, 1157–1175. [\[CrossRef\]](#)
4. Schultz, T.W.; Cronin, M.T.D.; Walker, J.D.; Aptula, A.O. Quantitative structure–activity relationships (QSARs) in toxicology: A historical perspective. *J. Mol. Struct. THEOCHEM* **2003**, *622*, 1–22. [\[CrossRef\]](#)
5. Schultz, T.W.; Cronin, M.T.D.; Netzeva, T.I. The present status of QSAR in toxicology. *J. Mol. Struct. THEOCHEM* **2003**, *622*, 23–38. [\[CrossRef\]](#)
6. Gombar, V.K.; Mattioni, B.E.; Zwickl, C.; Deahl, J.T. Computational Approaches for Assessment of Toxicity: A Historical Perspective and Current Status. In *Computational Toxicology*; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2006; pp. 183–215.
7. Hansch, C.; Maloney, P.P.; Fujita, T.; Muir, R.M. Correlation of Biological Activity of Phenoxyacetic Acids with Hammett Substituent Constants and Partition Coefficients. *Nature* **1962**, *194*, 178–180. [\[CrossRef\]](#)
8. Hansch, C.; Hoekman, D.; Leo, A.; Weininger, D.; Selassie, C.D. Chem-bioinformatics: Comparative QSAR at the interface between chemistry and biology. *Chem. Rev.* **2002**, *102*, 783–812. [\[CrossRef\]](#)
9. Katritzky, A.R.; Kuanar, M.; Slavov, S.; Hall, C.D.; Karelson, M.; Kahn, I.; Dobchev, D.A. ChemInform Abstract: Quantitative Correlation of Physical and Chemical Properties with Chemical Structure: Utility for Prediction. *ChemInform* **2011**, *42*. [\[CrossRef\]](#)
10. Fujita, T.; Winkler, D.A. Understanding the Roles of the “two QSARs”. *J. Chem. Inf. Model.* **2016**, *56*, 269–274. [\[CrossRef\]](#)
11. Wang, L.; Ding, J.; Pan, L.; Cao, D.; Jiang, H.; Ding, X. Quantum chemical descriptors in quantitative structure–activity relationship models and their applications. *Chemom. Intell. Lab. Syst.* **2021**, *217*, 104384. [\[CrossRef\]](#)
12. Tropsha, A. Best practices for QSAR model development, validation, and exploitation. *Mol. Inform.* **2010**, *29*, 476–488. [\[CrossRef\]](#)
13. Verina, J.; Malde, A.; Khedkar, S.; Iyer, R.; Coutinho, E. Local indices for similarity analysis (LISA)-A 3D-QSAR formalism based on local molecular similarity. *J. Chem. Inf. Model.* **2009**, *49*, 2695–2707. [\[CrossRef\]](#)
14. Verma, R.P.; Hansch, C. Use of ¹³C NMR chemical shift as QSAR/QSPR descriptor. *Chem. Rev.* **2011**, *111*, 2865–2899. [\[CrossRef\]](#)
15. McFarland, J.W. On the parabolic relationship between drug potency and hydrophobicity. *J. Med. Chem.* **1970**, *13*, 1192–1196. [\[CrossRef\]](#)
16. Itskowitz, P.; Tropsha, A. K nearest neighbors QSAR modeling as a variational problem: Theory and applications. *J. Chem. Inf. Model.* **2005**, *45*, 777–785. [\[CrossRef\]](#)
17. Ruggiu, F.; Gizzi, P.; Galzi, J.L.; Hibert, M.; Haiech, J.; Baskin, I.; Horvath, D.; Marcou, G.; Varnek, A. Quantitative structure–property relationship modeling: A valuable support in high-throughput screening quality control. *Anal. Chem.* **2014**, *86*, 2510–2520. [\[CrossRef\]](#)
18. Nieto-Draghi, C.; Fayet, G.; Creton, B.; Rozanska, X.; Rotureau, P.; De Hemptinne, J.C.; Ungerer, P.; Rousseau, B.; Adamo, C. A General Guidebook for the Theoretical Prediction of Physicochemical Properties of Chemicals for Regulatory Purposes. *Chem. Rev.* **2015**, *115*, 13093–13164. [\[CrossRef\]](#)
