Journal of Computational Methods in Molecular Design, 2015, 5 (3):106-119

Scholars Research Library
(http://scholarsresearchlibrary.com/archive.html)

ISSN : 2231- 3176
CODEN (USA): JCMMDA

Quantitative structure-toxicity relationship study of some polychlorinated aromatic compounds using molecular descriptors

Hassan Samuel[*1], Adamu Uzairu[1], Paul Mamza[1] and Okunola Oluwole Joshua[2]

[1]Department of Chemistry, Faculty of Science, Ahmadu Bello University Zaria, Nigeria
[2]Department of Applied Chemistry, Faculty of Science, Federal University Dutsin-ma Katsina State, Nigeria
[*]Corresponding E-mail: hassansamuel84@gmail.com

ABSTRACT

Polychlorinated aromatic compounds represent a large group of industrial and byproduct compounds which are resistant to chemical and biological degradation and highly toxic. QSAR analysis was performed on 74 molecules of three classes of polychlorinated aromatic compounds (polychlorinated dibenzo-p-dioxin (PCDDs), polychlorinated dibenzofuran (PCDF) and polychlorinated biphenyl (PCB)). A large number of about 1700 molecular descriptors was obtained from DFT (B3LYP/6-311+G) level of calculation for each molecule and used in Genetic function algorithm (GFA) approach to generate 5 models, out of which the one with the highest statistical significance (Model-1: R^2 = 0.9673, $R^2_{adjusted}$ = 0.9592, R^2_{cv} = 0.9402, R^2_{pred}. = 0.7209, F-test = 118.48, LOF = 0.4377) was selected as the best. From the model generated, it seems to be very clear that polarizability, SP-7, ETA_Epsilon_5, GRAVH_3, and MOMI-R contribute positively to the toxicity of these compounds while MaxHBint5, ETA_dApha_B, ETA_Epsinlon-2, n5Ring and GRAV_2 contribute negatively. This validated model brings important insight to aid the prediction and identification of other toxic polychlorinated aromatic compounds.*

Keywords: QSAR, Genetic Function Algorithm, Molecular descriptors, Polychlorinated aromatic compounds, toxicit of polychlorinated compounds.

INTRODUCTION

In recent years, due to the increasing impact the pharma and food chemistry in special have on the human and environment life, the scientific and economical interest forced the international communities (OECD- Organization of economic cooperation and development, EUC-European Commissions, just to name a few) to adopt memorandums regulating the design and use of chemicals towards lower toxicity and higher biodegrability [1]. Polychlorinated dibenzofurans (PCDFs), polyhalogenateddibenzo-p-dioxins (PHDDs) and polychlorinated biphenyls (PCBs) are chemicals of concern because of their elevated concentrations in adipose and hepatic tissues and their persistence in an individual for extended lengths of time. With heavier congeners, it may stay with an individual for decades because they are resistant to metabolic, thermal and environmental breakdown. Polychlorinated aromatic compounds are not commercially produced but are formed as trace amounts of undesired impurities in the manufacture of other chemicals [2-3].

The chemical/industrial sources of these chemicals include the manufacture of chlorinated compounds such as phenoxy herbicides, chlorinated benzenes, chlorinated aliphatic compounds, chlorinated catalysts and halogenated diphenyl ethers, PCBs, the pulp and paper industry, and dry cleaning distillation residues [4]. They could also be produced when organic compounds containing chlorine are burned and a series of chemical reactions take place under specific conditions [5]. These combustion sources include incinerators for municipal solid waste and hazardous waste, steelworks, metal refinery factories, power stations, coal and oil industries, sintering plants, cement, lime, glass and brick production, and recycling plants [6,7].The use and disposal of these compounds can cause the release of dioxins into the environment.

Polychlorinated aromatic compounds are considered as persistent and widespread environmental contaminants with high hydrophobicity, which can cause a great diversity of biological effects including hepatotoxicity, endocrine effects, immunotoxicity, body weight loss, teratogenicity, carcinogenicity and the induction of diverse enzymes such a aryl hydrocarbon hydroxylase (AHH) and 7-ethoxyresorufinOdeethylase (EROD) in various organisms [8,9].

Food is the major source for human exposure to PCBs and dioxins, especially fatty foods: dairy products (butter, cheese, fatty milk), meat, egg, and fish. Food of animal origin accounts for 95 % of total exposure. The current average body burden of dioxins is about 5–50 ng/kg (as WHO TEq in fat; pg/g = ng/kg) or 100–1000 ng (WHO-TEq) per person which is close to the lowest concentrations possibly causing health effects. Some subgroups within the society (e.g., nursing babies and people consuming plenty of fish) may be exposed to higher than average amounts of these compounds and are thus at greater risk. Dioxin concentrations have been screened in five WHO international studies, and in Central Europe the concentrations have decreased in breast milk from about 40 ng/kg (as TEq in milk fat) in 1987 to below 10 ng/kg in 2006. PCBs have decreased at about the same rate. The decrease in environmental concentrations is due to cessation of PCB use and improved incineration technology [10]. Due to the problems of assessing the fate and toxicity of large number of chemicals, alternative method has been sought to classical in vivo animal texting. In the area of computer – aided toxicity prediction, quantitative structure activity relationship (QSAR) have been seen as an attractive method for toxicity and fate assessment [11].The study of the quantitative relationship between toxicity/activity and molecular structure (QSTR/QSAR) is an important area of research in computational chemistry and has been widely used in the prediction of toxicity and other biological activities of organic compounds [12, 13].

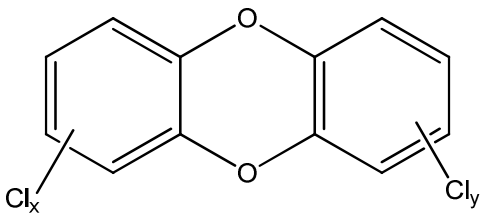
In this study, genetic function approximation (GFA) which is a statistical modeling algorithm that builds functional models of experimental data. Since its inception, several applications of this algorithm in the area of quantitative structure–activity relationship modeling have been reported [14]. The genetic function approximation (G FA) algorithm is a genetic algorithm (GA) derived from the previously reported G/SPLINES algorithm and has been recently applied to the generation of QSAR models [15-17]. The main purpose of this work is to find out how accurate QSAR analysis (using Material studio 7.0 software and the statistical tool Genetic functional algorithm) predicted the toxicity of polychlorinated aromatic compounds, and also to find out the descriptors responsible for producing such toxicity other than the once reported by [18, 19].

MATERIALS AND METHODS

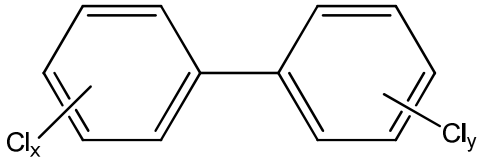
QSAR METHODOLOGY

Chemical data and biological activity

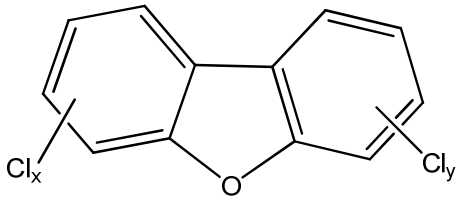
A data set of 74 molecules (25 PCDDs, 34 PCDFs and 15 PCBs) has been taken from the literature [20-22]. The toxicities of the compounds expressed in EC_{50} have been converted to log EC_{50}. The structures of the compounds were drawn using Chemsketch software. The general structural formulae of the three series are shown in Fig.2. IUPAC names and toxicity data of all the compounds are listed in Table-1.



