MOLECULAR DOCKING AND QUANTITATIVE STRUCTURE-ACTIVITY RELATIONSHIP (QSAR) MODELS FOR ESTIMATING ARYL HYDROCARBON RECEPTOR BINDING AFFINITY OF SOME POLYCHLORINATED AROMATIC COMPOUNDS

Supported by tetfund TERTIARY EDUCATION TRUST FUND

Stephen Ejeh*, AdamuUzairu and Gideon A. Shallangwa
Department of Chemistry, Ahmadu Bello University, Zaria, Nigeria
*Correspondingauthor:ejehstephen@gmail.com

Received: June 13, 2017 **Accepted:** September 19, 2017

Abstract: In order to perform the screening of new potential pollution points and to estimate their impact on the environment, a molecular docking and quantitative structure-activity relationship (QSAR) model of some polychlorinated aromatic compounds was developed for further understanding of the mechanism of toxicity. From molecular docking, hydrogen-bonding, hydrophobic, Van der Waals, Pi-sigma, Amide-Pi stacked, Alkyl, Pi-Alkyl and $\pi - \pi$ interactions were observed to be characteristic interactions between compounds and aryl hydrocarbon receptor (AhR). Based on the mechanism of interactions, an optimum 3D-QSAR model with good robustness ($R^2 = 0.930$)and predictability ($R^2_{pred} = 0.824$) was developed by Genetic function approximation (GFA). Additionally, the developed QSAR model indicated that the distribution of charge, the distant intramolecular, molecular size, shape profiles, polarizability and electropological states of compounds were related to the binding affinities to AhR. Hence, the model can be used for the screening of ligands for AhR binding activity.

Keywords: Aryl hydrocarbon receptor (AhR), genetic function approximation, QSAR

Introduction

Several members of the broad class of polyhalogenated aromatic compounds (PHAs), including polychlorinated dibenzo-p-dioxins (PCDDs), dibenzofurans (PCDFs), and biphenyls (PCBs), produce characteristic toxicity syndromes (Hansch*et al.*, 2002). The chemical mechanisms for the toxicity are complicated, and most remain poorly understood. They are classed as persistent organic pollutants (POPs) under the Stockholm Convention (Márcia, 2004). They are a large class of ubiquitous organic pollutants and have received great attention because of their carcinogenic, teratogenic and mutagenic properties (Gallegos *et al*., 2001; Kumar *et al*., 2001; Borosky&Laali, 2005; Xue&Warshawsky, 2005). Being hydrophobic and lipophilic, these compounds accumulate in the marine environment in sediments and lipid-rich tissue of marine organisms, making these matrices preferred media for environmental monitoring (Gerhard & Rainer, 2012).Previous studies have demonstrated that several toxic and biochemical effects caused by dioxin-like chemicals are mediated through a particular protein complex known as the aryl hydrocarbon receptor AhR (Landers &Bunce, 1991; Lucier*et al.*, 1993; Nebert*et al*., 1993; Hestermann*et al.*, 2000). Thus, binding to AhR is a key step for contaminants exhibiting their toxicity (Hilscherova*et al.*, 2000). Hence, AhR activated by the ligands play a key role in adverse effects (Ohura *et al*., 2010).

The aryl hydrocarbon receptor (Ah receptor or AhR) is a ligand-activated transcription factor belonging to the basic-helix-loop-helix (bHLH) PAS family (Schmidt & Bradfield, 1996; Burbach*et al.*, 1992; Kewley*et al*., 2004; Whitlock, 1993; Huang *et al*., 1993). It is a nuclear receptor, located in the cytoplasm and exists as one component of the complex (Chen and Perdew, 1994); the other components being two molecules of heat shock protein (hsp90), an X-associated protein and a co-chaperone protein (Perdew, 1988). The AhR protein is a cytosolic transcription factor that is normally inactive, being attached to several co-chaperones. Upon binding to some natural ligands (e. g. bilirubin, prostaglandin G) or synthetic chemicals (such as polychloro compounds or aromatic hydrocarbons), the chaperones dissociate allowing the translocation of AhR into the nucleus leading to changes in gene transcription, and resulting in the induction of metabolizing enzymes that cause the production of metabolites which should be more easily excreted but in this case are more toxic (Martin *et al*., 2016).