19. Shahlaei, M.; Fassihi, A. QSAR analysis of some 1-(3,3-diphenylpropyl)-piperidiny amides and ureas as CCR5 inhibitors using genetic algorithm-least square support vector machine. *Med. Chem. Res.* **2013**, *22*, 4384–4400. [\[CrossRef\]](#)
20. Katritzky, A.R.; Maran, U.; Lobanov, V.S.; Karelson, M. Structurally Diverse Quantitative Structure–Property Relationship Correlations of Technologically Relevant Physical Properties. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1–18. [\[CrossRef\]](#)
21. Shahlaei, M.; Madadkar-Sobhani, A.; Fassihi, A.; Saghale, L.; Shamshirian, D.; Sakhi, H. Comparative quantitative structure–activity relationship study of some 1-aminocyclopentyl-3-carboxyamides as CCR2 inhibitors using stepwise MLR, FA-MLR, and GA-PLS. *Med. Chem. Res.* **2012**, *21*, 100–115. [\[CrossRef\]](#)
22. Simarro, P.P.; Franco, J.; Diarra, A.; Postigo, J.A.R.; Jannin, J. Update on field use of the available drugs for the chemotherapy of human African trypanosomiasis. *Parasitology* **2012**, *139*, 842–846. [\[CrossRef\]](#)
23. Vigneresse, J.L. Revisiting immiscibility through DFT chemical descriptors. *Theor. Chem. Acc.* **2020**, *139*, 1–15. [\[CrossRef\]](#)
24. Gupta, S.P. Quantitative Structure–Activity Relationship Studies on Na⁺, K⁺-ATPase Inhibitors. *Chem. Rev.* **2012**, *112*, 3171–3192. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Gholivand, K.; Ebrahimi Valmoozi, A.A.; Mahzouni, H.R.; Ghadimi, S.; Rahimi, R. Molecular docking and QSAR studies: Noncovalent interaction between acephate analogous and the receptor site of human acetylcholinesterase. *J. Agric. Food Chem.* **2013**, *61*, 6776–6785. [\[CrossRef\]](#) [\[PubMed\]](#)
26. Ahmed, A.; Sandler, S.I. Physicochemical properties of hazardous energetic compounds from molecular simulation. *J. Chem. Theory Comput.* **2013**, *9*, 2389–2397. [\[CrossRef\]](#) [\[PubMed\]](#)
27. Gupta, S.P. QSAR Studies on Hydroxamic Acids: A Fascinating Family of Chemicals with a Wide Spectrum of Activities. *Chem. Rev.* **2015**, *115*, 6427–6490. [\[CrossRef\]](#)
28. Hu, J.; Zhang, X.; Wang, Z. A review on progress in QSPR studies for surfactants. *Int. J. Mol. Sci.* **2010**, *11*, 1020–1047. [\[CrossRef\]](#)
29. Bo, W.; Chen, L.; Qin, D.; Geng, S.; Li, J.; Mei, H.; Li, B.; Liang, G. Application of quantitative structure–activity relationship to food-derived peptides: Methods, situations, challenges and prospects. *Trends Food Sci. Technol.* **2021**, *114*, 176–188. [\[CrossRef\]](#)
30. Belfield, S.J.; Enoch, S.J.; Firman, J.W.; Madden, J.C.; Schultz, T.W.; Cronin, M.T.D. Determination of “fitness-for-purpose” of quantitative structure–activity relationship (QSAR) models to predict (eco)-toxicological endpoints for regulatory use. *Regul. Toxicol. Pharmacol.* **2021**, *123*, 104956. [\[CrossRef\]](#)
31. Cherkasov, A.; Muratov, E.N.; Fourches, D.; Varnek, A.; Baskin, I.I.; Cronin, M.; Dearden, J.; Gramatica, P.; Martin, Y.C.; Todeschini, R.; et al. QSAR modeling: Where have you been? Where are you going to? *J. Med. Chem.* **2014**, *57*, 4977–5010. [\[CrossRef\]](#)

32. Bruce, C.L.; Melville, J.L.; Pickett, S.D.; Hirst, J.D. Contemporary QSAR Classifiers Compared. *J. Chem. Inf. Model.* **2007**, *47*, 219–227. [\[CrossRef\]](#)