Polychlorinated dibenzo-P-dioxins (PCDDs)



Polychlorinated biphenyls (PCBs)



Polychlorinated dibenzofurans (PCDFs)

Fig.2: General structural formulae for the compounds (PCDDs, PCDFs and PCBs)

Table 1: Experimental Biological Activities in pEC_{50} of the compounds, (PCDDs, PCDFs and PCBs)

S/N	IUPAC NAME	pEC_{50}
1	2,3,7,8 Tetrachlorodibenzo-p-dioxin	8.00
2	1,2,3,7,8 pentachlorodibenzo-p-dioxin	7.10
3	2,3,6,7 Tetrachlorodibenzo-p-dioxin	6.80
4	2,3,6 Trichlorodibenzo-p-dioxin	6.66
5	1,2,3,4,7,8Hexachlorodibenzo-p-dioxin	6.55
6	1,3,7,8-Tetrachlorodibenzo-p-dioxin	6.10
7	1,2,4,7,8-pentachlorodibenzo--dioxin	5.96
8	1,2,3,4-Tetrachlorodibenzo-p-dioxin	5.89
9	2,3,7-Trichlorodibenzo-p-dioxin	7.15
10	1,2,3,4,7-pentachlorodibenzo-p-dioxin	5.19
11	1,2,4-Trichlorodibenzo-p-dioxin	4.89
12	2,8-dichlorodibenzo-p-dioxine	5.49
13	1,2,3,4,6,7,8,9-Octachlorodibenzo-o-dioxin	5.00
14	1-chlorodibenzo-p-dioxin	4.00
15	2,3,7,8-Tetrabromodibenzo-p-dioxin	8.82
16	2,3- Dibromo 7,8-chlorodibenzo-p-dioxin	8.83
17	2,8- Dibromo -3,7-dichlorodibenzo-p-dioxin	9.35
18	2,Bromo-3,7,8-trichlorodibenzo-p-dioxin	7.94
19	1,3,7,9-Tetrabromodibenzo-p-dioxin	7.03
20	1,3,7,8-Tetrabromodibenzo-p-dioxin	8.70
21	1,2,4,7,8-Pentabromodibenzo-p-dioxin	7.77
22	1,2,3,7,8-Pentabromodibenzo-p-dioxin	8.18
23	2,3,7-Tribromodibenzo-p-dioxin	8.93
24	2,7-Dibromodibenzo-p-dioxin	7.81
25	2-Bromodibenzo-p-dioxin	6.53
26	1-chlorodibenzofuran	4.53
27	2-chlorodibenzofuran	3.55
28	3-chlorodibenzofuran	4.38
29	4-chlorodibenzofuran	3.00
30	2,3-Dichlorodibenzofuran	5.36
31	2,6-dichlorodibenzofuran	3.61
32	2,8-Dichlorodibenzofuran	5.05
33	1,3,6-Trichlorodibenzofuran	5.36
34	1,3,8-Trichlorodibenzofuran	4.07
35	2,3,4-Trichlorodibenzofuran	4.72
36	2,3,7-Trichlorodibenzofuran	7.10
37	2,3,8-Trichlorodibenzofuran	6.00
38	2,6,7 Trichlorodibenzofuran	6.35
39	2,3,4,6 Tetrachlorodibenzofuran	6.46
40	2,3,4,8-Tetrachlorodibenzofuran	6.70
41	2,3,7,8-Tetrachlorodibenzofuran	7.39
42	1,2,4,8-Tetrachlorodibenzofuran	5.00
43	1,2,4,7,9-Pentachlorodibenzofuran	4.70
44	1,2,3,7,8-Pentachlorodibenzofuran	7.13

45	1,2,4,7,8-Pentachlorodibenzofuran	5.89
46	2,3,4,7,8-Pentachlorodibenzofuran	7.82
47	1,2,3,4,7,8-Hexachlorobenzofuran	6.64
48	1,2,3,6,7,8-Hexachlorobenzofuran	6.57
49	2,3,4,67,8-Hexachlorobenzofuran	7.33
50	1,2,4,6,7,9-Hexachlorodibenzofaran	5.08
51	2,3,6,8-Tetrachlorodibenzofuran	6.66
52	1,2,3,6-Tetrachlorodibenzofuran	6.46
53	1,2,3,7-Tetrachlorodibenzofuran	6.96
54	1,3,4,7,8-Pentachlorodibenzofuran	6.70
55	2,3,4,7,9-Pentachlorodibenzofuran	6.70
56	1,2,3,7,9-Pentachlorodibenzofuran	6.40
57	Dibenzofuran	3.00
58	2,3,4,7-Tetrachlorobiphenyl	7.60
59	1,2,4,6,8-Pentachlorobiphenyl	5.51
60	2,3,4,4'-Tetrachlorobiphenyl	4.94
61	3,3',4,4'-Tetrachlorobiphenyl	6.15
62	3,4,4',5-Tetrachlorobiphenyl	4.55
63	2',3,4,4',5-Pentachlorobiphenyl	4.85
64	2,3,3',4,4'-Pentachlorobiphenyl	5.37
65	2,3',4,4',5-Pentachlorobiphenyl	5.04
66	2,3,4,4',5-Pentachlorobiphenyl	5.39
67	3,3'4,4'5-Pentachlorobiphenyl	6.92
68	2,2'4,4',5,5'-Hexachlorobiphenyl	4.26
69	2,3,3',4,4',5-Hexachlorobiphenyl	5.15
70	2,3',4,4',5,5'-Hexachlorobiphenyl	4.80
71	2,3,3',4,4',5'-Hexachlorobiphenyl	5.30
72	2,2',4,4'-Tetrachlorobiphenyl	
73	2,3,4,5-Tetrachlorobiphenyl	3.85
74	2,3',4,4',5',6-Hexachlorobiphenyl	4.00

Geometry optimization and calculation of molecular descriptors

Complete geometry optimization of the 74 molecules of polychlorinated aromatic compounds was performed using Spatan "14"1.1.2 software. Density functional theory (DFT) was used as the level of theory, 6-31G* as the basis set and MMFF as Geometry. The second step in developing the model was the numerical description of molecular structures by defining descriptors. These descriptors were responsible for encoding important features of the structures. A large number of about 1700 molecular descriptors (0D, 1D, 2D and 3D) were calculated. Quantum chemical descriptors and some of the constitutional descriptors were calculated using Spatan "14"1.1.2 software while topological descriptors and geometrical descriptors were calculated using PaDel-Descriptor 2.18 software.

Statistical method/correlation analysis

Because of the large number of the descriptors calculated, a stepwise multiple linear regression procedure on the forward-selection and backward-elimination method was used for the selection and elimination of the descriptors. From the square correlation matrix obtained, pairs of variables that falls within the range $0.35 \leq r \leq 0.9$ were selected and used by the statistical tool to generate the models.

Development of QSAR models

Genetic Function Algorithm (GFA).

In this work, all the models were developed using genetic function approximation (GFA) technique. The genetic function approximation algorithm was initially anticipated by: (1) Holland's genetic algorithm and (2) Friedman's multivariate adaptive regression splines (MARS) algorithm. In this algorithm, an individual or model is represented as one dimensional string of bits. A distinctive feature of GFA is that it produces a population of several models instead of generating a single model, as do most other statistical methods. Genetic algorithm makes superior models to those developed using stepwise regression techniques because it selects the basis function genetically [23].

