World Journal of Organic Chemistry, 2015, Vol. 3, No. 1, 16-26
Available online at http://pubs.sciepub.com/wjoc/3/1/4

DOI:10.12691/wjoc-3-1-4

Cytotoxicity, 2D- and 3D- QSAR Study of some Halogen Containing Hydroxy and Amino Substituted Aromatic Compounds

M. Idrish Ali, M. Abul Kashem Liton*

Department of Chemistry, Mawlana Bhashani Science and Technology University, Santosh, Tangail, Bangladesh
*Corresponding author: litonchem@hotmail.com

Received December 09, 2015; Revised December 24, 2015; Accepted January 04, 2016

Abstract A set of 24 halogen containing hydroxy and amino substituted aromatic compounds were subjected to 2D- and 3D-QSAR studies. 3D-QSAR was studied at a 2.0 Å 3D grid spacing using molecular interaction fields (MIFs) analysis. The best predictive models by MIFs gave the cross-validated correlation coefficient, Q^2 of 0.668 and squared correlation coefficient, R^2 of 0.979 and the models by MLR, PCR and PLSR methods for 2D-QSAR provided a highly significant squared correlation coefficient (R^2) values of 0.904, 0.785, 0.903 and cross-validated correlation coefficients (Q^2) of 0.824, 0.662 and 0.718 respectively. The statistically significant model was established from a training set of 18 molecules, which were validated by evaluation of test set of 6 compounds. The calculated cytotoxic activities through MIFs model showed a very good agreement with experimental values. The information provided by QSAR analysis may give valuable clues to design and find the new potential drugs.

Keywords: *QSAR, LOO, MLR, PCR, PLSR and MIFs*

Cite This Article: M. Idrish Ali, and M. Abul Kashem Liton, "Cytotoxicity, 2D- and 3D- QSAR Study of some Halogen Containing Hydroxy and Amino Substituted Aromatic Compounds." *World Journal of Organic Chemistry*, vol. 3, no. 1 (2015): 16-26. doi: 10.12691/wjoc-3-1-4.

1. Introduction

Aromatic halides are a crucial step in the preparation of various synthetic intermediates or final products. Brominated and chlorinated aromatic compounds are widely used as intermediates in the manufacture of pharmaceuticals, agrochemicals and other chemical products [1]. The manufacture of a range of bulk and fine chemicals, including flame retardants, disinfectants, antibacterial and antiviral drugs, involve halogenations [2]. The halogen containing aromatic amines are also important building blocks in synthetic organic chemistry for the synthesis of a number of natural and bio-active substances. Numerous biologically active molecules and a large number of industrially valuable products such as pesticides, insecticides, herbicides, fire retardants, and other firsthand materials carry amino and halogen functionalized aromatic units in their structure [3]. These amino halides are involved in many famous organic reactions such as Stille, Suzuki, Heck, and Sonogashira and contribute access to accumulation of amino functionalized intermediates of enormous synthetic utility [4]. As well as halogenated phenol compounds are toxic and their most general property toxicity is useful. These compounds and their derivatives are used as a bactericide and fungicide and preservative. Due to the vast uses of phenolic compounds for industry productions, these compounds can spread through air and water, with strong carcinogenicity, teratogenicity and mutagenicity [5,6], which will cause a great damage to environment and also for living organisms. Again heterocyclic compounds are abundant in nature and are of great significance to life because their structural subunits exist in many natural products such as vitamins, hormones, and antibiotics [7]; hence, they have attracted considerable attention in the design of biologically active molecules [8] and advanced organic chemistry. Specially nitrogen containing heterocyclic aromatic compounds are an important class of compounds in the medicinal chemistry and also contributed to the society from biological and industrial point which helps to understand life processes [9].

Taking into consideration of this biological activity, we have synthesized some halogen containing hydroxy and amino substituted aromatic compounds and those were tested for cytotoxic activity. Though experimental procedure is the direct way to measure the cytotoxicity of organic compounds but it has many limitations, such as requirement of myriads of trial organisms, high expense and time consuming [10]. Consequently, development of drugs is lengthy, laborious and expensive process. So computer aided drug design (CADD) can help us to predict the toxicity quickly and almost exactly. The most important step is to find the possible structural feature of the compounds with desired biological activity. In this work, two dimensional quantitative structure-activity relationship (2D-QSAR) and three dimensional quantitative structure-activity relationship (3D-QSAR) were carried out to study the biological activity for the synthesized

compounds. With the rapid development of computational chemistry; many methods, algorithms and techniques have been discovered and applied in QSAR studies [11,12]. QSAR can predict the bioactivity based on structural parameters of compounds, and so now QSARs are being applied in many disciplines for designing drugs. It is a mathematical model that was used to evaluate the biological activity of a compound from its physiochemical properties of molecular structures.

2. Experimental

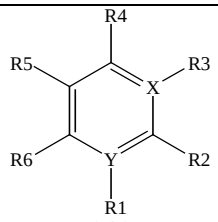
2.1. General Methods

A new method was employed for bromination and chlorination of some amino and hydroxy substituted aromatic compounds by the polymer-supported trimethylammonium dichlorobromide reagent at solid/solution phase system. This method was developed from the pioneering work by Merrifield [13] polymer-supports. The polymer could be recycled for the repeated reactions. All the reactions have been done under inert conditions at room temperature. All the reagents and solvents used in these experiments were pure, dried and purchased from E-Merck (Germany) and Fluka. Thin layer chromatographic analysis was performed on E-Merck 60 F 254 pre-coated aluminum thin layer chromatographic plates and column chromatography was done for separation of the products using silica gel (0.040-0.063 nm). Melting points were determined on a Fisher-John's electrochemical melting point apparatus and these were uncorrected. All the products were characterized by IR, ^{1}H-NMR, ^{13}C-NMR, ^{13}C-DEPT NMR spectroscopic and Mass spectrophotometric analysis. The IR spectra were recorded with KBr (Potassium Bromide) disk or thin film on DR-8001, Shimadzu FT-IR spectrometer and NMR spectra were also recorded in $CDCl_3$ on a Bruker 400 MHz spectrometer using TMS (Trimethylsilane) as an internal standard.

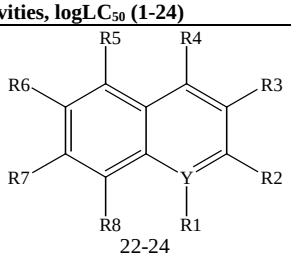
2.2. Synthesis

Twenty four (24) compounds were synthesized according to the literature procedure of Andreas Kirschning [14] and have been published by our research group [15]. All the synthesized compounds illustrated in Table 1 were confirmed with m.p. (melting point), IR, ^{1}H-NMR, ^{13}C-NMR, ^{13}C-DEPT NMR and Mass spectrophotometric analysis. Subsequently, cytotoxic activity of all synthesized compounds has been estimated with brine shrimps lethality bio-assay. The synthetic portion is not included in this manuscript due to the main focus on 2D- and 3D-QSAR study.