33. Ferrins, L.; Gazdik, M.; Rahmani, R.; Varghese, S.; Sykes, M.L.; Jones, A.J.; Avery, V.M.; White, K.L.; Ryan, E.; Charman, S.A.; et al. Pyridyl benzamides as a novel class of potent inhibitors for the kinetoplastid *Trypanosoma brucei*. *J. Med. Chem.* **2014**, *57*, 6393–6402. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Escuder-Gilabert, L.; Sagrado, S.; Villanueva-Camañas, R.M.; Medina-Hernández, M.J. Quantitative Retention-Structure and Retention-Activity Relationship Studies of Local Anesthetics by Micellar Liquid Chromatography. *Anal. Chem.* **1998**, *70*, 28–34. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Guo, X.D.; Zhang, L.J.; Qian, Y. Systematic Multiscale Method for Studying the Structure–Performance Relationship of Drug-Delivery Systems. *Ind. Eng. Chem. Res.* **2012**, *51*, 4719–4730. [\[CrossRef\]](#)
36. Burden, F.R. Quantitative Structure-Activity Relationship Studies Using Gaussian Processes. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 830–835. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Khadikar, P.V.; Mather, K.C.; Singh, S.; Phadnis, A.; Shrivastava, A.; Mandaloi, M. Study on quantitative structure–toxicity relationships of benzene derivatives acting by narcosis. *Bioorg. Med. Chem.* **2002**, *10*, 1761–1766. [\[CrossRef\]](#)
38. Padmanabhan, J.; Parthasarathi, R.; Subramanian, V.; Chattaraj, P.K. Group Philicity and Electrophilicity as Possible Descriptors for Modeling Ecotoxicity Applied to Chlorophenols. *Chem. Res. Toxicol.* **2006**, *19*, 356–364. [\[CrossRef\]](#)
39. Parthasarathi, R.; Padmanabhan, J.; Subramanian, V.; Maiti, B.; Chattaraj, P.K. Chemical Reactivity Profiles of Two Selected Polychlorinated Biphenyls. *J. Phys. Chem. A* **2003**, *107*, 10346–10352. [\[CrossRef\]](#)
40. Parthasarathi, R.; Subramanian, V.; Roy, D.R.; Chattaraj, P.K. Electrophilicity index as a possible descriptor of biological activity. *Bioorg. Med. Chem.* **2004**, *12*, 5533–5543. [\[CrossRef\]](#)
41. Russom, C.L.; Bradbury, S.P.; Broderius, S.J.; Hammermeister, D.E.; Drummond, R.A. Predicting modes of toxic action from chemical structure: Acute toxicity in the fathead minnow (*Pimephales Promelas*). *Environ. Toxicol. Chem.* **1997**, *16*, 948–967. [\[CrossRef\]](#)
42. Zhao, Y.H.; Cronin, M.T.D.; Dearden, J.C. Quantitative Structure-Activity Relationships of Chemicals Acting by Non-polar Narcosis—Theoretical Considerations. *Quant. Struct. Relatsh.* **1998**, *17*, 131–138. [\[CrossRef\]](#)
43. Karelson, M.; Lobanov, V.S.; Katritzky, A.R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. *Chem. Rev.* **1996**, *96*, 1027–1044. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Chakraborty, A.; Giri, S.; Duley, S.; Anoop, A.; Bultinck, P.; Chattaraj, P.K. Aromaticity in all-metal annular systems: The counter-ion effect. *Phys. Chem. Chem. Phys.* **2011**, *13*, 14865–14878. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Chattaraj, P.K.; Roy, D.R. Update 1 of: Electrophilicity Index. *Chem. Rev.* **2007**, *107*, PR46–PR74. [\[CrossRef\]](#)
46. Chattaraj, P.K.; Sarkar, U.; Roy, D.R. Electrophilicity Index. *Chem. Rev.* **2006**, *106*, 2065–2091. [\[CrossRef\]](#)
47. Geerlings, P.; De Proft, F.; Langenaeker, W. Conceptual Density Functional Theory. *Chem. Rev.* **2003**, *103*, 1793–1874. [\[CrossRef\]](#)
48. Parr, R.G.; Yang, W. *Density-Functional Theory of Atoms and Molecules*; Oxford University Press: New York, NY, USA; Oxford, UK, 1989.