A quantitative risk assessment becomes increasingly important in the modern society and is slowly incorporated into legislation of different countries. For instance, the European Union (EU) has introduced the Registration, Evaluation and Authorization of Chemicals (REACH) program for assessment of human and environmental risk of all chemicals that are produced or imported in the amount greater than 1 ton per year. It is clear that if such a risk assessment is performed purely experimentally, it would require a huge amount of resources as well as time. Therefore, the introduced REACH by European Union, encouraged the use of QSAR modeling and other alternatives especially for the risk assessment of chemicals that are produced or imported in smaller quantities (Martin *et al.*, 2016).

In this study, the three-dimension crystal structure of AhR was obtained by homologous modeling. Molecular docking was performed to define a model for the further understanding of the binding interactions between ligands and receptor interactions. Based on the mechanism of interaction, an optimal QSAR model of AhR binding affinity of PCBs, PCDDs and PCDFs was developed based on the experimental data taken from Martin *et al.* (2016) and Huifeng*et al.* (2011). From the developed QSAR model, critical molecular structural features related to the AhR binding affinity of PCBs, PCDDs and PCDFs were identified. Furthermore, the developed model was externally validated.

Materials and Methods

The following materials were utilized in this research: ChemDraw Ultra 12.0, Cambridge Soft Corp. (www.cambridgesoft.Corn), USA;PyRx software (http://pyrx.sourceforge.net, Scripps Institute); Accelerys Discovery Studio 4.5 version, Ligplot and AutoDock 4.2. (Anitha*et al.*, 2013) with MGL tools (http://mgltools.scripps.edu/downloads) installed on a Dell personal computer (PC) equipped with 8GBRAM capacity, processor intel CORE ™i5, hard disc capacity of 1000 GB and CPU@ 2.20GHz2.20GHz running on 64-bit Operating System and data set of 71 molecules was collected from published literature (Azeddine*et al*., 2014; Huifeng*et al*., 2011; Martin *et al*., 2016). The chemical structures and the biological response (pEC_{50}) of these 71 molecules are

presented in Table 1. The entire set of compounds was divided into two subsets: training set (70%) consisting of 49 molecules were used to build the actual models, and test set (30%) consisting of 22 molecules, not found in the training set, were used to validate the models once they were built. Members of each set were assigned using Kennard Stone method.

Table 1: The chemical structures and biological response (Observed and predicted pEC_{50}) of polychlorodibenzodioxins, polychlorodibenzofurans and polychlorobiphenyls