Table 1. List of the synthesized compounds with biological activities, logLC_{50} (1-24)



1-21



22-24

Comp. ID	X	Y	R1	R2	R3	R4	R5	R6	R7	R8	log LC_{50}
1	C	C	H	OH	CH_3CO	H	Br	H	-	-	1.24
2	C	C	Br	OH	CH_3CO	H	H	H	-	-	1.25
3	C	C	Br	OH	CH_3CO	H	Br	OH	-	-	0.39
4	C	C	Br	OH	CH_3CO	OH	Br	H	-	-	0.49
5	C	C	Br	H	CH_3CO	H	Br	OH	-	-	0.73
6	C	C	Br	H	CH_2ClCO	H	Br	OH	-	-	0.03
7	C	C	Br	H	Cl	H	Cl	OH	-	-	0.49
8	C	C	Br	H	CH_3CO	H	Br	NH_2	-	-	1.10
9	C	C	Br	H	CH_2ClCO	H	Br	NH_2	-	-	0.30
10	C	C	Br	H	Cl	H	Br	NH_2	-	-	0.48
11	C	C	H	Br	OH	Br	H	Br	-	-	0.27
12	C	C	H	Br	OH	Br	H	NO_2	-	-	1.12
13	C	C	H	Br	OH	H	H	NO_2	-	-	0.98
14	C	C	H	Br	NH_2	Br	H	Br	-	-	0.60
15	C	C	H	Br	NH_2	Cl	H	Br	-	-	0.83
16	C	C	H	Br	NH_2	Br	H	NO_2	-	-	1.74
17	C	C	H	Br	NH_2	Cl	H	NO_2	-	-	1.79
18	C	N	-	OH	Br	H	Br	CH_3	-	-	1.01
19	N	N	-	NH_2	-	H	Br	H	-	-	0.26
20	C	N	-	NH_2	Br	H	H	NH_2	-	-	0.54
21	C	N	-	NH_2	Cl	H	Cl	H	-	-	0.65
22	-	C	Cl	H	Cl	OH	H	H	H	H	2.37
23	-	N	-	H	H	H	Br	H	Cl	OH	1.61
24	-	N	-	H	Cl	H	Cl	H	H	OH	2.60

2.3. Cytotoxicity Calculation

The median lethal concentration (LC_{50}) with 95% confidence intervals of the test samples in Table 3 and reference standard (Vincristine Sulfate) in Table 4 were calculated using the probit analysis program [16,17] of IBM SPSS Statistics20 software packages. According to the research methodology for QSAR analysis, all the experimental LC_{50} values (µg/mL) were converted to the tenth based logarithm of LC_{50}, i.e., $logLC_{50}$.

2.4. Test Animal

Brine shrimps were used as test animal for the investigation of cytotoxic activity [18,19,20] and its scientific name is *Artemia Salina*.

2.5. Hatching of Shrimp

Brine shrimp (*A. Salina*) eggs were hatched in a vessel containing sterile artificial seawater prepared by dissolving 38 g of table salt in 1L distilled water. The vessel was kept under an inflorescent bulb and facilitated with good aeration for 48 h at room temperature. After hatching, nauplii released from the egg shells were collected at the bright side of the vessel (near the light source) by using micropipette. The larvae were isolated from the eggs by aliquoting them in small beaker containing the seawater.

2.6. Brine Shrimp Lethality Bioassay

The brine shrimp lethality bioassay was used to predict the cytotoxic activity [21,22] of the compounds. For the experiments, 4 mg of each test sample was dissolved in dimethylsulfoxide (DMSO) and solutions of varying concentrations (400, 100, 50, 10, 5, and 2 µg/mL) were obtained by the serial dilution technique using simulated seawater. The solutions were then added to the pre-marked glass vials containing 20–25 live brine shrimp nauplii in 10 mL simulated seawater. After 24 h, the vials were inspected using a magnifying glass, and the number of survived nauplii in each vial was counted. The mortality endpoint of this bioassay was defined as the absence of controlled forward motion during 30 s of observation [23]. From this data, the percent of lethality of the brine shrimp nauplii for each concentration and control was calculated. Vincristine Sulfate (reference standard) and DMSO were used as positive control and negative control respectively. All the procedures were replicated three times.

3. Computational Details

3.1. Data Set

2D- and 3D-QSAR modelling were applied to a set of 24 molecules, which were divided into a training set [24] of 18 molecules, and a test set of 6 molecules in a random manner (the test set is marked by * in Table 7). According to research methodology, all experimental LC_{50} values (µg/mL) were converted to the tenth-based logarithm of LC_{50}, i.e., $logLC_{50}$ and used as the dependent variable in QSAR studies. The structures of these compounds and their experimental biological activities are shown in Table 1.

Table 2. Stepwise multi-linear regression (MLR) was employed to select the best QSAR model

Model No	Descriptor (s) used in the model	R^2	RMSE	Coefficient	*t*-stat	*p*-value	Status of descriptor
1	XLogP	0.627	0.469	-0.518592	-6.5431	0.0000	
2	LUMO	0.773	0.348	0.544693	1.7268	0.1047	ns
	XLogP			-0.582043	-.9870	0.0000	
3	IP	0.822	0.321	0.394681	1.9627	0.0699	ns
	LUMO			0.863928	2.6042	0.0208	
	XLogP			-0.511907	-6.0719	0.0000	
4	IP	0.904	0.244	0.666700	3.8302	0.0021	
	LUMO			0.991600	3.8702	0.0019	
	$PPSA_1$			0.005200	3.3182	0.0056	
	XLogP			-0.355600	-4.4583	0.0006	
5	HF	0.905	0.251	-0.001380	-0.5300	0.6058	ns
	IP			0.639900	3.4397	0.0049	
	LUMO			0.975800	3.6778	0.0032	
	$PPSA_1$			0.005030	3.0354	0.0104	
	XLogP			-0.374600	-4.1837	0.0013	
6	HF	0.906	0.262	-0.001420	-0.5154	0.6165	ns
	TE			0.000032	0.0907	0.9294	ns
	IP			0.634200	3.1028	0.0101	
	LUMO			0.941800	2.0219	0.0682	ns
	$PPSA_1$			0.005000	2.8941	0.0146	
	XLogP			-0.372100	-3.8121	0.0029	
7	HF	0.906	0.262	-0.001420	-0.5154	0.6165	ns
	TE			0.000032	0.0907	0.9294	ns
	IP			0.634200	3.1028	0.0101	
	HOMO			0.000024	0.0691	0.9613	ns
	LUMO			0.941800	2.0219	0.0682	ns
	$PPSA_1$			0.005000	2.8941	0.0146	
	XLogP			-0.372100	-3.8121	0.0029	

ns: not significant (p-value>0.05).

3.2. Calculating Descriptors

At first all the 3D-structures of the compounds were generated by Gauss View03 software, and minimization of 3D- structure was performed by the MOPAC-2012 software [25] using the semi-empirical (PM6) method [26]. All geometric variables were finally optimized for each compound using a Gaussian03W program [27] (ver. 6.0) at B3LYP/6-31G(d,p) level of theory [28] and the low-energy conformers were ensured with all real frequencies by the frequency calculations, and these lowest energy conformers were selected for descriptors calculation. Briefly, the CDK Descriptor calculator [29] (v1.3.4) was employed to calculate the molecular descriptors and overall, more than 180 theoretical descriptors were calculated. These descriptors can be classified into several groups, including: (i) constitutional, (ii) geometrical, (iii) topological, (iv) electronic, (v) BCUT (Burden–CAS-University of Texas) and vi) WHIM (Weighted Holistic Invariant Molecular). QM (Quantum chemical) descriptors like HOMO and LUMO energies, heat of formation, dipole moment, square dipole moment, surface area, surface volume, energy gap and total energy were further calculated using Gaussian03W (ver. 6.0) program [27] at B3LYP/6-31G(d,p) level of theory [28]. From 210 different descriptors including quantum chemical descriptors which having less than 0.7 correlations were retained for further analyses and finally 7 descriptors were selected for 2D-QSAR study (Table 2).