49. Geerlings, P.; Chamorro, E.; Chattaraj, P.K.; De Proft, F.; Gázquez, J.L.; Liu, S.; Morell, C.; Toro-Labbé, A.; Vela, A.; Ayers, P. Conceptual density functional theory: Status, prospects, issues. *Theor. Chem. Acc.* **2020**, *139*, 36. [\[CrossRef\]](#)
50. Parthasarathi, R.; Padmanabhan, J.; Elango, M.; Chitra, K.; Subramanian, V.; Chattaraj, P.K. p K a Prediction Using Group Philicity. *J. Phys. Chem. A* **2006**, *110*, 6540–6544. [\[CrossRef\]](#)
51. Parr, R.G.; Szentpály, L.V.; Liu, S. Electrophilicity Index. *J. Am. Chem. Soc.* **1999**, *121*, 1922–1924. [\[CrossRef\]](#)
52. Parthasarathi, R.; Padmanabhan, J.; Sarkar, U.; Maiti, B.; Subramanian, V.; Chattaraj, P.K. Toxicity analysis of benzidine through chemical reactivity and selectivity profiles: A DFT approach. *Internet Electron. J. Mol. Des.* **2003**, *2*, 798–813.
53. Bansal, S.; Dixit, R. Performance enhancement of the pattern recalling efficiency of Hopfield neural network using genetic algorithm for cursive handwritten character recognition. *Int. J. Appl. Pattern Recognit.* **2016**, *3*, 59. [\[CrossRef\]](#)
54. Gridach, M. Character-level neural network for biomedical named entity recognition. *J. Biomed. Inform.* **2017**, *70*, 85–91. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Hertz, J.; Krogh, A.; Palmer, R.G. *Introduction to the Theory of Neural Computation*; CRC Press: Boca Raton, FL, USA, 2018; Volume 44, ISBN 9780429968211.
56. Pal, R.; Jana, G.; Sural, S.; Chattaraj, P.K. Hydrophobicity versus electrophilicity: A new protocol toward quantitative structure–toxicity relationship. *Chem. Biol. Drug Des.* **2019**, *93*, 1083–1095. [\[CrossRef\]](#) [\[PubMed\]](#)
57. Bianucci, A.M.; Micheli, A.; Sperduti, A.; Starita, A. A Novel Approach to QSPR/QSAR Based on Neural Networks for Structures. In *Soft Computing Approaches in Chemistry*; Springer: Berlin/Heidelberg, Germany, 2003; pp. 265–296.
58. Hansch, C.; Fujita, T. p-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. *J. Am. Chem. Soc.* **1964**, *86*, 1616–1626. [\[CrossRef\]](#)
59. Kubinyi, H. Methods and Principles in Medicinal Chemistry. In *QSAR: Hansch Analysis and Related Approaches*; Wiley-VCH: Weinheim, Germany; New York, NY, USA, 1993; p. 438, ISBN 352730035X.