S/N	Name	Observed pEC_{50}	Predicted pEC_{50}
1[b]	2-Chlorodibenzofuran	3.55	3.54
2[a]	3-Chlorodibenzofuran	4.38	4.17
3[a]	4-Chlorodibenzofuran	3.00	3.50
4[a]	2,3-Dichlorodibenzofuran	5.33	5.05
5[a]	1,3,6-Trichlorodibenzofuran	5.36	4.87
6[a]	1,3,8-Trichlorodibenzofuran	4.07	4.84
7[a]	2,3,4-Trichlorodibenzofuran	4.72	5.69
8[a]	2,3,8-Trichlorodibenzofuran	6.00	5.99
9[a]	2,6,7-Trichlorodibenzofuran	6.35	6.20
10[a]	2,3,4,6-Tetrachlorodibenzofuran	6.46	6.46
11[a]	2,3,4,8-Tetrachlorodibenzofuran	6.70	6.64
12[a]	2,3,7,8-Tetrachlorodibenzofuran	7.39	7.33
13[a]	1,2,4,8-Tetrachlorodibenzofuran	5.00	4.89
14[a]	1,2,4,7,9-Pentachlorodibenzofuran	4.70	4.94
15[a]	1,2,3,7,8-Pentachlorodibenzofuran	7.13	6.83
16[a]	1,2,4,7,8-Pentachlorodibenzofuran	5.89	6.09
17[a]	2,3,4,7,8-Pentachlorodibenzofuran	7.82	7.73
18[b]	1,2,3,4,7,8-Hexachlorodibenzofuran	6.64	7.08
19[a]	1,2,3,6,7,8-Hexachlorodibenzofuran	6.57	6.75
20[a]	2,3,4,6,7,8-Hexachlorodibenzofuran	7.33	7.27
21[a]	2,3,6,8-Tetrachlorodibenzofuran	6.66	6.67
22[a]	1,2,3,6-Tetrachlorodibenzofuran	6.46	5.97
23[a]	1,2,3,7-Tetrachlorodibenzofuran	6.96	6.30
24[a]	1,3,4,7,8-Pentachlorodibenzofuran	6.70	6.83
25[a]	2,3,4,7,9-Pentachlorodibenzofuran	6.70	6.15
26[b]	1,2,3,7,9-Pentachlorodibenzofuran	6.40	6.08
27[a]	Dibenzofuran	3.00	3.04
28[a]	8-Hydroxy-2,3,4-trichlorodibenzofuran	6.52	6.65
29[a]	3,8-Dihydroxy-2-chlorodibenzofuran	5.43	5.63
30[a]	8-Hydroxy-2-monochlorodibenzofuran	6.52	5.96
31[b]	8-Hydroxy-3,4,6-trichlorodibenzofuran	6.52	5.83
32[b]	8-Hydroxy-3,4-dichlorodibenzofuran	6.22	6.06
33[a]	8-Hydroxy-3-monochlorodibenzofuran	6.37	6.59
34[a]	2,3,7,8-Tetrachlorodibenzo-p-dioxin	8.00	7.79
35[a]	1,2,3,7,8-Pentachlorodibenzo-p-dioxin	7.10	7.26
36[a]	2,3,6,7-Tetrachlorodibenzo-p-dioxin	6.80	6.82
37[a]	2,3,6-Trichlorodibenzo-p-dioxin	6.66	6.06
38[a]	1,2,3,4,7,8-Hexachlorodibenzo-p-dioxin	6.55	6.46
39[a]	1,3,7,8-Tetrachlorodibenzo-p-dioxin	6.10	6.72
40[a]	1,2,4,7,8-Pentachlorodibenzo-p-dioxin	5.96	5.99
41[b]	1,2,3,4-Tetrachlorodibenzo-p-dioxin	5.89	5.08
42[a]	1,3,7-Trichlorodibenzo-p-dioxin	7.15	7.36
43[b]	1,2,3,4,7-Pentachlorodibenzo-p-dioxin	5.19	5.89
44[b]	1,2,4-Trichlorodibenzo-p-dioxin	4.89	4.45
45[a]	1,2,3,4,6,7,8,9-Octachlorodibenzo-p-dioxin	5.00	4.83
46[a]	1-Chlorodibenzo-p-dioxin	4.00	4.20
47[a]	2,3,7,8-Tetrabromodibenzo-p-dioxin	8.82	9.02
48[b]	2,3-Dibromo-7,8-dichlorodibenzo-p-dioxin	8.83	8.66
49[b]	2,8-Dibromo-3,7-dichlorodibenzo-p-dioxin	9.35	9.05
50[b]	2-Bromo-3,7,8-trichlorodibenzo-p-dioxin	7.94	8.69
51[a]	1,3,7,9-Tetrabromodibenzo-p-dioxin	7.03	6.68
52[b]	1,3,7,8-Tetrabromodibenzo-p-dioxin	8.70	8.27
53[a]	1,2,4,7,8-Pentabromodibenzo-p-dioxin	7.77	7.56
54[a]	1,2,3,7,8-Pentabromodibenzo-p-dioxin	8.18	8.64
55[b]	2,3,7-Tribromodibenzo-p-dioxin	8.93	8.67
56[b]	2,7-Dibromodibenzo-p-dioxin	7.81	8.05
57[a]	2-Bromodibenzo-p-dioxin	6.53	6.45
58[a]	2,2',4,4'-Tetrachloro-1,1'-biphenyl	3.89	3.59
59[b]	2,3,4,4'-Tetrachloro-1,1'-biphenyl	4.55	4.98
60[a]	2,3,4,5-Tetrachloro-1,1'-biphenyl	3.85	4.47
61[a]	3,3',4,4'-Tetrachloro-1,1'-biphenyl	6.15	5.93
62[b]	2,3,3',4,4'-Pentachloro-1,1'-biphenyl	5.37	5.27
63[b]	2,3,4,4',5-Pentachloro-1,1'-biphenyl	5.39	5.17
64[b]	2,3',4,4',5-Pentachloro-1,1'-biphenyl	5.04	5.14
65[b]	3,3',4,4',5-Pentachloro-1,1'-biphenyl	6.89	6.06
66[b]	2,3',4,4',5'-Pentachloro-1,1'-biphenyl	4.85	4.96
67[b]	2,2',4,4',5,5'-Hexachloro-1,1'-biphenyl	4.10	4.10
68[a]	2,3,3',4,4',5-Hexachloro-1,1'-biphenyl	5.15	5.43
69[b]	2,3,3',4,4',5'-Hexachloro-1,1'-biphenyl	5.33	5.19
70[a]	2,3',4,4',5,5'-Hexachloro-1,1'-biphenyl	4.80	5.16
71[a]	2,3',4,4',5',6-Hexachloro-1,1'-biphenyl	4.00	3.55