Table 3. Cytotoxic activity of the synthesized compounds against brine shrimp nauplii

Comp ID	Concentration of the test solution (µg/mL)	Probit values	LC_{50}	$logLC_{50}$	95% confidence limit	
					Lower	Upper
1	2, 5, 10, 50, 100, 400	4.33, 4.64, 4.87, 5.36, 5.55, 5.77	17.47	1.24	11.75	25.70
2	2, 5, 10, 50, 100, 400	4.29, 4.64, 4.87, 5.36, 5.55, 5.77	17.83	1.25	12.02	25.70
3	2, 5, 10, 50, 100, 400	4.95, 5.13, 5.13, 5.67, 5.77, 6.08	2.467	0.39	1.10	4.68
4	2, 5, 10, 50, 100, 400	4.90, 5.05, 5.28, 5.58, 5.71, 5.92	3.12	0.49	1.15	5.89
5	2, 5, 10, 50, 100, 400	4.82, 4.95, 5.13, 5.41, 5.55, 5.67	5.39	0.73	1.95	10.47
6	2, 5, 10, 50, 100, 400	5.08, 5.20, 5.36, 5.67, 5.67, 5.84	1.08	0.03	0.13	3.02
7	2, 5, 10, 50, 100, 400	4.87, 5.08, 5.31, 5.58, 5.71, 5.92	3.06	0.49	1.12	5.89
8	2, 5, 10, 50, 100, 400	4.75, 4.87, 4.98, 5.18, 5.31, 5.44	12.64	1.10	4.47	26.92
9	2, 5, 10, 50, 100, 400	5.00, 5.15, 5.25, 5.50, 5.67, 5.84	2.01	0.30	0.83	4.21
10	2, 5, 10, 50, 100, 400	4.87, 5.08, 5.31, 5.61, 5.71, 5.92	3.03	0.48	1.10	5.75
11	2, 5, 10, 50, 100, 400	5.03, 5.15, 5.25, 5.47, 5.64, 5.84	1.87	0.27	0.35	4.57
12	2, 5, 10, 50, 100, 400	4.75, 4.85, 4.98, 5.18, 5.31, 5.44	13.17	1.12	4.90	28.18
13	2, 5, 10, 50, 100, 400	4.77, 4.90, 5.03, 5.23, 5.36, 5.50	9.51	0.98	3.24	19.95
14	2, 5, 10, 50, 100, 400	4.82, 5.03, 5.25, 5.52, 5.71, 5.92	4.00	0.60	1.70	7.24
15	2, 5, 10, 50, 100, 400	4.80, 4.92, 5.08, 5.33, 5.50, 5.61	6.77	0.83	2.51	13.18
16	2, 5, 10, 50, 100, 400	4.62, 4.72, 4.80, 5.00, 5.10, 5.20	54.71	1.74	22.91	213.80
17	2, 5, 10, 50, 100, 400	4.59, 4.70, 4.77, 5.00, 5.10, 5.18	62.06	1.79	26.30	251.19
18	2, 5, 10, 50, 100, 400	4.75, 4.90, 5.03, 5.20, 5.36, 5.50	10.26	1.01	3.72	20.89
19	2, 5, 10, 50, 100, 400	5.03, 5.15, 5.25, 5.52, 5.64, 5.84	1.83	0.26	0.35	4.37
20	2, 5, 10, 50, 100, 400	4.85, 5.05, 5.28, 5.55, 5.74, 5.92	3.50	0.54	1.38	6.46
21	2, 5, 10, 50, 100, 400	4.80, 5.00, 5.23, 5.52, 5.71, 5.88	4.43	0.65	1.95	7.76
22	2, 5, 10, 50, 100, 400	4.42, 4.50, 4.64, 4.80, 4.92, 5.05	232.60	2.37	85.11	2089.30
23	2, 5, 10, 50, 100, 400	4.59, 4.70, 4.80, 5.03, 5.15, 5.31	40.78	1.61	19.50	102.33
24	2, 5, 10, 50, 100, 400	4.42, 4.53, 4.56, 4.75, 4.87, 5.00	395.10	2.60	120.23	8709.64

Table 4. Cytotoxicity of the reference standard (Vincristine Sulfate) on brine shrimp nauplii

Concentration tested (µg/mL)	Probit	LC_{50} (µg/mL)	$logLC_{50}$	95% Confidence limit
2	4.75	4.71	0.67	2.88-6.92
5	5.00			
10	5.25			
50	5.67			
100	6.04			
400	6.41			

3.3. Molecular Modelling and Alignment

The molecules were superimposed using the atom-based alignment by the open3DALIGN tools [30,31] given in Figure 6. From the data set, the compound, 6 shown in Figure 5 was selected as template to construct other compounds because of its high biological activity and the alignment was completed by open3DALIGN workstation. Except for some special notes, default values were chosen. Then their geometries were optimized by the RMS gradient [32] criterion method on MMFF94s force-field implemented in TINKER by using command option of open3DALIGN. The energy convergence criterion in alignment is 0.01 kcal/mol.

3.4. MIFs Study

The Molecular Interaction Fields (MIFs) are the interaction energies between a probe atom (or a molecule) and a set of aligned molecules. To generate the MIFs, a probe atom is systematically moved from one point to another for each aligned molecule within a defined 3D grid [33]. At each grid point, the interaction energy is calculated between the probe and the target molecule. In this study, the 24 aligned molecules were placed in various 3D cubic lattice spacing. The steric (van der Waals) and electrostatic (Coulombic) interaction energies were calculated for each molecule at the respective grid point using an alkyl carbon probe (default) with automatically assigned charges using OpenBabel utilities. Energies lower than -40.0 kcal/mol and greater than 40.0 kcal/mol were cut off because a few high values in the dataset may severely bias the model. After alignment, the molecules were put to 2.0 Å of 3D grid [33] spacing. The steric and electrostatic fields were then calculated using a 'CR' Alkyl Carbon atom with default charge and the cutoff energy was set between -40 to 40 kcal/mol. Regression analysis of the resulting field matrix was performed by Partial Least Squares regression (PLSR) [34] technique. To obtain the 3D-QSAR models, PLS analysis was performed using both steric and electrostatic fields. Cross-validation in PLS was carried out using the leave-one-out method (LOO) [35,36] to check the predictive ability of the models and to determine the optimal number of components to be used in the final 3D-QSAR models.

3.5. Statistical Methods

For 2D-QSAR study, preliminary model selection was performed by BuildQSAR (version 2.1.0.0) [37] software program. Next, the MATLAB (ver. 7.6.0.324) software package [38] was applied for detailed statistical analysis of the QSAR models. 3D-QSAR analysis by MIFs study was performed on an open3DQSAR tools [39] using Partial Least Square (PLS) technique through the NIPALS algorithm methodology [40]. Statistical calculations allowed for the selection of the models with the following characteristics: high squared correlation coefficient (R^2), high cross-validated correlation coefficient (Q^2), high Fischer's value (*F-test*), low standard deviation (*s*), correlation matrix (Table 5) among the parameters and the least number of descriptors involved.

Table 5. Correlation matrix among the descriptors

	E_{IP}	E_{LUMO}	$PPSA_1$	XLogP
E_{IP}	1.000			
E_{LUMO}	0.613	1.000		
$PPSA_1$	0.089	0.105	1.000	
XLogP	0.571	0.441	0.441	1.000

Table 6. Statistical results of training and test set by MLR, PCR, PLSR regression methods (2D-QSAR) and MIFs studies (3D-QSAR)

Model No.	Method	Training set (18)				Test set (6)		Optimal no of variables used	3D-Grid spacing	Fields Contribution	
		R^2	Q^2_{LOO}	s	F test	R^2_{pred}	SDEP			Electro-static	Steric
2D-QSAR											
1	MLR	0.904	0.824	0.207	30.479	0.867	0.245	4 Des	-	-	-
2	PCR	0.785	0.662	0.310	23.422	0.784	0.312	3 PCs			
3	PLSR	0.903	0.718	0.208	43.556	0.873	0.239	3 LVs	-	-	-
3D-QSAR (atom-based alignment)											
4	MIFs	0.979	0.668	0.097	111.639	0.795	0.304	5 PLSC	2.0 Å	0.436	0.564

MLR: Multi linear regression, PCR: Principle component regression, PLSR: Partial least square regression, LOO: Leave one out, s: Standard deviation, SDEP: Standard error of prediction, F test: Fischer's statistics, MIFs: Molecular interaction energy fields, Des: Descriptors, PC: Principle components, PLSC: PLS Components (LVs), LVs: Latent variables.