60. Pauling, L. *The Nature of the Chemical Bond*, 3rd ed.; Cornell University Press: Ithaca, NY, USA, 1960.
61. Sen, K.; Jorgenson, C. Electronegativity. In *Structure and Bonding*; Springer: Berlin/Heidelberg, Germany, 1987.
62. Pearson, R.G. *Chemical Hardness*; Wiley-VCH: Weinheim, Germany, 1997.
63. Chattaraj, P.K. (Ed.) *Chemical Reactivity Theory*; CRC Press: Boca Raton, FL, USA, 2009; ISBN 9780429137228.

64. Parr, R.G.; Donnelly, R.A.; Levy, M.; Palke, W.E. Electronegativity: The density functional viewpoint. *J. Chem. Phys.* **1977**, *68*, 3801–3807. [\[CrossRef\]](#)
65. Pauling, L. The nature of the chemical bond. IV. the energy of single bonds and the relative electronegativity of atoms. *J. Am. Chem. Soc.* **1932**, *54*, 3570–3582. [\[CrossRef\]](#)
66. Mulliken, R.S. A New Electroaffinity Scale; Together with Data on Valence States and on Valence Ionization Potentials and Electron Affinities. *J. Chem. Phys.* **1934**, *2*, 782–793. [\[CrossRef\]](#)
67. Allred, A.L.; Rochow, E.G. A scale of electronegativity based on electrostatic force. *J. Inorg. Nucl. Chem.* **1958**, *5*, 264–268. [\[CrossRef\]](#)
68. Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871. [\[CrossRef\]](#)
69. Kohn, W.; Sham, L.J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140*, A1133–A1138. [\[CrossRef\]](#)
70. Ingold, C.K. 266. Significance of tautomerism and of the reactions of aromatic compounds in the electronic theory of organic reactions. *J. Chem. Soc.* **1933**, 1120–1127. [\[CrossRef\]](#)
71. Ingold, C.K. Principles of an Electronic Theory of Organic Reactions. *Chem. Rev.* **1934**, *15*, 225–274. [\[CrossRef\]](#)
72. Maynard, A.T.; Huang, M.; Rice, W.G.; Covell, D.G. Reactivity of the HIV-1 nucleocapsid protein p7 zinc finger domains from the perspective of density-functional theory. *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 11578–11583. [\[CrossRef\]](#)
73. Atoms, D.T.; Parr, M.R.G.; Yang, W. Book Review. *Density Funct. Theory Atoms Mol.* **1989**, *47*, 10101.
74. Koopmans, T. Über die Zuordnung von Wellenfunktionen und Eigenwerten zu den Einzelnen Elektronen Eines Atoms. *Physica* **1934**, *1*, 104–113. [\[CrossRef\]](#)
75. Mačernis, M. Nonlinear correlations between ν 1 Raman band and global scalar properties for different length carotenoids. *Lith. J. Phys.* **2019**, *58*, 358–378. [\[CrossRef\]](#)
76. Flores-Holguín, N.; Frau, J.; Glossman-Mitnik, D. Conceptual DFT-Based Computational Peptidology of Marine Natural Compounds: Discodermins A–H. *Molecules* **2020**, *25*, 4158. [\[CrossRef\]](#)
77. Roy, D.R.; Sarkar, U.; Chattaraj, P.K.; Mitra, A.; Padmanabhan, J.; Parthasarathi, R.; Subramanian, V.; Van Damme, S.; Bultinck, P. Analyzing toxicity through electrophilicity. *Mol. Divers.* **2006**, *10*, 119–131. [\[CrossRef\]](#)
78. Roy, D.R.; Parthasarathi, R.; Maiti, B.; Subramanian, V.; Chattaraj, P.K. Electrophilicity as a possible descriptor for toxicity prediction. *Bioorg. Med. Chem.* **2005**, *13*, 3405–3412. [\[CrossRef\]](#)
79. Pal, R.; Pal, G.; Jana, G.; Chattaraj, P.K. An In Silico QSAR Model Study Using Electrophilicity as a Possible Descriptor Against T. Brucei. *Int. J. Chemoinform. Chem. Eng.* **2019**, *8*, 57–68. [\[CrossRef\]](#)
80. Frisch, M.J.; Trucks, G.W.; Schlegel, H.B.; Scuseria, G.E.; Robb, M.A.; Cheeseman, J.R.; Scalmani, G.; Barone, V.; Petersson, G.A.; Nakatsuji, H.; et al. *Gaussian 16*; Gaussian Inc.: Wallingford, UK, 2016.