The GFA algorithm approach has several important advantages over other techniques: (1) it builds multiple models rather than a single model. (2) It automatically selects which features are to be used in the models. (3) It is better at discovering combinations of features that take advantage of correlations between multiple features. (4) It incorporates Friedman's lack-of fit (LOF) error measure, which estimates the most appropriate number of features, resists over fitting, and allows control over the smoothness of fit. (5) It can use a large variety of equation term types in construction of its models, e.g., splines, step functions, high order polynomials. (6) It provides, through study of

the evolving models, additional information not available from standard regression analysis, such as the preferred model length and useful partitions of the data set [24-25].

QSAR analysis in computational research is responsible for the generation of models to correlate biological activity and physicochemical properties of a series of compounds. The underlying assumption is that the variations of biological activity within a series can be correlated with changes in measured or computed molecular features of the molecules. In the present study, QSAR model generation was performed by GFA technique. The application of the GFA algorithm allows the construction of high-quality predictive models and makes available additional information not provided by standard regression techniques, even for data sets with many features [27-27]. GFA was performed using 100,000 crossovers, smoothness value of 2.0 and other default settings for each combination. The number of terms in the equation was fixed to 10 including constant in the training set. The set of equations generated were evaluated on the basis of some statistical parameters.

Statistical/Validation Parameter

Lack of fit (LOF)

A "fitness function" or lack of fit (LOF) was used to estimate the quality of the model, so that best model receives the best fitness score. The error measurement term is determine by equation-1

$$LOF = \frac{LSE}{(1-\frac{c+d*p}{M})^2} \quad \dots\dots\dots (1)$$

where 'c' is the number of basic functions (other than constant term); 'd' is smoothing parameter (adjustable by the user); 'M' is the number of samples in the training set; LSE is least squares error and 'p' is the total numbers of the features contained in all basis functions [28].

Coefficient of multiple determination (R^2)

To assess the goodness-of-fit, the coefficient of multiple determination is used. R^2 estimates the proportion of the variation in the response that is explained by the predictor.

$$R^2 = 1 - \frac{\sum_{i=1}^{I}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{I}(y_i-\bar{y})} \dots\dots\dots (2)$$

Where yi is the observed dependent variable, $\bar{y}$ the mean value of the dependent variable and $\hat{y}$ the calculated dependent variable. If there is no linear relationship between the dependent variable and the descriptors then R^2 = 0.00; if there is a perfect fit then R^2 = 1.00. R^2 values higher than 0.5 indicates that the explained variance by the model is higher the unexplained one.

Adjusted R^2 (R^2_{adj})

The value of R^2 can generally be increased by adding additional predictor variables to the model, even if the added variable does not contribute to reduce the unexplained variance of the dependent variable. It follows R^2 should be used with caution. This can be avoided by using another statistical parameter the so-called adjusted R^2 (R^2_{adj}).

$$R^2_{adj} = 1 - (1 - R^2)(\frac{I-1}{I-K}) \dots\dots\dots (3)$$

R^2_{adj} is interpreted similarly to the R^2 value, except that it takes into consideration the number of degrees of freedom. The value of R^2_{adj} decreases if an added variable to the equation does not reduce the unexplained variable.

Standard error of estimate (SEE)

$$\text{SEE} = \sqrt{\frac{\sum_{i=1}^{I}(y_i-\hat{y}_i)^2}{(I-(K+1))}} \dots\dots\dots (4)$$

The smaller the value of SEE is, the higher the reliability of the prediction. However, it is not recommended to have the standard error of estimate smaller than the experimental error of the biological data, because it is an indication of over fitted model.

F-value
The F-value is determined using equation-6

$$F = \frac{\sum_{i=0}^{I}(y_i-\bar{y})^2/(K-1)}{\sum_{i=1}^{I}(y_i-\hat{y}_i)^2/(I-K)} \dots \dots \dots (6)$$

The higher the F-value, the greater the probability that the equation is significant [29].

Validation Parameters
Cross-validation squared correlation coefficient R^2 (R^2_{cv})

Cross-validation squared correlation coefficient R^2 (LOO-Q^2) is calculated according to the formula:

$$Q^2 = 1 - \frac{\sum(Y_{pred}-Y)^2}{\sum(Y-\bar{Y})^2} \dots \dots \dots (7)$$

In Eq. (2), $Y_{pred}$and Y indicate predicted and observed activity values respectively and $\bar{Y}$ indicate mean activity value. A model is considered acceptable when the value of Q^2 exceeds 0.5. [30].

In the case of this research, external validation techniques (LMO-Leave Many Out) was applied in which the 23 compounds of the test set were used for the external validation and the predicted R^2 for the validation was calculated using equation-2.

Predicted R^2 (R^2_{pred})
The predictive R^2 was calculated based on only molecules not included in the training set (test set). Models are generated based on training set compounds and predictive capacity of the models was judged based on the predictive R^2 (R^2_{pred}) value which was calculated using eqn-5.

$$R^2_{pred} = 1 - \frac{\sum(Y_{pred(test)}-Y_{test})^2}{\sum(Y_{(test)}-\bar{Y}_{training})^2} \dots \dots \dots \dots (5).$$

In Eq. (5), $Y_{pred(test)}$ and Y_{test} indicate predicted and observed activity values respectively of the test set compounds and $\bar{Y}_{training}$ indicates mean activity of the training set. For a QSAR model, the value of R^2_{pred} should be more than 0.5. All the statistical parameters calculated, agree with the criteria reported in Table-2

Table-2: criteria for selection of good model

S/N	CRITERIA FOR SELECTION OF MODEL
1	N = number of molecules (> 20 molecules)
2	K= number of descriptors in a model (statistically N/5 descriptor in a model)
3	df = degree of freedom (N-K-1) (higher is better).
4	R^2 =coefficient of determination (> 0.7)
5	R^2cv = cross-validation square correlation (> 0.5)
6	R^2_{adj} = adjusted squared correlation coefficient (> 0.5)
7	R^2_{pred} = predicted coefficient of determination (> 0.5)
8	SEE = standard error of estimate (smaller is better)
9	F-test = F-test for statistical significance of the model (higher is better, for some set of descriptors and compounds)

RESULTS AND DISCUSSION

Generation of models
Genetic function approximation was used to performed QSAR regression on 74 molecules of polychlorinated aromatic compounds using pEC_{50} as dependent variable and calculated molecular descriptors as independent variables described by the equations in Table-2. 51 molecules were used as training set to generate the 5 models which are presented in Table-3 .The remaining 23 compounds were used as test set for external prediction and the predicted toxicities are presented in Table-4. Model-1 was selected as the best on the basis of its statistical parameters.