4. Results and Discussions

4.1. 2D-QSAR Analysis

The correlation between various descriptors [41] with biological activity is the most important means of structure–activity relationship (SAR) study. Thus the equation should use the minimum number of descriptors to obtain the best fit. To achieve this, a popular procedure is used to find out the saturation point, a point beyond which there is no considerable improvement in regression coefficient (R^2) values has been observed. By interpreting the resulting descriptors, it is possible to gain some insight into factors that are likely to govern the cytotoxic activity. The best QSAR model constructed with stepwise multiple linear regression (MLR) method is shown by the following equation:

Model-1 (MLR)

$$\log LC_{50} = 0.0052 \times PPSA_1 + 0.6667 \times E_{IP} + 0.9916 \times E_{LUMO} - 0.3556 \times X \log P - 4.9968 \quad (1)$$

$$n_{train} = 18; R^2 = 0.904; s = 0.207;$$
$$Ftest = 30.479; p < 0.0001;$$
$$Q^2 = 0.824; n_{test} = 6; R^2_{pred} = 0.867; SDEP = 0.245$$

where, n is the number of observations, R^2 is the squared correlation coefficient, s is the standard deviation, p is the statistical significance >99.9% with Fisher's statistic F, Q^2 is the cross-validated square correlation coefficient (internal validation), predictive correlation coefficient (External validation) is indicated by R^2_{pred} and $SDEP$ is the standard deviation error of prediction.

The seven descriptors used in the stepwise multiple linear regression (MLR) methods can be classified as follows: partition coefficient by atom-additive method [42] ($XlogP$), the charged partial surface area, CPSA [43] ($PPSA_1$) and quantum chemical [44-53] or QC (E_{HOMO}, E_{LUMO}, E_{IP}, ΔH_{FORM} and E_T) which are shown in Table 2. The contribution of each descriptor can be validated by means of its p- and t-value since it reveals the significance of the parameter within the models. An important observation during generating QSAR models by stepwise multiple linear regression (MLR) method is the p-value. From Table 2, an excellent improvement of R^2 and $RMSE$ values are observed for model-4 than that of model-1, 2 and 3 whereas a considerable improvement of R^2 and $RMSE$ are not observed for Model-5, 6, 7 than that of Model-4. In the meanwhile, the p-values of some descriptors for model-2, 3, 5, 6 and 7 are insignificant shown in Table 2. Therefore, model-1 and 4 are only valid models in this stepwise multiple regressions (MLR) method where model-4 is the best MLR model which is denoted above by Model-1 (MLR). The selection of descriptors is based on R^2 and $RMSE$ of one-parameter correlation, and the parameter's t- and p- values, and inter-correlation among the descriptors are also been considered for the descriptor selection in the best MLR model. A final model is obtained by four descriptors ($PPSA_1$, E_{IP}, E_{LUMO} and $XlogP$) with optimal values of R^2, $Ftest$, Q^2 and lowest value of s. Next principle components regression (PCR) and partial least square regression (PLSR) were also applied to generate the models [54,55] for quantitative structure–activity relationship (QSAR) between a set of molecular descriptors used in the MLR method and experimental activity.

Model-2 (PCR)

$$\log LC_{50} = 0.0035 \times PPSA_1 + 0.1541 \times E_{IP} + 0.0553 \times E_{LUMO} - 0.4073 \times X \log P - 0.5366 \quad (2)$$

$$n_{train} = 18; R^2 = 0.785; s = 0.310;$$
$$Ftest = 23.422; p < 0.0001;$$
$$Q^2 = 0.662; n_{test} = 6; R^2_{pred} = 0.784; SDEP = 0.312$$

Model-3 (PLSR)

$$\log LC_{50} = 0.0050 \times PPSA_1 + 0.6353 \times E_{IP} + 1.0017 \times E_{LUMO} - 0.3704 \times X \log P - 4.6286 \quad (3)$$

$$n_{train} = 18; R^2 = 0.903; s = 0.208;$$
$$Ftest = 43.556; p < 0.0001;$$
$$Q^2 = 0.718; n_{test} = 6; R^2_{pred} = 0.873; SDEP = 0.239$$

In the above MLR, PCR, PLSR models, R^2 value multiplied by 100 gives explained variance in biological activity of the compounds. The CSPA descriptor ($PPSA_1$) encodes features responsible for polar interactions between molecules. The definition by Jurs [43] of this descriptor (partial positive surface area) is represented as $PPSA_1 = \sum_i SA_i^+$ where SA_i^+ is the surface area of the positively charged atom i. It is positively correlated with the activity, indicating that increasing this value in molecules could increase biological activity. The molecule is considered as an assembling of hard spheres defined by the van der Waals radii of the atoms. The solvent-accessible surface area is traced out by the centre of a solvent sphere (usually water) that rolls over the van der Waals surface of the molecule. The Ionization Potential (E_{IP}) is defined as the minimum energy required to removing an electron from an isolated molecule (or atom) in its ground state to form an ion in the gas phase [56]. The E_{IP} parameter can be related to the electron transfer pathway, in which low values of E_{IP} favor the electron transfer process in a molecule and high values of E_{IP} decrease the electron donating capacity of a molecule rather increase electron accepting tendency of a molecule. Sjoberg *et al.* [57,58] extended their concept about ionization potential to the local electron affinity, which is the acceptor equivalent. These properties can also be described in terms of spherical harmonics in conjunction with a distance-dependent overlap term to calculate donor-acceptor interactions. Thus, surface-based descriptors calculated from the local ionization potential and electron affinity are able to describe intermolecular donor/acceptor interactions [59]. The electronic parameter E_{LUMO}, which denotes energy of the lowest unoccupied molecular orbital, directly related to the electron affinity and characterizes the susceptibility of the molecule towards attack by nucleophiles. An electron deficiency in the lowest unoccupied molecular orbital (LUMO), shows its low energy and low electron density, which gives a molecule a favorable condition for the acceptance of an electron. The positive slope of LUMO in this model indicates that the higher the E_{LUMO} energy of the molecule, the higher the

biological activity of the molecules. *XlogP* is the atom type descriptor [60] used to characterize the hydrophobicity (logP) of molecules. The atomic contribution of individual atom types was proposed by Ghose and Crippen [61] toward the overall hydrophobicity of molecules where carbon, hydrogen, oxygen, nitrogen, sulfur and halogens were classified into atoms. Hydrogen and halogens are classified by the hybridization and oxidation state of the carbon they are bonded to, and carbon atoms are classified by their hybridization state and the chemical nature of their neighbouring atoms. The negative slope of *XlogP* in above models represents that biological activity increases with lowering of hydrophobicity i.e. decreases with an increase in lipophilicity.

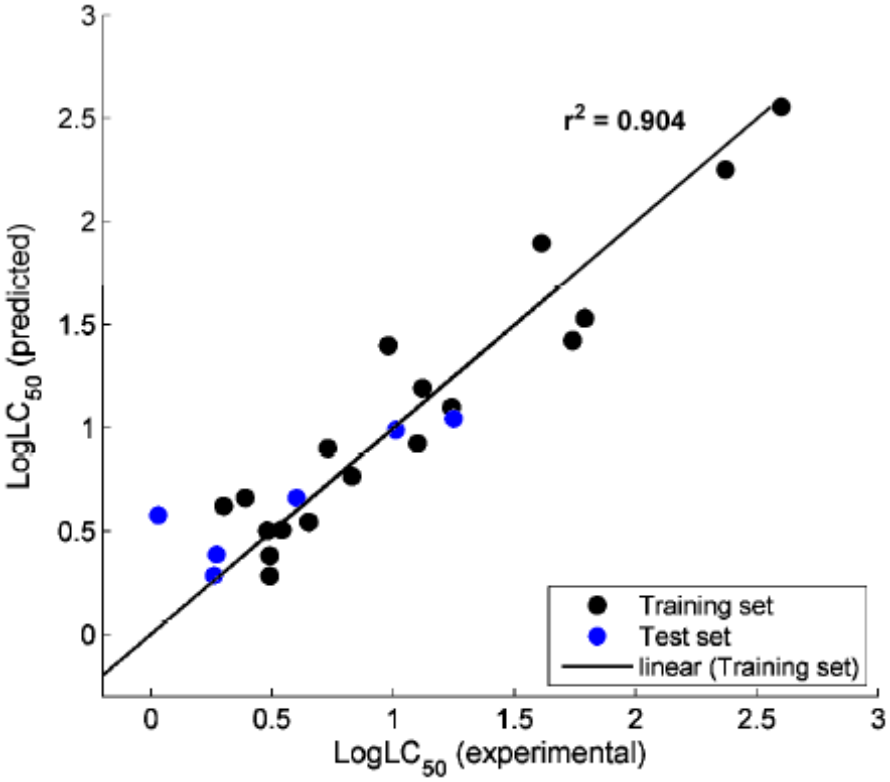


Figure 1. Graph of actual versus predicted activities for training and test set molecules by MLR method