[a] Training set; [b] Test set

Calculation of descriptors

Different types of descriptors were calculated for each molecule in Table 1 using default settings within Spartan'14 version 1.1.2 (Wavefunction, 2013) and PaDEL-Descriptor version 2.18 (Yap, 2011). These descriptors include steric, electrostatic, electronic, spatial, structural, and thermodynamic.

Generation of QSAR models

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 (descriptors) of the molecules. In the present study, we have used genetic function approximation (GFA) technique to generate different 1D, 2D, and 3DQSAR models from various descriptors available within Spartan'14 version 1.1.2 (Wavefunction, 2013) and PaDEL version 2.18 (Yap, 2011) modeling software in order to deduce correlation between the structure and biological activity of the present series of molecules.

Homologous modeling

The amino acid residue sequence (sequence GI: 7304873) of AhR conservative domain (the number of residues is from 278 to 384) for mouse was obtained by searching NCBI. The sequence analysis and molecular modeling were completed through both Internet resources and PCs. Homologous 3D model of AhR was built on SWISS-MODEL net server. The nuclear magnetic resonance (NMR) structure of the human PAS domain of the hypoxia-inducible factor 2R (HIF-2R) available in the Protein Data Bank (http://www.rcsb.org/pdb, PDB ID: 1P97) was used as the 3D coordinate template for the homology modeling of AhR (Huifeng*et al*., 2011).