81. Duchowicz, P. Linear Regression QSAR Models for Polo-Like Kinase-1 Inhibitors. *Cells* **2018**, *7*, 13. [\[CrossRef\]](#)
82. Peter, S.C.; Dhanjal, J.K.; Malik, V.; Radhakrishnan, N.; Jayakanthan, M.; Sundar, D. Quantitative Structure-Activity Relationship (QSAR): Modeling Approaches to Biological Applications. In *Encyclopedia of Bioinformatics and Computational Biology*; Elsevier: Amsterdam, The Netherlands, 2019; pp. 661–676.
83. Liu, P.; Long, W. Current mathematical methods used in QSAR/QSPR studies. *Int. J. Mol. Sci.* **2009**, *10*, 1978–1998. [\[CrossRef\]](#) [\[PubMed\]](#)
84. Luco, J.M.; Ferretti, F.H. QSAR Based on Multiple Linear Regression and PLS Methods for the Anti-HIV Activity of a Large Group of HEPT Derivatives. *J. Chem. Inf. Comput. Sci.* **1997**, *37*, 392–401. [\[CrossRef\]](#)
85. Liman, W.; Oubahmane, M.; Hdoufane, I.; Bjij, I.; Villemain, D.; Daoud, R.; Cherqaoui, D.; El Allali, A. Monte Carlo Method and GA-MLR-Based QSAR Modeling of NS5A Inhibitors against the Hepatitis C Virus. *Molecules* **2022**, *27*, 2729. [\[CrossRef\]](#) [\[PubMed\]](#)
86. Liu, S.S.; Yin, C.S.; Wang, L.S. Combined MEDV-GA-MLR method for QSAR of three panels of steroids, dipeptides, and COX-2 inhibitors. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 749–756. [\[CrossRef\]](#) [\[PubMed\]](#)
87. Ambure, P.; Gajewicz-Skretna, A.; Cordeiro, M.N.D.S.; Roy, K. New workflow for QSAR model development from small data sets: Small dataset curator and small dataset modeler. Integration of data curation, exhaustive double cross-validation, and a set of optimal model selection techniques. *J. Chem. Inf. Model.* **2019**, *59*, 4070–4076. [\[CrossRef\]](#) [\[PubMed\]](#)
88. Tsou, L.K.; Yeh, S.H.; Ueng, S.H.; Chang, C.P.; Song, J.S.; Wu, M.H.; Chang, H.F.; Chen, S.R.; Shih, C.; Chen, C.T.; et al. Comparative study between deep learning and QSAR classifications for TNBC inhibitors and novel GPCR agonist discovery. *Sci. Rep.* **2020**, *10*, 16771. [\[CrossRef\]](#) [\[PubMed\]](#)
89. Mao, J.; Akhtar, J.; Zhang, X.; Sun, L.; Guan, S.; Li, X.; Chen, G.; Liu, J.; Jeon, H.N.; Kim, M.S.; et al. Comprehensive strategies of machine-learning-based quantitative structure-activity relationship models. *iScience* **2021**, *24*, 103052. [\[CrossRef\]](#)
90. Chakravarti, S.K.; Alla, S.R.M. Descriptor Free QSAR Modeling Using Deep Learning with Long Short-Term Memory Neural Networks. *Front. Artif. Intell.* **2019**, *2*, 17. [\[CrossRef\]](#)
91. Stitou, M.; Toufik, H.; Bouachrine, M.; Bih, H.; Lamchouri, F. Machine learning algorithms used in Quantitative structure—Activity relationships studies as new approaches in drug discovery. In Proceedings of the 2019 International Conference on Intelligent Systems and Advanced Computing Sciences (ISACS), Taza, Morocco, 26–27 December 2019; pp. 1–8. [\[CrossRef\]](#)
92. Ancuceanu, R.; Dinu, M.; Neaga, I.; Laszlo, F.G.; Boda, D. Development of QSAR machine learning-based models to forecast the effect of substances on malignant melanoma cells. *Oncol. Lett.* **2019**, *17*, 4188–4196. [\[CrossRef\]](#)