Table-3. 5 generated models by GFA

No	Equation	Definition
1	Y = 0.317123855 * X26 + 2.673194291 * X94 - 1.361102037 * X144 - 1.453070371 * X165 - 96.631409262 * X167 + 69.872692578 * X170 - 7.877708429 * X228 - 21.439713345 * X277 + 124.942300592 * X281 + 1.835776438 * X293 - 631.436918720	X26 : AA : Polarizability X94 : CQ : SP-7 X144 : EO : maxHBint5 X165 : FJ : ETA_dAlpha_B X167 : FL : ETA_Epsilon_2 X170 : FO : ETA_Epsilon_5 X228 : HU : n5Ring X277 : JR : GRAV-2 X281 : JV : GRAVH-3 X293 : KH : MOMI-R
2	Y = 0.337543195 * X26 + 1.986019421 * X94 - 1.841849479 * X144 - 2.591615577 * X165 - 105.290996384 * X167 + 150.560145568 * X175 - 8.806840418 * X228 - 22.985643463 * X277 + 135.442609778 * X281 + 0.491497214 * X345 - 657.102081723	X26 : AA : Polarizability X94 : CQ : SP-7 X144 : EO : maxHBint5 X165 : FJ : ETA_dAlpha_B X167 : FL : ETA_Epsilon_2 X175 : FT : ETA_Psi_1 X228 : HU : n5Ring X277 : JR : GRAV-2 X281 : JV : GRAVH-3 X345 : MH : WA.eneg
3	Y = 0.312654553 * X26 + 2.629675114 * X94 - 1.401918491 * X144 - 1.502764637 * X165 - 116.167215005 * X167 + 66.429804381 * X175 - 7.847220091 * X228 - 21.154896743 * X277 + 125.332874970 * X281 + 1.696580396 * X293 - 602.797141503	X26 : AA : Polarizability X94 : CQ : SP-7 X144 : EO : maxHBint5 X165 : FJ : ETA_dAlpha_B X167 : FL : ETA_Epsilon_2 X175 : FT : ETA_Psi_1 X228 : HU : n5Ring X277 : JR : GRAV-2 X281 : JV : GRAVH-3 X293 : KH : MOMI-R
4	Y = 0.336800494 * X26 + 2.942208255 * X94 - 0.640140511 * X108 - 1.579876967 * X144 - 1.116608578 * X165 + 81.907227030 * X175 - 8.260828699 * X228 - 14.048865479 * X277 + 82.359135734 * X281 + 2.468795662 * X293 - 413.254169614	X26 : AA : Polarizability X94 : CQ : SP-7 X108 : DE : nHBa X144 : EO : maxHBint5 X165 : FJ : ETA_dAlpha_B X175 : FT : ETA_Psi_1 X228 : HU : n5Ring X277 : JR : GRAV-2 X281 : JV : GRAVH-3 X293 : KH : MOMI-R
5	Y = 0.334445535 * X26 - 1.292560192 * X35 + 2.861556067 * X94 - 1.304920784 * X144 - 1.234903575 * X165 - 136.309059489 * X177 - 8.316100485 * X228 - 17.174253798 * X277 + 96.839460363 * X281 + 2.237066146 * X293 - 454.420645683	X26 : AA : Polarizability X35 : AJ : apol X94 : CQ : SP-7 X144 : EO : maxHBint5 X165 : FJ : ETA_dAlpha_B X177 : FV : ETA_dPsi_B X228 : HU : n5Ring X277 : JR : GRAV-2 X281 : JV : GRAVH-3 X293 : KH : MOMI-R

Table-4: the best model selected

Model 1	Variables in the model
Y = 0.317123855 * X26 + 2.673194291 * X94 - 1.361102037 * X144 - 1.453070371 * X165 - 96.631409262 * X167 + 69.872692578 * X170 - 7.877708429 * X228 - 21.439713345 * X277 + 124.942300592 * X281 + 1.835776438 * X293 - 631.436918720	X26 : Polarizability X94 : SP-7 X144 : maxHBint5 X165 : ETA_dAlpha_B X167 : ETA_Epsilon_2 X170 : ETA_Epsilon_5 X228 : n5Ring X277 : GRAV-2 X281 : GRAVH-3 X293 : MOMI-R

Table-5.Chemical names along with the observed and calculated toxicity values in pEC_{50} of the training set compounds

S/N	Chemical name	Actual values	Predicted values	Residual values
1	1,2,3,7,8 pentachlorodibenzo-p-dioxin	7.10000000	6.89480200	0.205198
2	2,3,6,7 Tetrachlorodibenzo-p-dioxin	6.80000000	6.57743100	0.222569
3	2,3,6 Trichlorodibenzo-p-dioxin	6.66000000	6.76089900	-0.100899
4	1,2,3,4,7,8Hexachlorodibenzo-p-dioxin	6.55000000	6.44193700	0.108063
5	1,3,7,8 Tetrachlorodibenzo-p-dioxin	6.10000000	6.11652600	-0.016526
6	1,2,4,7,8 pentachlorodibenzo--dioxin	5.96000000	6.19664600	-0.236646
7	1,2,3,4 Tetrachlorodibenzo-p-dioxin	5.89000000	6.28801600	-0.398016
8	2,3,7 Trichlorodibenzo-p-dioxin	7.15000000	6.86770100	0.282299
9	1,2,4 Trichlorodibenzo-p-dioxin	4.89000000	4.88714900	0.002851
10	2,8-dichlorodibenzo-p-dioxin	5.49000000	5.15069900	0.339301
11	1,2,3,4,6,7,8,9-Octachlorodibenzo-o-dioxin	5.00000000	5.14484700	-0.144847
12	1-chlorodibenzo-p-dioxin	4.00000000	4.24100500	-0.241005
13	2,3,7,8 Tetrabromodibenzo-p-dioxin	8.82000000	9.40408400	-0.584084
14	2,3-Dibromo 7,8-chlorodibenzo-p-dioxin	8.83000000	8.99649900	-0.166499
15	2,8- Dibromo -3,7-dichlorodibenzo-p-dioxin	9.35000000	9.06707700	0.282923
16	2-Bromo-3,7,8-trichlorodibenzo-p-dioxin	7.94000000	8.30822700	-0.368227
17	1,3,7,9-Tetrabromodibenzo-p-dioxin	7.03000000	7.26343800	-0.233438
18	1,3,7,8-Tetrabromodibenzo-p-dioxin	8.70000000	8.47020000	0.229800
19	1,2,4,7,8-Pentabromodibenzo-p-dioxin	7.77000000	7.42604000	0.343960
20	1,2,3,7,8-Pentabromodibenzo-p-dioxin	8.18000000	8.09153100	0.088469
21	2,7-Dibromodibenzo-p-dioxin	7.81000000	7.82283900	-0.012839
22	2-Bromodibenzo-p-dioxin	6.53000000	6.15431200	0.375688
23	1-chlorodibenzofuran	4.53000000	4.36064600	0.169354
24	2-chlorodibenzofuran	3.55000000	4.22833800	-0.678338
25	3-chlorodibenzofuran	4.38000000	4.10933400	0.270666
26	4-chlorodibenzofuran	3.00000000	3.01669200	-0.016692
27	2,3-Dichlorodibenzofuran	5.36000000	5.63232200	-0.272322
28	1,3,6-Trichlorodibenzofuran	5.36000000	5.30147400	0.058526
29	1,3,8-Trichlorodibenzofuran	4.07000000	4.56537300	-0.495373
30	2,3,7-Trichlorodibenzofuran	7.10000000	6.96629000	0.133710
31	2,3,4,8-Tetrachlorodibenzofuran	6.70000000	6.40437800	0.295622
32	2,3,7,8-Tetrachlorodibenzofuran	7.39000000	7.77433500	-0.384335
33	1,2,3,7,8-Pentachlorodibenzofuran	7.13000000	6.97671300	0.153287
34	1,2,4,7,8-Pentachlorodibenzofuran	5.89000000	5.69544900	0.194551
35	2,3,4,7,8-Pentachlorodibenzofuran	7.82000000	7.82693800	-0.006938
36	1,2,3,4,7,8-Hexachlorobenzofuran	6.64000000	6.83519100	-0.195191
37	1,2,3,6,7,8-Hexachlorobenzofuran	6.57000000	6.76563400	-0.195634
38	2,3,4,6,7,8-Hexachlorobenzofuran	7.33000000	7.51370700	-0.183707
39	2,3,6,8-Tetrachlorodibenzofuran	6.66000000	6.24406500	0.415935
40	1,2,3,7-Tetrachlorodibenzofuran	6.96000000	6.86559800	0.094402
41	2,3,4,7,9-Pentachlorodibenzofuran	6.70000000	6.37662100	0.323379
42	1,2,3,7,9-Pentachlorodibenzofuran	6.40000000	6.39655100	0.003449
43	Dibenzofuran	3.00000000	2.89483700	0.105163
44	2,3,4,7-Tetrachlorobiphenyl	7.60000000	7.48510400	0.114896
45	1,2,4,6,8-Pentachlorobiphenyl	5.51000000	5.31940200	0.190598
46	2,3,4,4'-Tetrachlorobiphenyl	4.94000000	4.94644900	-0.006449
47	3,3',4,4'-Tetrachlorobiphenyl	6.15000000	6.06549900	0.084501
48	3,3'4,4'5-Pentachlorobiphenyl	6.92000000	6.42820400	0.491796
49	2,3,3',4,4',5-Hexachlorobiphenyl	5.15000000	5.63058300	-0.480583
50	2,3,3',4,4',5'-Hexachlorobiphenyl	5.30000000	5.42050100	-0.120501
51	2,3,4,5-Tetrachlorobiphenyl	3.85000000	3.89186900	-0.041869