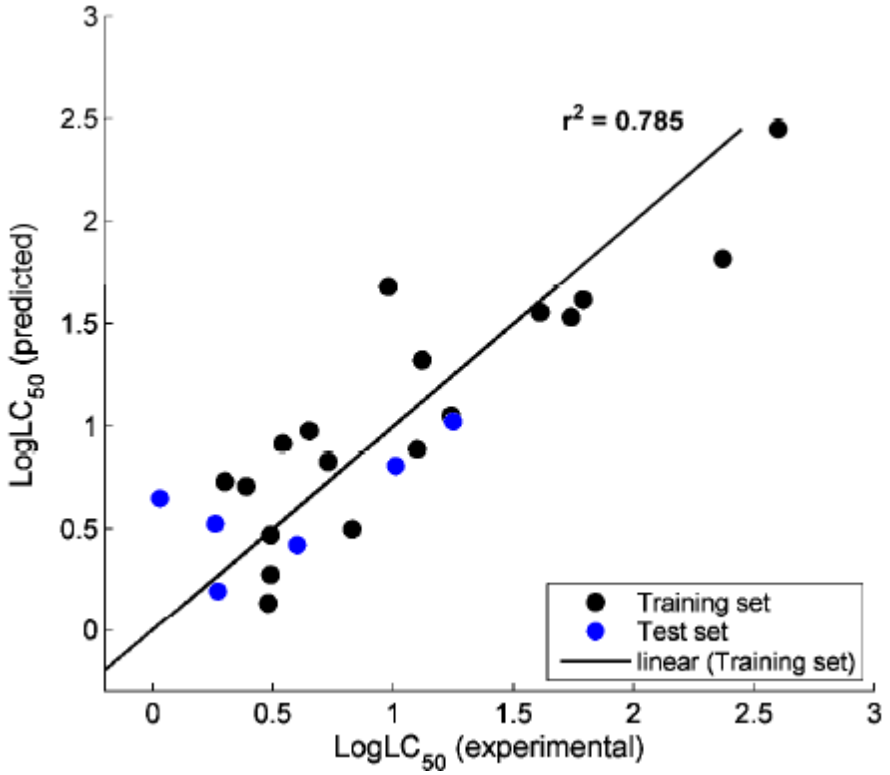


Figure 2. Graph of actual versus predicted activities for training and test set molecules by PCR method

In order to validate our results we have estimated the cytotoxic activity ($logLC_{50}$) of training and test sets using the model expressed by Eq. (1), (2) and (3) and compared them with the observed values. The data presented in Table 7 shows that the observed and predicted activities are very close to each other. This was further supported by the plot of MLR predicted $logLC_{50}$ values against the observed $logLC_{50}$ values (Figure 1) and the lowest values of the statistical parameters *s* and *SDEP* estimated in Table 6 are also in the favor of linear model constructed with $PPSA_1$, E_{IP}, E_{LUMO} and *XlogP* descriptors. Cross-validated square correlation coefficient (Q^2) by LOO technique was 0.824 which showed a good internal predictive ability of the model. Consequently Eq. (1) can be also considered as a perfect model for both high statistical significant and excellent predictive ability. It is observed that $PPSA_1$, E_{IP}, E_{LUMO} and *XlogP* are the best descriptors in the establishment of the QSAR model for halogen containing hydroxy and amino substituted aromatic compounds. The correlation of the experimental activities with the MLR calculated ones is illustrated in Figure 1. The optimal number of principal components (3 PCs) for PCR and optimal number of latent variables (3 LVs) for PLSR are used in 2D-QSAR analysis shown in Figure 7 and Figure 8 respectively. The correlation of the experimental and calculated activities with the PCR and PLSR methods are also shown in Figure 2 and Figure 3 respectively. The squared correlation coefficient R^2, standard deviation *s*, and Fischer Statistics *F* obtained with PCR and PLSR methods (Table 6) indicate that the models proposed to predict activities are significant and pertinent to that of MLR method [62]. The biological activities predicted by MLR, PCR and PLSR methods with respect to their experimental values are shown in Table 7.

Table 7. Predicted activities by MLR, PCR, PLSR and MIFs methods as well as experimental activities

Comp ID	$logLC_{50}$					Residual Errors			
	Exp.	2D-QSAR			3D-QSAR				
		MLR	PCR	PLSR	MIFs	MLR	PCR	PLSR	MIFs
1	1.24	1.10	1.05	1.08	1.09	0.14	0.19	0.16	0.15
2*	1.25	1.05	1.02	1.03	1.42	0.20	0.23	0.22	-0.17
3	0.39	0.66	0.71	0.65	0.40	-0.27	-0.32	-0.26	-0.01
4	0.49	0.28	0.47	0.27	0.64	0.21	0.02	0.22	-0.15
5	0.73	0.90	0.83	0.89	0.62	-0.17	-0.10	-0.16	0.11
6*	0.03	0.58	0.65	0.56	-0.10	-0.55	-0.62	-0.53	0.13
7	0.49	0.38	0.28	0.39	0.35	0.11	0.21	0.10	0.14
8	1.10	0.93	0.89	0.95	1.10	0.17	0.21	0.15	0.00
9	0.30	0.62	0.73	0.64	0.37	-0.32	-0.43	-0.34	-0.07
10	0.48	0.50	0.14	0.48	0.51	-0.02	0.34	0.00	-0.03
11*	0.27	0.39	0.20	0.37	0.33	-0.12	0.07	-0.10	-0.06
12	1.12	1.19	1.32	1.18	1.15	-0.07	-0.20	-0.06	-0.03
13	0.98	1.40	1.68	1.40	1.00	-0.42	-0.70	-0.42	-0.02
14*	0.60	0.66	0.42	0.65	0.96	-0.06	0.18	-0.05	-0.36
15	0.83	0.77	0.50	0.76	0.91	0.06	0.33	0.07	-0.08
16	1.74	1.42	1.53	1.42	1.76	0.32	0.21	0.32	-0.02
17	1.79	1.53	1.62	1.53	1.71	0.26	0.17	0.26	0.09
18*	1.01	0.99	0.81	1.00	1.49	0.02	0.20	0.01	-0.48
19*	0.26	0.29	0.52	0.30	0.62	-0.03	-0.26	-0.04	-0.36
20	0.54	0.51	0.92	0.51	0.55	0.03	-0.38	0.03	-0.01
21	0.65	0.55	0.98	0.56	0.63	0.10	-0.33	0.09	0.02
22	2.37	2.25	2.82	2.28	2.21	0.12	-0.45	0.09	0.16
23	1.61	1.90	1.56	1.90	1.82	-0.29	0.05	-0.29	-0.21
24	2.60	2.56	2.45	2.53	2.64	0.04	0.15	0.07	-0.04

*Test set; Exp.: Experimental.