Docking studies

The binding mode for the PCDDs, PCDFs, and PCBs to AhR was investigated by PyRx software using Autodock 4.2 {http://pyrx.sourceforge.net, Scripps Institute} (Dong-Chan & Jae-Ki, 2016; Trott& Olson, 2010).The three dimensional (3D) crystal structure of AhR (PDB ID = 1P97) was downloaded from PDB (http://www.rcsb.org/pdb). Before docking, all water molecules and the cofactors were removed. The 3D structures of the compounds were optimized by energy minimization using Discovery Studio 4.5 version (DS, Accelrys Software Inc., USA) (Lei *et al*., 2015). The X, Y, Z grid of the compounds binding site on the AhR was indentified and calculated with the "centroid program" in the Discovery Studio 4.5, X= 19.337, Y= 42.037, Z=35.848, (Rohmah*et al*., 2015). PyRx was used to dock the compounds into the X/Y/Z grid of the AhR with the flexible docking option turned on. To examine the docking conformational space comprehensively, the search efficiency was set at 100%. The highest binding affinity (the lowest docking energy) score was chosen to explore the binding mode of docked compound in the AhR protein using PyRx program. For the analysis of the docking calculations, 9 conformers were considered for each ligand-macromolecule complex, and the resulting docking clusters were calculated with a 2.0 Å root mean squared deviation (RMSD) tolerance on the heavy atoms (Sambasivarao*et al*., 2014). The two-dimensional (2D) and 3D molecular interaction models of the docked compounds on the AhR protein and receptor surface shape modeling (involving hydrogen bonding) were displayed using Ligplot and Discovery Studio (Maryam*et al*., 2015; Usha*et al*., 2014).

Model validation

The internal and external validation parameters were used to evaluate the fitting ability, stability, reliability and predictive ability of the developed models. The validation parameters were compared with the minimum recommended value for a generally acceptable QSAR model (Ravinchandran*et al*.,

2011) showed in Table 2. The performance of external validation was characterized by the determination coefficient (R^2), root mean standard error (*RMSE*) and external explained variance (R^2_{ext}), which are defined as follows (Huifeng*et al*., 2011):

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

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Y_{obs} - Y_{pred})^2}{n}} \quad 2$$

$$R^2_{ext} = 1 - \frac{\sum(Y_{obs(Test)} - Y_{pred(Test)})^2}{\sum(Y_{obs(Test)} - \bar{Y}_{training})^2} \quad 3$$

WhereY_{obs}; Y_{pred}; $\bar{Y}_{training}$ are observed activity values, predicted activity values and the mean observed activity values of the samples in the training set, respectively. n is the total number of samples in the training set, $Y_{obs(Test)}$, $Y_{pred(Test)}$, $\bar{Y}_{training}$are observed activity values, predicted activity values and the mean observed activity values of the samples in the test set, respectively.

Table 2 Minimum recommended value of Validation Parameters for a generally acceptable QSAR model compared with model 1

Symbol	Name	Threshold Value	Values for model 1
R^2	Coefficient of determination	≥ 0.6	0.930
$P_{(95\%)}$	Confidence interval at 95% confidence level	< 0.05	P < 0.0001
Q^2	Cross validation coefficient	≥ 0.5	0.911
R^2 - Q^2	Difference between R^2 and Q^2	≤ 0.3	0.019
$N_{ext.test\ set}$	Minimum number of external test set	≥ 5	22
R^2_{ext}	Coefficient of determination for external test set	≥ 0.6	0.824
RMSE	Root mean standard error	smaller value is better	0.353