Table-6. Chemical names along with the observed and the calculated toxicities of the test set compounds expressed in logEC_{50}

S/N	Compound names	Actual values	Predicted values	Residual values
1	2,3,7,8 Tetrachlorodibenzo-p-dioxin	8.00	7.137742	0.862258
2	1,2,3,4,7 pentachlorodibenzo-p-dioxin	5.19	6.416795	-1.2468
3	2,3,7-Tribromodibenzo-p-dioxin	8.93	9.433407	-0.50341
4	2,6-dichlorodibenzofuran	3.61	5.33345	-1.72345
5	2,8-Dichlorodibenzofuran	5.05	5.814313	-0.76431
6	2,3,4-Trichlorodibenzofuran	4.72	6.341931	-1.62193
7	2,3,8-Trichlorodibenzofuran	6.00	6.078967	-0.07897
8	2,6,7 Trichlorodibenzofuran	6.35	5.891172	0.458828
9	2,3,4,6 Tetrachlorodibenzofuran	6.46	6.643481	-0.18348
10	1,2,4,8-Tetrachlorodibenzofuran	5.00	5.764051	-0.76405
11	1,2,4,7,9-Pentachlorodibenzofuran	4.7	4.070704	0.629296
12	1,2,4,6,7,9-Hexachlorodibenzofaran	5.08	3.868581	1.211419
13	1,2,3,6-Tetrachlorodibenzofuran	6.46	6.216	0.244
14	1,3,4,7,8-Pentachlorodibenzofuran	6.70	6.636922	0.063078
15	3,4,4',5-Tetrachlorobiphenyl	4.55	5.973791	-1.42379
16	2',3,4,4',5-Pentachlorobiphenyl	4.85	4.9348	-0.0848
17	2,3,3',4,4'-Pentachlorobiphenyl	5.37	5.668328	-0.29833
18	2,3',4,4',5-Pentachlorobiphenyl	5.04	5.392424	-0.35242
19	2,3,4,4',5-Pentachlorobiphenyl	5.39	5.43461	-0.04461
20	2,2'4,4',5,5'-Hexachlorobiphenyl	4.26	4.397758	-0.13776
21	2,3',4,4',5,5'-Hexachlorobiphenyl	4.80	4.397758	0.402242
22	2,2',4,4'-Tetrachlorobiphenyl	3.89	4.710143	-0.82014
23	2,3',4,4',5',6-Hexachlorobiphenyl	4.00	3.366028	0.633972

Table-7. Statistical/validation parameters of the generated models

Statistical parameters	Model 1	Model 2	Model 3	Model 4	Model 5
Friedman LOF	0.437664	0.439352	0.440601	0.443745	0.443884
R-squared	0.967342	0.967217	0.967123	0.966889	0.966878
Adjusted R-squared	0.959178	0.959021	0.958904	0.958611	0.958598
Cross validated R-squared	0.940154	0.923084	0.937807	0.933100	0.937382
Significant Regression	Yes	Yes	Yes	Yes	Yes
Significance-of-regression F-value	118.48328	118.012638	117.666633	116.80482	116.7668
Critical SOR F-value (95%)	2.080618	2.080618	2.080618	2.080618	2.080618
Replicate points	0	0	0	0	0
Computed experimental error	0.000000	0.000000	0.000000	0.000000	0.000000
Lack-of-fit points	40	40	40	40	40
Min expt. error for non-significant LOF (95%)	0.257926	0.258423	0.258790	0.259712	0.259752

Table-8: Univariate analysis of the toxicity data

Statistical parameter	values
Number of sample points	51
Range	6.35000000
Maximum	9.35000000
Minimum	3
Mean	6.28450980
Median	6.57000000
Variance	2.23024000
Standard deviation	1.50826000
Mean absolute deviation	1.21134000
Skewness	-0.26949300
Kurtosis	-0.48857800