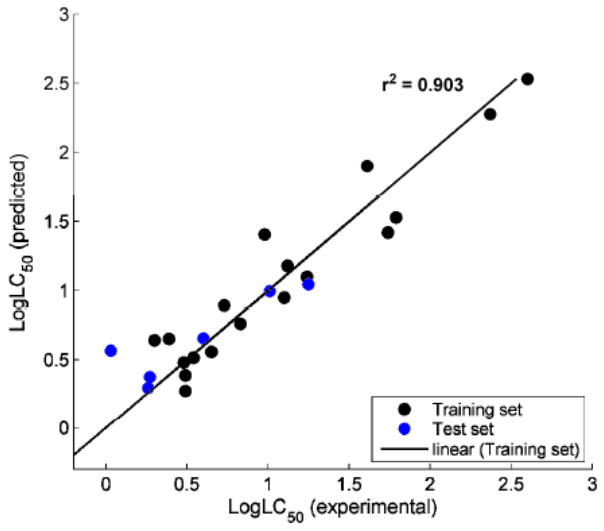


Figure 3. Graph of actual versus predicted activities for training and test set molecules by PLSR method

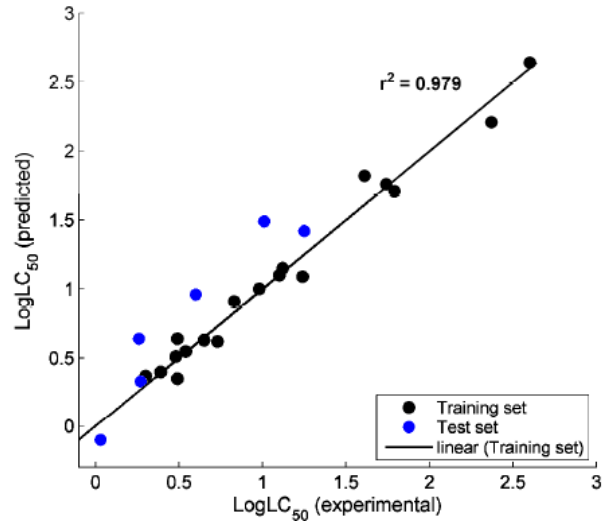


Figure 4. Graph of actual versus predicted activities for training and test set molecules by MIFs analysis

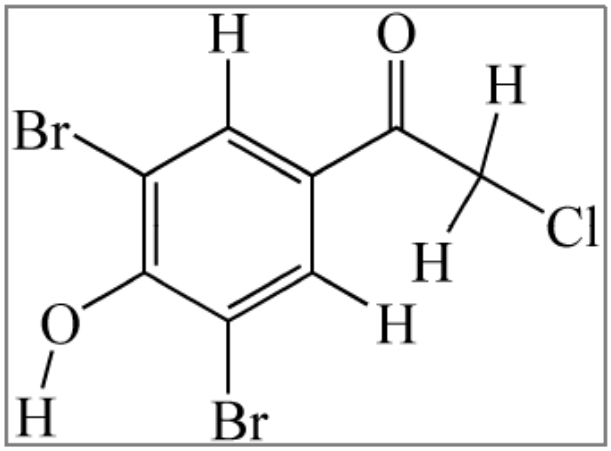


Figure 5. The most active compound, 6 used as a template for alignment of data set

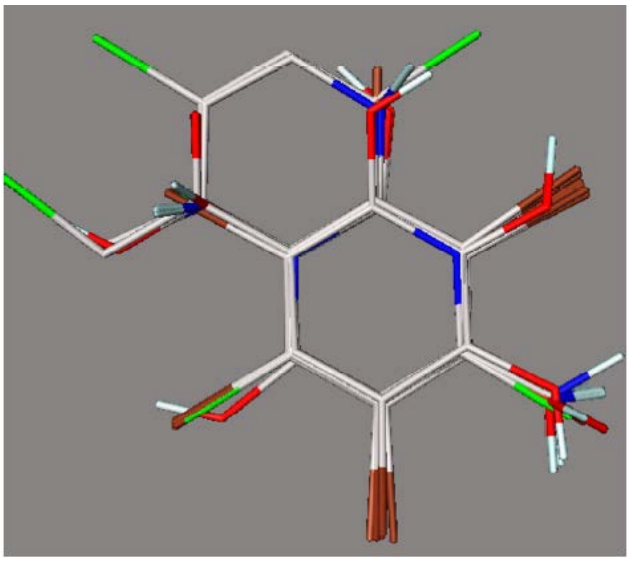


Figure 6. The atom-based alignment of the superimposed structure of all compounds used in the 3D-QSAR (MIFs) analysis

4.2. 3D-QSAR Analysis

To develop an effective 3D-QSAR model some parameters such as squared correlation coefficient (R^2), cross-validated correlation coefficient (Q^2), standard deviation (s) and F-statistic values have been taken under consideration. The squared correlation coefficient (R^2) and standard deviation (s) were carried out first for 3D-QSAR model. Then the number of components identified during the squared correlation coefficient (R^2) determination was used in the cross-validated PLS run. The optimal number of components was determined by selecting highest Q^2 value. A data set of 24 halogen containing hydroxy and amino substituted aromatic compounds was divided into a training set of 18 compounds for developing the MIFs model, and a test set of 6 compounds for evaluating the predictive ability of the model. The leave-one-out cross-validated PLS analysis gave a Q^2 of 0.668, using five PLS components (LVs), and the non-cross-validated PLS analysis yields a higher R^2 of 0.979 with a low standard deviation (s) of 0.097 and high *F-value* of 111.639. The model was externally validated by high R^2_{pred} of 0.795 with the low standard error of prediction (*SDEP*) of 0.304 for the test set. Both steric and electrostatic fields at 2.0Å 3D grid spacing were used to construct the model and the contributions of steric and electrostatic fields in the model were 56.4% and 43.6% respectively. The experimental and calculated activities ($logLC_{50}$) using the best 3D-QSAR MIFs model for training and test (indicated by * marks) set are shown in Table 7. Figure 4 showed a plot of observed versus calculated activity of training and test set molecules through MIFs analysis [63]. So the derived model was satisfactory with respect to high statistical results and predictive ability. Our present studies have established that the model derived through MIFs studies is quite reliable and significant. We have investigated that the MIFs analysis at 2.0Å 3D grid spacing by Open3DQSAR tools has presented an excellent statistical results in terms of Q^2 and R^2 values for halogen containing hydroxy and amino substituted aromatic compounds analogues and showed a high degree of agreement with the experimental values.

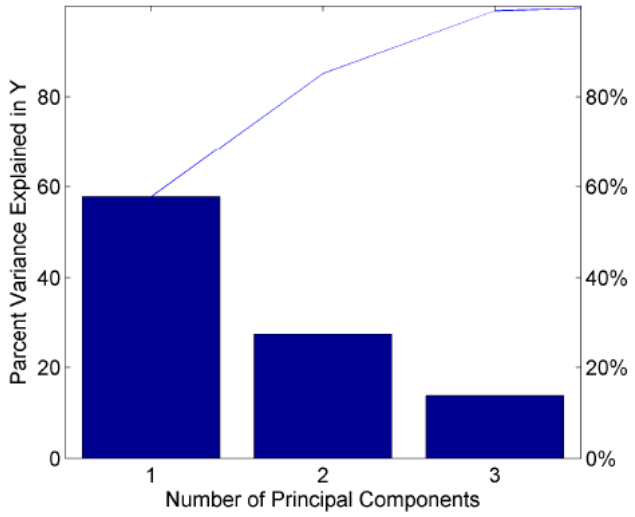


Figure 7. Optimal number of principal components (PCs) used in PCR for 2D-QSAR analysis

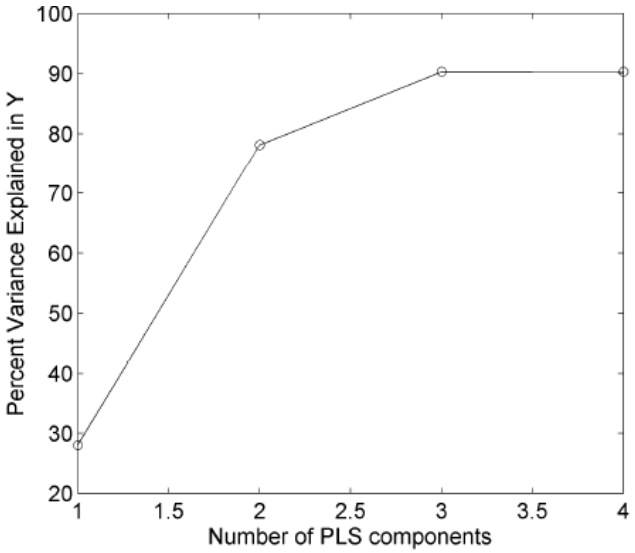


Figure 8. Optimal number of Latent variables (LVs) used in PLSR for 2D-QSAR analysis

5. Conclusion

The brine shrimp lethality bioassay is considered as a useful tool for rapid and preliminary assessment of toxicity of the compounds. Further studies are required to determine the more accurate bioactivity and, to find the mode of pharmacological activities. Significant regression equations were obtained by MLR, PCR, PLSR and MIFs methods with respect to their experimental cytotoxic activities. The best regression equation was obtained on the following descriptors: partial positive surface area ($PPSA_1$), potential energy (E_{IP}), lowest unoccupied molecular orbital energy (E_{LUMO}) and partition coefficient by atom-additive method (*XlogP*). These variables allowed physical explanation of electronic molecular properties contributing to the cytotoxic activity as the electronic character relates directly to the electron distribution of interacting molecules. The predicted biological activities by MLR, PCR, PLSR and MIFs showed a very good agreement with experimental values but the activities obtained by MIFs analysis were relatively better among them. The LOO cross-validation indicates that the model is significant, robust and has a good predictive ability.