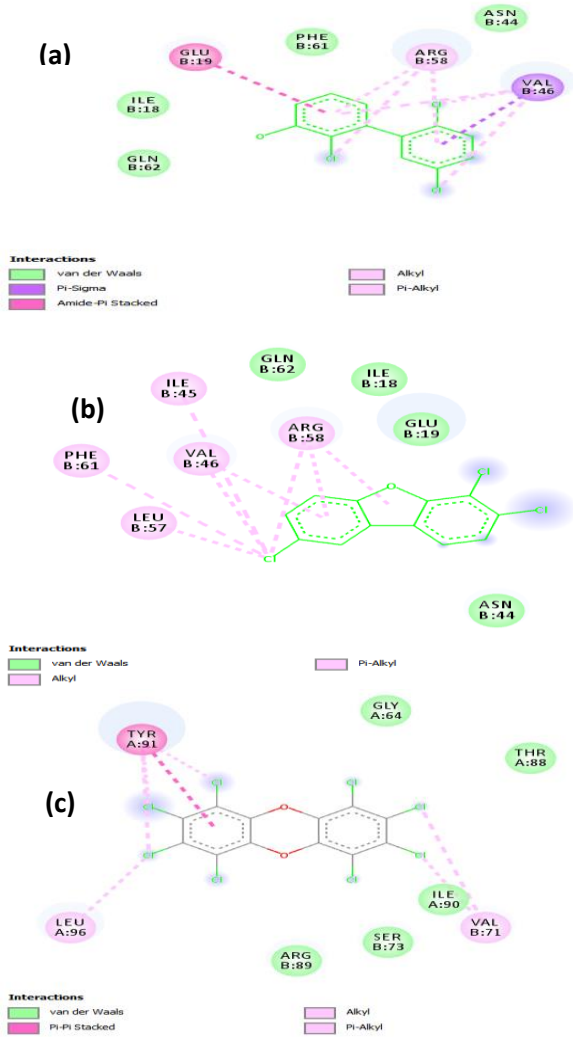

Fig. 1: Receptor – Ligand interactions on a 2D diagram of **(a)** PCB, **(b)**PCDF and **(c)**PCDD with receptor

Results and Discussions

The target sequence of the ligand bind domain of the mouse (278- 384) AhR was used as a query to search for homologues protein structure belonging to the category of nuclear receptors that could serve as templates. The x-ray crystalline structure of the high affinity heterodimer of HIF-2R alpha and ARNT c-terminal pas domains (resolution: 1.65A^o) with the artificial ligand THS017 (PDB ID 1P97) showed detectable degree of similarity with the query sequence.

The image of PCB, PCDF and PCDD docked into the homology model of mouse AhR is shown in Fig. 1(a, b & c). The result of docking of PCB, PCDF and PCDD into mAhR showed that the residues lining the ligand binding cavity include Glu19, Phe61, Arg58, Asn44, Ile18, Val46, Gln62, Ile45, Leu57, Gly64, Tyr91, Thr88, Ile90, Val71, Ser73, Arg89, Leu96. Also Interactions such as Van der Waals, Pi-sigma, Amide-Pi stacked, Alkyl, Pi-Alkyl and Pi-Pi stacked were observed between receptor and ligand. This is also in accordance with the literature report of Huifeng*et al*. (2011); thus, validating our model further.

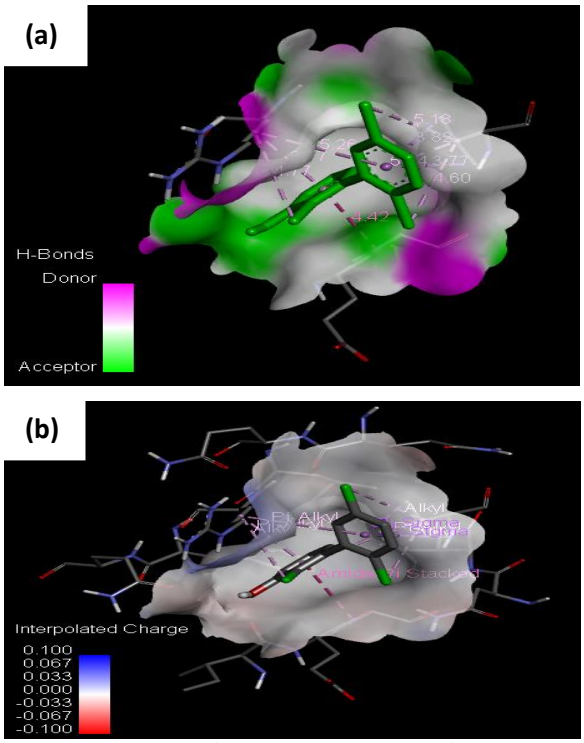

Fig. 2: Docking views of PCB in the binding site of AhR**(a)** H-bonds and **(b)** interpolated charge

Acting as an 'anchor', the hydrogen-bonding intensely determines the 3D space position of the benzene ring in the binding pocket, and facilitates the hydrophobic interaction of the PCB with Phe61, Glu19, Ile319, Arg58 and Val46, as shown in Fig. 2(a). The molecular surface potential indicates the charge distribution in a molecule (Politzer*et al*., 1984), which gauge the basicity and nucleophilicity of a molecule (Colominas*et al.*, 1998). As shown in Fig. 2(b), the binding site has negative potentials, from which it can also be concluded that the positive potentials of the PCB molecules facilitate them to bind with receptor.