Table-9 values of the descriptors used in the selected model

S/N	p	SP-7	minHBint5	ETA_dAlpha_B	ETA_Epsilon_2	ETA_Epsilon_5	n5Ring	GRAV-2	GRAVH-3	MOMI-R
1	61.63	7.185	0.8984	22.9790	0.5291	0.871	0.4444	42.242	12.218	6.9852
2	61.63	7.849	0.9207	27.0342	0.53885	0.871	0.4444	43.892	12.507	7.2141
3	59.39	8.003	0.9321	27.321	0.53885	0.7723	0.4444	43.893	12.506	7.0553
4	62.75	6.738	0.8647	19.581	0.51821	0.9203	0.375	40.525	11.916	6.4812
5	60.51	8.526	0.9543	31.143	0.54762	0.8216	0.4444	45.483	12.782	7.4407
6	61.63	6.983	0.9049	23.201	0.5291	0.871	0.4444	42.243	12.218	6.9739
7	59.4	7.734	0.9322	27.2260	0.53885	0.7722	0.4444	43.893	12.507	7.2520
8	60.53	6.600	0.8414	19.2939	0.51821	0.8216	0.4444	40.525	11.916	6.4875
9	59.38	7.604	0.8636	23.8245	0.5291	0.7723	0.375	42.242	12.218	6.4679
10	58.28	7.940	0.8973	27.4095	0.53885	0.7229	0.4444	43.893	12.507	7.0710
11	64.96	6.813	0.8413	19.9452	0.51821	1.0191	0.5	40.525	11.915	6.2940
12	57.13	6.015	0.8191	15.6297	0.50595	0.6736	0.5	38.732	11.597	6.2049
13	61.99	9.866	1.1305	39.4347	0.5628	0.7091	0.4444	48.507	13.301	7.6460
14	61.25	5.817	0.7918	12.3852	0.49206	0.7654	0.4286	36.851	11.259	5.4654
15	61.24	7.185	0.9407	14.8188	0.66667	0.7654	0.4444	47.829	13.252	7.9693
16	60.88	7.185	0.9119	18.8989	0.59788	0.7935	0.4444	45.122	12.756	7.4719
17	61.98	7.185	0.9154	18.899	0.59788	0.7091	0.4444	45.122	12.756	7.5033
18	61.82	7.185	0.9042	20.9390	0.56349	0.7091	0.4444	43.706	12.493	7.2480
19	63.51	6.781	0.9790	15.0493	0.66667	0.7303	0.5	47.829	13.251	7.9030
20	63.55	6.983	0.9493	14.9323	0.66667	0.7303	0.4444	47.829	13.252	7.9577
21	59.01	7.734	0.9857	16.8090	0.70175	0.6667	0.4444	50.553	13.725	8.2323
22	57.5	7.850	0.9708	16.7255	0.70175	0.6455	0.4444	50.552	13.724	8.242
23	56.4	6.600	0.8661	13.1738	0.62745	0.6358	0.4444	44.941	12.742	7.4339
24	56.42	6.015	0.8446	11.5452	0.58333	0.6358	0.4444	41.854	12.189	6.4238
25	56.42	5.818	0.7997	10.2368	0.53333	0.6358	0.4285	38.520	11.580	5.6197
26	50.4	5.504	0.9954	8.35188	0.5034	0.6358	0.3333	36.569	11.204	5.0549
27	57.55	5.433	0.9840	8.02229	0.5034	0.6875	0.4285	36.569	11.204	5.2679
28	58.66	5.427	0.9840	7.96089	0.5034	0.7392	0.4285	36.569	11.204	5.1256
29	58.7	5.570	0.9954	8.26993	0.5034	0.7392	0.3333	36.569	11.204	5.1663
30	58,7	6.019	1.0063	11.6458	0.51746	0.7392	0.4286	38.464	11.545	5.6225
31	59.81	5.913	1.0291	11.8161	0.51746	0.7909	0.4285	38.463	11.545	5.7123
32	59.82	5.776	1.0177	11.6102	0.51746	0.7909	0.4286	38.464	11.545	6.0089
33	60.93	6.228	1.1096	15.8458	0.52976	0.8426	0.4286	40.269	11.867	6.1781
34	60.94	6.091	1.0514	15.6334	0.52976	0.8426	0.5	40.269	11.867	6.1080
35	60.78	6.689	1.0399	15.7116	0.52976	0.8426	0.4285	40.269	11.867	5.9662
36	62.05	6.354	1.0634	15.2198	0.52976	0.8943	0.375	40.269	11.866	6.1495
37	62.04	6.361	1.0399	15.3223	0.52976	0.8943	0.5	40.269	11.866	6.2819
38	62.04	6.513	1.0513	15.5286	0.52976	0.8943	0.5	40.269	11.866	6.0816
39	59.81	7.168	1.1589	19.6480	0.54062	0.7909	0.4286	41.997	12.171	6.3935
40	59.81	7.031	1.0736	19.4259	0.54062	0.7909	0.5	41.996	12.172	6.5899
41	60.93	6.946	1.0970	19.0551	0.54062	0.8426	0.375	41.996	12.171	6.6103
42	60.93	6.853	1.0850	19.8096	0.54062	0.8426	0.4285	41.996	12.172	6.4965
43	55.27	7.174	1.1656	23.9134	0.55027	0.5841	0.5	43.656	12.462	6.6331
44	59.81	7.548	1.1307	23.3562	0.55027	0.7909	0.375	43.656	12.462	6.6939
45	60.95	7.438	1.1421	23.5761	0.55027	0.8426	0.5	43.656	12.462	6.8714
46	59.83	7.616	1.1307	23.1860	0.55027	0.8426	0.375	43.656	12.462	6.9490
47	59.97	8.230	1.1644	27.5404	0.5589	0.7156	0.375	45.255	12.740	7.0311
48	61.09	8.218	1.1878	27.5241	0.5589	0.7649	0.375	45.255	12.740	7.0098
49	62.16	8.287	1.2497	27.3540	0.5589	0.8143	0.375	45.255	12.740	7.1719
50	62.04	7.930	1.2611	28.0611	0.5589	0.8143	0.5	45.255	12.740	6.8331
51	59.83	6.750	1.1085	19.3050	0.54062	0.7155	0.5	41.996	12.172	6.8053

Table-10: definition of descriptors that were found in the 5 models

Descriptor	Definition
Polarizability	Determine the dynamical response of a bound system to external fields, and provide insight into a molecule's internal structure.
SP-7	Chi path descriptor with a simple path order 7
MaxHBint5	Maximum E-state descriptor of strength for potential hydrogen bonds of path length 5
ETA_dApha_B	Extended Topochemical Atomic descriptor which is defined as a measure of count of hydrogen bond acceptor atoms and/or polar surface area
ETA_Epsilon_2	Extended Topochemical Atomic descriptor which is defined as a measure of electronegative atom count 2
ETA_Epsilon_5	Extended Topochemical Atomic descriptor which is defined as a measure of electronegative atom count 5
n5Ring	Ring count descriptor which indicates 5 member rings
GRAV_2	Gravitational index descriptor which is defined as square root of gravitational index of heavy atom
GRAVH_3	Gravitational index descriptor which is defined as cube root of hydrogen-included gravitational index
MOMI-R	Moment of initia along the radius of gyration
ETA_psi_1	Measure of hydrogen bond propensity the molecules and/or polar surface area.
WA.eng	Non directional WHIM, weighted by Mulliken atomic electronegativites
apol	Sum of atomic polarizabilities (including implicit hydrogen)
nHBa	Electrotopological state atom type descriptor which is defined as count of E-state for hydrogen bond acceptors
ETA_dpsi_B	Measure of hydrogen bonding propensity of the molecules

Table-11: The definition of the descriptors used in model-1 and their regression coefficients

Descriptor notation	Definition	Regression coefficient
Polarizability (p)	Determine the dynamical response of a bound system to external fields, and provide insight into a molecule's internal structure.	0.317123855
SP-7	Chi path descriptor with a simple path order 7	2.673194291
MaxHBint5	Maximum E-state descriptor of strength for potential hydrogen bonds of path length 5	-1.36110204
ETA_dApha_B	Extended Topochemical Atomic descriptor which is defined as a measure of count of hydrogen bond acceptor atoms and/or polar surface area	-1.45307037
ETA_Epsilon_2	Extended Topochemical Atomic descriptor which is defined as a measure of electronegative atom count 2	-96.6314093
ETA_Epsilon_5	Extended Topochemical Atomic descriptor which is defined as a measure of electronegative atom count 5	69.8726926
n5Ring	Ring count descriptor which indicates 5 member rings	-7.87770843
GRAV_2	Gravitational index descriptor which is defined as square root of gravitational index of heavy atom	-21.4397133
GRAVH_3	Gravitational index descriptor which is defined as cube root of hydrogen-included gravitational index	124.9423006
MOMI-R	Moment of initia along the radius of gyration	1.835776438