93. Keyvanpour, M.R.; Shirzad, M.B. An Analysis of QSAR Research Based on Machine Learning Concepts. *Curr. Drug Discov. Technol.* **2021**, *18*, 17–30. [\[CrossRef\]](#)
94. Mosier, P.D.; Jurs, P.C. QSAR/QSPR studies using probabilistic neural networks and generalized regression neural networks. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1460–1470. [\[CrossRef\]](#) [\[PubMed\]](#)
95. Lowe, R.; Mussa, H.Y.; Mitchell, J.B.O.; Glen, R.C. Classifying molecules using a sparse probabilistic kernel binary classifier. *J. Chem. Inf. Model.* **2011**, *51*, 1539–1544. [\[CrossRef\]](#) [\[PubMed\]](#)
96. Ghasemi, J.; Saaidpour, S. Artificial neural network-based quantitative structural property relationship for predicting boiling points of refrigerants. *QSAR Comb. Sci.* **2009**, *28*, 1245–1254. [\[CrossRef\]](#)
97. Shahlaei, M.; Fassihi, A.; Saghaie, L. Application of PC-ANN and PC-LS-SVM in QSAR of CCR1 antagonist compounds: A comparative study. *Eur. J. Med. Chem.* **2010**, *45*, 1572–1582. [\[CrossRef\]](#)
98. Ajmani, S.; Jadhav, K.; Kulkarni, S.A. Three-dimensional QSAR using the k-nearest neighbor method and its interpretation. *J. Chem. Inf. Model.* **2006**, *46*, 24–31. [\[CrossRef\]](#)
99. Myint, K.Z.; Wang, L.; Tong, Q.; Xie, X.Q. Molecular fingerprint-based artificial neural networks QSAR for ligand biological activity predictions. *Mol. Pharm.* **2012**, *9*, 2912–2923. [\[CrossRef\]](#)
100. Yaffe, D.; Cohen, Y. Neural network based temperature-dependent quantitative structure property relations (QSPRs) for predicting vapor pressure of hydrocarbons. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 463–477. [\[CrossRef\]](#)
101. Burden, F.R.; Ford, M.G.; Whitley, D.C.; Winkler, D.A. Use of Automatic Relevance Determination in QSAR Studies Using Bayesian Neural Networks. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1423–1430. [\[CrossRef\]](#)
102. Espinosa, G.; Yaffe, D.; Arenas, A.; Cohen, Y.; Giralt, F. A fuzzy ARTMAP-based Quantitative Structure-Property Relationship (QSPR) for predicting physical properties of organic compounds. *Ind. Eng. Chem. Res.* **2001**, *40*, 2757–2766. [\[CrossRef\]](#)
103. Burden, F.R.; Winkler, D.A. Robust QSAR models using bayesian regularized neural networks. *J. Med. Chem.* **1999**, *42*, 3183–3187. [\[CrossRef\]](#)
104. Koutsoukas, A.; Lowe, R.; Kalantarmotamedi, Y.; Mussa, H.Y.; Klaffke, W.; Mitchell, J.B.O.; Glen, R.C.; Bender, A. Erratum: “in silico target predictions: Defining a benchmarking data set and comparison of performance of the multiclass naïve bayes and parzen-rosenblatt window”. *J. Chem. Inf. Model.* **2014**, *54*, 2180–2182. [\[CrossRef\]](#)
105. Ghasemi, F.; Mehridehnavi, A.; Fassihi, A.; Pérez-Sánchez, H. Deep neural network in QSAR studies using deep belief network. *Appl. Soft Comput. J.* **2018**, *62*, 251–258. [\[CrossRef\]](#)
106. Netzeva, T.I.; Schultz, T.W.; Aptula, A.O.; Cronin, M.T.D. Partial Least Squares modelling of the acute toxicity of aliphatic compounds to *Tetrahymena pyriformis*. *SAR QSAR Environ. Res.* **2003**, *14*, 265–283. [\[CrossRef\]](#) [\[PubMed\]](#)