Acknowledgement

The authors wish to thank Chairman, Department of Chemistry, Mawlana Bhashani Science and Technology University, Santosh, Tangail-1902, Bangladesh to use the departmental computational chemistry laboratory and for his kind support during the work.

References

[1] Armarego, W. L. F.; Chai, C. L. L.; Purification of Laboratory Chemicals, 5th ed., Butterworth Heinemann, London, 2003.

[2] Ullmann's Encyclopedia of Industrial Chemistry. 6th ed., Wiley-VCH: Weinheim. Electronic release; 1998.

[3] Seevers, R. H.; Counsell, R. E.; Radioiodination Techniques for Small Organic Molecules, *Chem. Rev.*, 82, 575-590, 1982.

[4] (a) Corbet, J. P.; Mignani, G.; Selected patented cross-coupling reaction technologies, *Chem. Rev.*, 106(7), 2651-2710, 2006. (b) Miyaura, N.; Suzuki, A.; Palladium-catalyzed Cross-coupling Reactions of Organoboron Compounds, *Chem. Rev.*, 95, 2457-2483, 1995.

[5] Zang, Y. B.; Research Advance of Phenol Adsorption of Modified Bentonite, *Chinese. Agri. Sci. Bull.*, 28, 282-285, 2012.

[6] Zhan, P. R.; Wang, H. T.; Chen, Z. X.; Extraction and Residue Analysis of Volatile Phenols in Fish, *J. Agro-Environ. Sci.* 27, 801-804, 2008.

[7] Ju, Y.; Varma, R. S.; Aqueous N-heterocyclization of Primary Amines and Hydrazines with Dihalides: Microwave-assisted Syntheses of N-Azacycloalkanes, Isoindole, Pyrazole, Pyrazolidine, and Phthalazine Derivatives, *J. Org. Chem.*,71(1), 135-141, 2006.

[8] Lokhande, P. D.; Waghamare, B. Y.; Sakate, S. S.; Regioselective One-pot Synthesis of 3,5-Diarylpyrazoles, *Indian. J. Chem.*, 44(11), 2338-2342, 2005.

[9] García-Valverde, M.; Torroba, T.; Special issue: Sulfur-Nitrogen Heterocycles, *Molecules*, 10(2), 318-320, 2005.

[10] Hou, T.; Xu, X.; A new molecular simulation software package - Peking University Drug Design System (PKUDDS) for structure-based drug design, *J. Mol. Graph. Model.* 19, 455-465, 2001.

[11] He, L.; Jurs, P. C.; Assessing the reliability of a QSAR model's predictions, *J. Mol. Graph. Model*. 23, 503-523, 2005.

[12] Eldred, D. V.; Weikel, C. L.; Jurs, P. C.; Kaiser, K. L. E.; Prediction of fathead minnow acute toxicity of organic compounds from molecular structure, *Chem. Res. Toxicol*. 12, 670-678, 1999.

[13] Merifield, R. B.; Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide, *J. Am. Chem. Soc.*, 85, 2149-2154, 1963.

[14] Kirschning, A.; Monenschein, H.; Wittenberg, R.; Functionalized Polymers-Emerging Versatile Tools for Solution-Phase Chemistry and Automated Parallel Synthesis, *Angew. Chem. Int. Ed. Engl.*,40(4), 650-679, 2001.

[15] Khan, M. A. E.; Ali, M. I.; Hashem, M. A.; Polymer-supported Ammonium Dichlorobromide(I) Reagent Promoted Halogenation of Hydroxy and Amino Substituted Aromatic Compounds, *J. Bang. Chem. Soc.*, 22(1), 55-65, 2009.

[16] Finney, J. D.; Probit Analysis. Cambridge University Press, UK., 1971.

[17] Rickman, R.; Mitchell, N; Dingman, J.; Dalen, J. E.; Changes in serum cholesterol during the Stillman Diet, *J. Am. Med. Assoc.*, 228, 54-58, 1974.

[18] Martin, T. M.; Harten, P.; Venkatapathy, R.; Das, S.; Young, D. M; A Hierarchical Clustering Methodology for the Estimation of Toxicity, Toxi*col. Mech. Meth.*, 18, 251-66, 2008.

[19] Islam, M. K.; Eti, I.Z.; Chowdury, J. A.; Cytotoxic studies on two Meliaceae plants: Chukrasiatabularis and Aglaia roxburghiana, *J. Sci. Res.*, 1, 399-403, 2009.

[20] Wang, H. H. K.; Xia, Y.; Yang, Z.Y.; Morris-Natschke, S. L.; Lee, K. H.; A recent advances in the discovery and development of flavonoids and their analogues as antitumor and anti-HIV agents, Plenum Press, New York, 1998, 191-225.

[21] McLaughlin, J. L.; Rogers, L. L.; Anderson, J. E.; The use of biological assays to evaluate botanicals, *Drug. Inform. J.*, 32, 513-524, 1998.

[22] Middleton, P.; Stewart, F.; Al-Qahtani, S.; Egan, P.; Rourke, C. O.; Sarker, S. D.; Antioxidant, antibacterial activities and general toxicity of Alnus glutinosa, Fraxinus excelsior and Papaver rhoeas, *Iranian. J. Pharm. Res.*2, 81-86, 2005.

[23] Meyer, B. N.; Ferrigni, N. R.; Putnam, J. E.; Brine shrimp: a convenient general bioassay for active plant constituents, *Planta. Medica*., 45, 31-34, 1982.

[24] Golbraikh, A.; Tropsha, A.; Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection, *Mol. Divers.*, 5, 231-243, 2002.

[25] a) Stewart, J. J. P.; Optimization of parameters for semiempirical methods I. Math., *J. Comp. Chem.*, 10, 209-220, 1989. b) Stewart, J. J. P.; Optimization of parameters for semiempirical methods II. Applications., *J. Comp. Chem.*, 10, 22-64, 1989.

[26] Stewart, J. J. P.; Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and application to 70 elements, *J. Mol. Model.* 13, 1173-1213, 2007.

[27] Frisch, M. J. et al.; Gaussian 03, Revision B.01, Gaussian Inc, Pittsburgh PA. 2003.

[28] Becke, A. D.; Density-Functional Thermochemistry. III. The Role of Exact Exchange, *J. Chem. Phys.*, 98, 5648-5652, 1993.

[29] a) Steinbeck, C.; Hoppe, C.; Kuhn, S.; Floris, M.; Guha, R.; Willighagen, E. L.; Recent developments of the chemistry development kit (CDK) - an open-source java library for chemo- and bioinformatics, *Curr. Pharm. Des.*, 12(17), 2111-2120, 2006. b) Steinbeck, C.; Han, Y.; Kuhn, S.; Horlacher, O.; Luttmann, E.; Willighagen, E. L.; The Chemistry Development Kit (CDK): an open-source Java library for Chemo- and Bioinformatics, *J. Chem. Inf. Comput. Sci.*, 43(2), 493-500, 2003.

[30] Richmond, N. J.; Willett, P.; Clark, R. D.; Alignment of three-dimensional molecules using an image recognition algorithm, *J. Mol. Graph. Model.*, 23(2), 199-209, 2004.