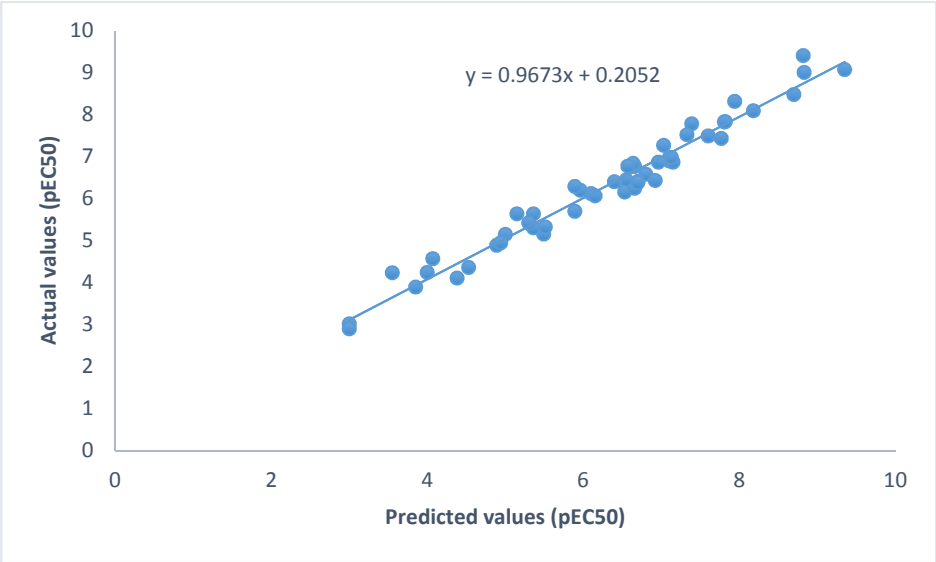


Fig. 2: Linear relationship of observed and predicted toxicities of data of the training set

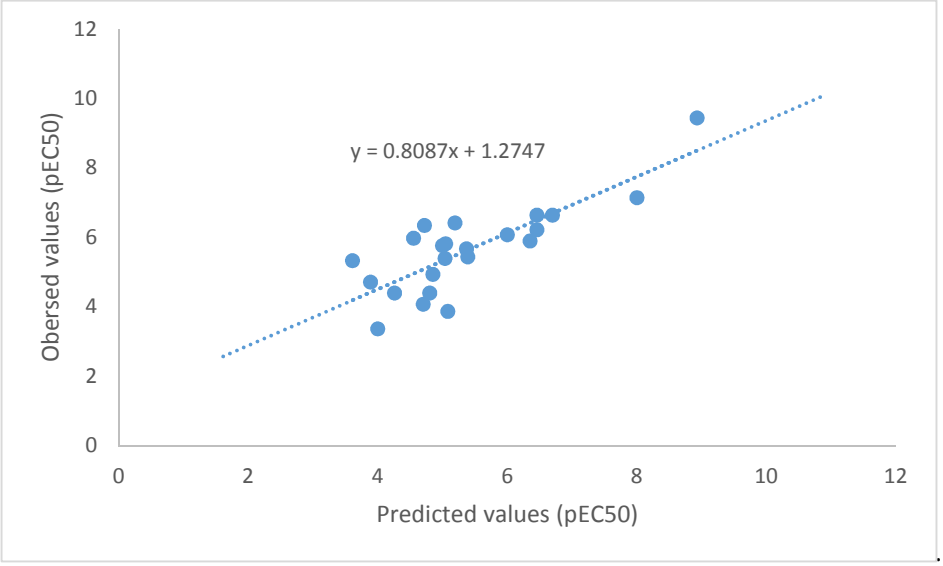


Fig. 3: Linear relationship of observed and predicted toxicities of data of the test set

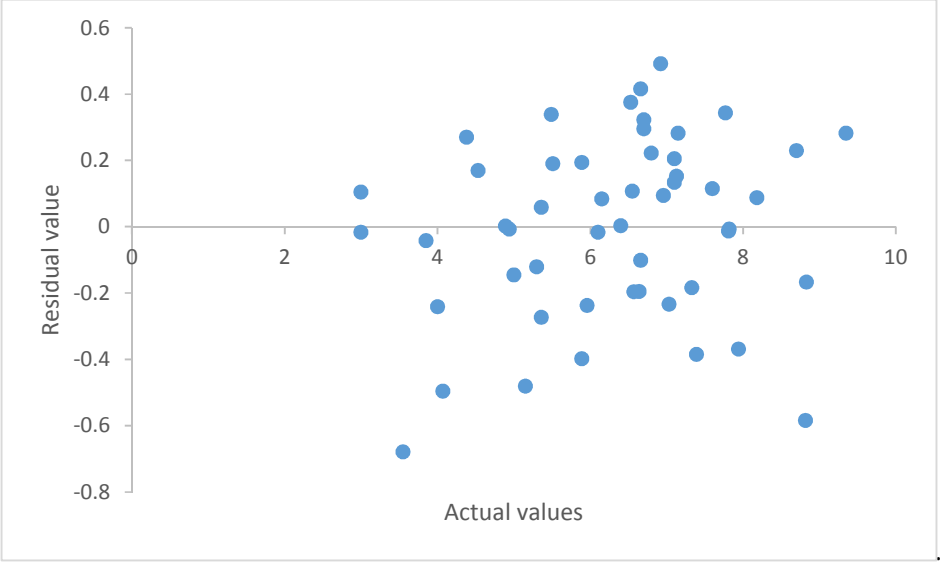


Fig.4. residual versus actual values

DISCUSSION

Table-3 shows 5 models generated by GFA using all the compounds of the training set. The toxicity (Y) was used as independent variables and the descriptors (Xi) as dependent variables. Each model contains10 descriptors as this agrees with the second criteria reported in Table-2. Among the 5 models generated, model-1 in Table-4 was selected as the best on the basis of the various statistical parameters reported in Table-7. Table-5 shows the predicted toxicities of the training set in pEC50 which has a good agreement with the experimental toxicities. Table-6 reported the predicted toxicities of test set for external validation with which the predicted (R^2_{pred}.) was calculated as 0.7209 and is in excellent agreement with criteria reported in Table-2

Table-7 shows the statistical/validation parameters of all the 5 models. The statistical quality of the models were determined by the validation parameters like LOF which is the measure of quality of fit, R^2, R^2_{adj}.,R^2_{cv}, F-test and the larger the value the better the model and the external validation parameter R^2_{pred}. The statistical parameters

model-1 appear to be statistically better than those of the other models. For model-1: LOF=0.4377, R_2 = 0.9673, $R^2_{adj.}$ = 0.9592, R_{2cv} = 0.9402, F-test = 118.48 and $R^2_{pred.}$ = 0.7209. All of these parameters are in a very good agreement with criteria reported in Table-2

Table-8 shows the statistical parameters of univariate analysis that describe the toxicity data. The most important parameters here are skewness and kurtosis. Skewness is the third moment of the distribution, which indicates the symmetry of distribution. As skewness is positive, the distribution of data value within the column is skewed toward positive values. For a symmetry distribution, the skewness is close to zero. Kurtosis is the fourth moment of the distribution which indicates the profile of the column of data relative to normal distribution [32].

Descriptor contribution

(Huifeng et al., 2011) reported that radius of gyration (RGyr) and Ist component accessibilities directional index/weight by atomic polarizabilities (E_{Ip}) are among the descriptors that are responsible for producing toxicties of polychlorinated aromatic compounds. Another previous work by (Nandan et al., 2013) shown that the descriptors Winner index (W), Balban index (J), polarizability (α) and index of refraction (η) have high responsibilities in producing toxicity of some polychlorinated aromatic compounds.