107. Schultz, T.W.; Cronin, M.T.D.; Netzeva, T.I.; Aptula, A.O. Structure-toxicity relationships for aliphatic chemicals evaluated with *Tetrahymena pyriformis*. *Chem. Res. Toxicol.* **2002**, *15*, 1602–1609. [\[CrossRef\]](#) [\[PubMed\]](#)
108. Toropov, A.A.; Benfenati, E. QSAR modelling of aldehyde toxicity against a protozoan, *Tetrahymena pyriformis* by optimization of correlation weights of nearest neighboring codes. *J. Mol. Struct. THEOCHEM* **2004**, *679*, 225–228. [\[CrossRef\]](#)
109. Castro, E.A.; Toropov, A.A.; Nesterova, A.I.; Nazarov, A.U. QSAR study of the toxic action of aliphatic compounds to the bacteria *Vibrio fischeri* based on correlation weighting of local graph invariants. *J. Mol. Struct. THEOCHEM* **2003**, *639*, 129–135. [\[CrossRef\]](#)
110. Chattaraj, P.K.; Roy, D.R.; Giri, S.; Mukherjee, S.; Subramanian, V.; Parthasarathi, R.; Bultinck, P.; Van Damme, S. An atom counting and electrophilicity based QSTR approach. *J. Chem. Sci.* **2007**, *119*, 475–488. [\[CrossRef\]](#)
111. Pecka, J.; Ponc, R. Simple analytical method for evaluation of statistical importance of correlations in QSAR studies. *J. Math. Chem.* **2000**, *27*, 13–22. [\[CrossRef\]](#)
112. Parthasarathi, R.; Padmanabhan, J.; Subramanian, V.; Maiti, B.; Chattaraj, P.K. Toxicity analysis of 33′44′5-pentachloro biphenyl through chemical reactivity and selectivity profiles. *Curr. Sci.* **2004**, *86*, 535–542.
113. Russell, R.C.G.; Williams, N.S.; Bulstrode, C.J.K. (Eds.) *Bailey & Love’s Short Practice of Surgery*, 23rd ed.; Arnold: London, UK, 2000.
114. Jana, G.; Pal, R.; Sural, S.; Chattaraj, P.K. Quantitative structure-toxicity relationship: An “in silico study” using electrophilicity and hydrophobicity as descriptors. *Int. J. Quantum Chem.* **2020**, *120*, e26097. [\[CrossRef\]](#)
115. Masand, V.H.; Mahajan, D.T.; Maldhure, A.K.; Rastija, V. Quantitative structure–activity relationships (QSARs) and pharmacophore modeling for human African trypanosomiasis (HAT) activity of pyridyl benzamides and 3-(oxazolo [4,5-b]pyridin-2-yl)anilides. *Med. Chem. Res.* **2016**, *25*, 2324–2334. [\[CrossRef\]](#)
116. Héberger, K. Sum of ranking differences compares methods or models fairly. *TrAC Trends Anal. Chem.* **2010**, *29*, 101–109. [\[CrossRef\]](#)
117. Kollár-Hunek, K.; Héberger, K. Method and model comparison by sum of ranking differences in cases of repeated observations (ties). *Chemom. Intell. Lab. Syst.* **2013**, *127*, 139–146. [\[CrossRef\]](#)
118. Roy, D.R.; Giri, S.; Chattaraj, P.K. Arsenic toxicity: An atom counting and electrophilicity-based protocol. *Mol. Divers.* **2009**, *13*, 551–556. [\[CrossRef\]](#) [\[PubMed\]](#)
119. Thanikaivelan, P.; Subramanian, V.; Raghava Rao, J.; Unni Nair, B. Application of quantum chemical descriptor in quantitative structure activity and structure property relationship. *Chem. Phys. Lett.* **2000**, *323*, 59–70. [\[CrossRef\]](#)