[31] Taminau, J.; Thijs, G.; De Winter, H.; Pharao: pharmacophore alignment and optimization, *J. Mol. Graph. Model*., 27, 161-169, 2008.

[32] Petitjean, M.; Interactive maximal common 3D substructure searching with the combined SDM/RMS algorithm, *Computers. Chem.*, 22(6), 463-465, 1998.

[33] Cramer, III R. D.; Milne, M.; The Lattice Model: A general paradigm for shape-related structure/activity correlation. Abstracts of Papers of the *Am. Chem. Soc.*, Computer Chemistry Section: no. 44, 1979.

[34] Stahle, L.; Wold, S. J.; Partial least squares analysis with cross-validation for the two-class problem: A Monte Carlo study, *J. Chemom.*, 1, 185-196, 1987.

[35] Centner, V.; Massart, D. L.; de Noord, O. E.; de Jong, S.; Vandeginste, B. M.; Sterna, C.; Elimination of uninformative variables for mutivariate calibration, *Anal. Chem.*, 68, 3851-3858, 1996.

[36] Grohmann, R.; Schindler, T.; Toward robust QSPR models: Synergistic utilization of robust regression and variable elimination, *J. Comput. Chem.*, 29(6), 847-860, 2008.

[37] De Oliveira, D. B.; Gaudio, A. C.; Build QSAR: a new computer program for QSAR analysis, *Quant. Struct-Activ. Relat.*, 19 (6), 599-601, 2001.

[38] MATLAB and Statistics Toolbox Release 2008a, The MathWorks, Inc., Natick, Massachusetts, United States. 2008.

[39] Tosco, P.; Balle, T.; Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields, *J. Mol. Mod.*, 17, 201-208, 2011.

[40] Wold, S.; Sjöström, M.; Eriksson, L.; PLS-regression: a basic tool of chemometrics, *Chemom. Intell. Lab. Syst.*, 58,109-130, 2001.

[41] (a) Kier, L. B.; Hall, L. H.; Molecular Connectivity in Structure-Activity Analysis, John Wiley and Sons, New York, 1986. (b) Kier, L. B.; Hall, L. H.; Moleculer Structure Description: The Electrotopological State, Academic Press, New York, 1999.

[42] Wang, R.; Gao, Y.; Lai, L.; Calculating partition coefficient by atom-additive Method, *Perspectives in Drug Discovery and Design*, 19, 47-66, 2000.

[43] Stanton, D. T.; Jurs, P. C.; Development and Use of Charged Partial Surface Area Structural Descriptors in Computer Assissted Quantitative Structure Property Relationship Studies, *Anal. Chem.*, 62, 2323-2329, 1990.

[44] Cartier, A.; Rivail, J. L.; Electronic descriptors in quantitative structure-activity relationships, *Chemom. Intell. Lab. Syst.*, 1(4), 335-347, 1987.

[45] Brown, R. E.; Simas, A. M.; On the applicability of CNDO indices for the prediction of chemical reactivity, Theor. Chim. Acta., 62, 1-16, 1982.

[46] Gruber, C.; Buss, V.; Quantum-mechanically calculated properties for the development of quantitative structure-activity relationships (QSAR'S). pKa-values of phenols and aromatic and aliphatic carboxylic acids, Chemosphere, 19, 1595-1609, 1989.

[47] Bodor, N.; Gabanyi, Z.; Wong, C. K.; A new method for the estimation of partition coefficient, *J. Am. Chem. Soc.*, 111, 3783-3786, 1989.

[48] Cocchi, M.; Menziani, M. C.; De Benedetti, P. G.; Cruciani, G.; Theoretical versus empirical molecular descriptors in monosubstituted benzenes—a chemometric study, *Chemom. Intell. Lab. Syst.*, 14, 209-224, 1992.

[49] Hanai, T.; Hatano, H.; Nimura, N.; Kinoshita, T.; Computational chemical analysis of the retention of phenols in reversed-phase liquid chromatography, *Analyst.*, 119, 1167-1170, 1994.

[50] Clare, B. W.; Structure-Activity Correlations for Psychotomimetics. III. Tryptamines, *Aust. J. Chem.*, 48, 1385-1400, 1995.

[51] Osmialowski, K.; Halkiewicz, J.; Radecki, A.; Kaliszan, R.; Quantum chemical parameters in correlation analysis of gas liquid chromatographic retention indices of amines, *J. Chromatogr.*, 346, 53-60, 1985.

[52] Osmialowski, K.; Halkiewicz, J.; Kaliszan, R.; Quantum chemical parameters in correlation analysis of gas liquid chromatographic retention indices of amines. II. Topological electronic index, *J. Chromatogr.*, 361, 63-69, 1986.

[53] Saura-Calixto, F; Garcı´a-Raso, A.; Raso, M. A.; Study of the Applications of Magnitudes of Energy and Charge of Molecular Orbitals to GC Retention-Esters, *J. Chromatogr. Sci.*, 22(1), 22-26, 1984.

[54] Geladi, P.; Kowalski, B. R.; Partial least-squares regression: a tutorial, *Anal. Chim. Acta*., 185, 1-17, 1986.

[55] Cramer, R. D.; Bunce, J. D.; Patterson, D. E.; Frank, I. E.; Crossvalidation, bootstrapping, and partial least squares compared with multiple regression in conventional QSAR studies, *Quant. Struct-Act. Relat*., 7, 18-25, 1988.

[56] McNaught, A. D.; Wilkinson, A.; Compendium of Chemical Terminology: IUPAC Recommendations (online version). 2nd ed., Oxford: Blackwell Science; 1997.

[57] a) Sjoberg, P.; Murray, J. S.; Brinck, T.; Politzer, P.; Average local ionization energies on the molecular surfaces of aromatic systems as guides to chemical reactivity, *Can. J. Chem.*, 68(8), 1440-1443, 1990. b) Politzer, P.; Murray, J. S.; Grice, M. E.; Brinck, T.; Ranganathan, S.; Radial behavior of the average local ionization energies of atoms, *J. Chem. Phys*., 95(9), 6699-6704, 1991. c) Politzer, P.; Murray, J. S.; Concha, M. C.; The complementary roles of molecular surface electrostatic potentials and average local ionization energies with respect to electrophilic processes, *Int. J. Quant. Chem.*, 88(1), 19-27, 2002.

[58] Ehresmann, B.; Martin, B.; Horn, A. H. C.; Clark, T.; Local molecular properties and their use in predicting reactivity, *J. Mol. Model.*, 9, 342-347, 2003.

[59] Schürer, G.; Gedeck, P.; Gottschalk, M.; Clark, T.; Accurate parametrized variational calculations of the molecular electronic polarizability by NDDO-based methods, *Int. J. Quant. Chem.*, 75(1), 17-31, 1999.

[60] Wang, R.; Fu, Y.; Lai, L.; A New Atom-Additive Method for Calculating Partition Coefficients, *J. Chem. Inf. Comput. Sci.*, 37, 615-621, 1997.

[61] a) Ghose, A. K.; Crippen, G. M.; Atomic physicochemical parameters for three-dimensional structure-directed quantitative structure-activity relationships. I. Partition coefficients as a measure of hydrophobicity, *J. Comput. Chem.*, 7, 565-577, 1986. b) Ghose, A. K.; Crippen, G. M.; Atomic physicochemical parameters for three-dimensional-structure-directed quantitative structure-activity relationships. 2. Modeling dispersive and hydrophobic interactions, *J. Chem. Inf. Comput. Sci.*, 27, 21-35, 1987.

[62] Liton, M. A. K.; Salma, U.; Bhowmick, A. C.; Cytotoxicity and 2D-QSAR study of some heterocyclic compounds, *Arabian. J. Chem.*, 7, 639-646, 2014.

[63] Liton, M. A. K.; Bhowmick, A. C.; Ali, M. A.; 3D-QSAR MIFs Studies on 3,5-substituted-1,4,2-dioxazoles Derivatives Using Open3DQSAR Tools, *Univers. J. Chem.* 1(2), 71-76, 2013.