The present QSAR model study reveals that apart from the descriptors reported by (Huifeng et al., 2011; Nandan et al., 2013) which are responsible for producing toxicity of polychlorinated aromatic compounds, other descriptors were also found to be responsible for producing toxicity of polychlorinated aromatic compounds. Among these descriptors, Polarizability, SP-7, ETA_Epsilon_5, GRAVH_3, and MOMI-R which are used in model-1 contribute positively in producing toxicities of polychlorinated aromatic compounds. This indicate a positive impact on the toxicities of polychlorinated aromatic compounds, which means increasing the value of this descriptors produces higher toxicities of these compounds. In the other hand, the descriptors maxHBint5, ETA_dApha_B, ETA_Epsilon_2, n5Ring and GRAV_2 with negative coefficient used in model-1 contribute negatively, hence decreasing the values of these descriptors will provide higher toxicities of polychlorinated aromatic compounds. The interpretation of this model shows that each of these descriptors with positive coefficient is directly proportional to the toxicities of these molecules while each of those descriptors with negative coefficient is inversely proportional to the toxicities of the molecules [33]. Model-1 is presented in Table-4 and the descriptors used in model-1 are listed in Table-8. It is observed that both in this work and the once reported by (Huifeng et al., 2011; Nandan et al., 2013 polarizability and radius of gyration (RGyr) contribute in producing toxicity of polychlorinated aromatic compounds.

Figure-2 shows a plot describing the linear relationship between the experimental values in pEC_{50} and the calculated values. Most of the compounds of the training set are along the linear line of the plot. This indicates that the predicted values of pEC_{50} are in agreement with the experimental values. But for the test set, whose imprecise toxicity data were reported as shown in Figure-3, errors are higher than the training set. Figure 4 shows the plot of residuals versus experimental values of data set. The propagation of residuals on both sides of zero indicates that no systematic error exists in the development of GFA.

CONCLUSION

A genetic function approximation method was used to run the regression analysis and establish correlation's between different types of descriptors and experimental toxicity of three classes of polychlorinated aromatic (PCDDs, PCDFs and PCBs). QSAR models were developed and one of them was used to predict the toxicity efficiency of polychlorinated aromatic compounds. The prediction of toxicity efficiencies of these compounds matched with the experimental measurements. The developed models were found to be statistically significant as evidenced from their regression statistics.

Out of about 1700 molecular descriptors generated only these few were found to be the once responsible for producing toxicity of poly. These descriptors include: polalizability, Chi path descriptor with a simple path length order 7 (SP-7), Extended Topochemical Atomic descriptor (ETA_Epsilon_5) which is the measure of electronegative atom count 5, cubic root of hydrogen-included gravitational index (GRAVH-3) and moment of initia along the radius of gyration (MOMI-R). All the calculated molecular descriptors were aimed to encode some important information about the structural features of polychlorinated aromatic compounds which could influence the receptor binding affinity. Some of them provided good correlations and statistically reliable models.

Acknowledgement
Authors gratefully acknowledge the financial support from Aliyu Hassan. We also acknowledge the assistance of David Ebuka Arthur, Department of Chemistry, Ahmadu Bello University Zaria, Nigeria for providing the necessary facilities and support. Hassan Samuel shows deep sense of gratitude to his supervisor Prof. AdamuUzairu for introducing him to this fascinating aspect of chemistry.

REFERENCES

[1]European Chemicals Agency, **2011**, http://www.qsartoolbox.org
[2]M Athanasiadou; S N. Cuadra; G Marsh; A Bergman; K Jacobsson, *Envir. Health Perspec*. **2008**; 116: 400-408
[3] A Basler,*Envir. Sci. Pollut.Res.***1995**, 2, 117.
[4]R.Weber, C.Gaus, M.Tysklind, P.Johnston, M.Forter, H.Hollert Dioxin and POP-contaminated sites – contemporary and future relevance and challenges, *Envir. Sci. and Pollution Research*, **2008**; 15, 363-393.
[5]P.SKulkarni, J.GCrespo, C.AAfanso, *Envir. Int.*, **2008**, 34, 139-153.
[6] SFiedler, *Chemosphere*, **1996** 32, 55-64
[7] G JZheng; AOLeung; L P Jiao; MHWong, *Envir. Int.*,**2008**, 34, 1050-1061.
[8]JChovancova; AKocan S; Jursa,*Chemosphere*, **2005**; 61:1305–1311.
[9] J LDomingo,*Envir. Int.***2007**, 33, 397–405.
[10]T Jouko; V Terttu; T JTuomisto,Mannerheimintie, 166, Helsinki, Finland, **2011**.
[11] D JKnight; DBehenu,*Alter.Lab Anim.*,**2002**, 30, 7.
[12] A RKatritzky; RPentrukhin; DTatham; SBasak; E Benfenatim; M Karelaon; UMaran,*J Chem. Info. &Comput. Sci.*,**2001**; 40: 1–18.
[13]A R Katritzky; UMaran; V SLobanov; Mkrelson; YVhem,*Info. Comput.***2000**, 41, 345-390
[14]D Roggers; JDevillers (Ed.), "Genetic Algorithm in Molecular Modeling", Academic, London, **1996**, pp.87–107
[15]J H. Holland, Adaption in Natural and Artificial Systems, University of MichiganPress, Ann Arbor, MI, **1975.**
[16]D Rogers, Data analysis using G/SPLINES, in: Advances in Neural Processing Systems 4, Morgan Kaufmann, San Mateo, CA**, 1992.**
[17]T L Azeddine; A Rachid; H Majdouline; L Mohammed, *IJARCSSE***2014**, 4, 304-305.
[18] K D Paull; S H Friend; J N Weinstein,*J Chem. Inf. Comput. Sci.* 1998, 38, 189.
[19] K Nandan; K ranjan; Md B Ahmad; B Sah,*IJPSR*, **2013**, 4(7), 2691-2695.
[20] M Caprioara; M Diudea,*Indian J Chem*. **2003**, 42, 1368-1378.
[21] H Wu; F Li; J Zhao; X Liu; L Zhang, D,*Environ. Toxico. pharm.***2011**, 32, 478-485.
[22] A TBalaban; A Beteringhe, *ARKIVOC*, **2004,** (i) 163-182
[23] S P Niculescu; *Theochem*, **2003**, 622, 17-83.
[24]D Rogers; A J Hopfinger, *J. Chem. Inf. Comput. Sci.***1994**, 34, 854–866.
[25] V Patel; M T Chabria; P S Brahmkshatriya; B Mahajan; S Pael,*Pharmagene*, **2013,** 1, 234-282.
[26] C Putavy; J Devillers; D Domine; J Devillers (Ed), Genetic Algorithms in molecular modeling. Academic press London, **1996,** p 243.
[27] T Hou; X Xu, *Chem*. **2004**; 16: 35-41.
[28] R Kunal; P P Roy; S Paul; I Mitra, *Molecules*, **2009**, 14, 1660-1701.
[29]Sofie Van Damme. "Quantum Chemistry in QSAR, Quantum Chemical Descriptors, use, benefits and drawback". Thesis, department of inorganic and physical chemistry Faculty of Science, Universiteit Gent (**2009**), p 39.
[30]SWold, L Eriksson. In Chemometrics Methods in Molecular Desisign; van de Waterbeemd, H., Ed VCH, Weinheim, Germany, **1995**, pp. 309-318.
[31] A KPathak; A BMundada; AShrivastava, *Pharmacia*,**2011**, 1, P, 57.
[32]K F Khaled,*Corrosion sci.***2011**, 53, 3457-3465.
[33]R Navin; K J Sanmati,*ChemTech*, **2012**; 4, 1350-1360