QSAR development and validation

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. GFA was performed using 10,000 crossovers, smoothness value of 2.0 and other default settings for each combination. The number of terms in the equation was fixed to five including constant in the training set. The set of equations generated were evaluated on the following basis: (a) Friedman's Lack of fit **(**LOF) measure; (b) Variable terms in the equations; (c) Cross validated and non-cross validated R^2; (d) Predictive ability of the equation R^2_{pred}. GFA technique was used since it generates a population of equations rather than one single equation for correlation between biological activity and physicochemical properties. Table 3 shows the Summary of generated GFA equations from Materials Studio 8.0.

Table 3: Summary of generated GFA equations from materials studio 8.0

		Equation 1	Equation 2	Equation 3	Equation 4	Equation 5
1	Friedman LOF	0.6048	0.6142	0.6178	0.6241	0.6253
2	R-squared	0.9303	0.9293	0.9288	0.9281	0.9279
3	Adjusted R-squared	0.9222	0.9210	0.9206	0.9198	0.9196
4	Cross validated R-squared	0.9109	0.9135	0.9125	0.9131	0.9133
5	Significant Regression	Yes	Yes	Yes	Yes	Yes
6	Significance-of-regression F-value	114.8633	112.9695	112.2619	111.0372	110.8042
7	Critical SOR F-value (95%)	2.4591	2.4591	2.4591	2.4591	2.4591
8	Replicate points	0	0	0	0	0
9	Computed experimental error	0.0000	0.0000	0.0000	0.0000	0.0000
10	Lack-of-fit points	43	43	43	43	43
11	Min expt. error for non-significant LOF (95%)	0.3211	0.3236	0.3245	0.3262	0.3265

Table 4: Description of the descriptors used in the QSAR optimization model

Descriptors	Meaning
BCUTw-1h	nlow highest atom weighted BCUTS
SpMax7_Bhp	Largest absolute eigenvalue of Burden modified matrix - n 7 / weighted by relative polarizabilities
Gmin	Minimum E-State
RDF65e	Radial distribution function - 065 / weighted by relative Sanderson electronegativities
E1v	1st component accessibility directional WHIM index / weighted by relative van der Waals volumes

Table 5: Pearson's correlation matrix for descriptors used in the model 1 and pEC_{50}

	PEC50	*BCUTw-1h*	*SpMin7_Bhp*	*gmin*	*RDF65e*	*E1v*
PEC50	1					
BCUTw-1h	0.46601	1				
SpMin7_Bhp	0.055096	0.047723	1			
gmin	-0.36428	-0.66601	-0.69776	1		
RDF65e	-0.39905	-0.21615	0.020404	0.150597	1	
E1v	0.733047	0.321281	-0.25779	0.065759	0.08446	1

In a QSAR study, the quality of a model is expressed by its fitting and prediction ability. In order to build and test the model, a data set of 71 compounds was separated into a training set of 49 compounds, which was used to build model and a test set of 22 compounds, which was applied to validate the built model. The GFA analysis led to the derivation of five models, with five descriptors as shown in Table 3, but only the best model (Equation 1) was selected and reported due to small value of LOF, high value of R^2. The Description of the descriptors used in the QSAR optimization model are shown in Table 4. Fig. 3 is the graph of calculated pEC_{50} against the experimental values for the training sets. The R^2 value of the QSAR model was 0.930, indicating a high goodness-of-fit of the model. Q^2_{LOO} of the QSAR was as high as 0.911, implying good robustness of the model. The differences between R^2 and Q^2_{LOO} (0.019) did not exceed 0.3, indicating no over-fitting in the model (Golbraikh&Tropsha, 2002).

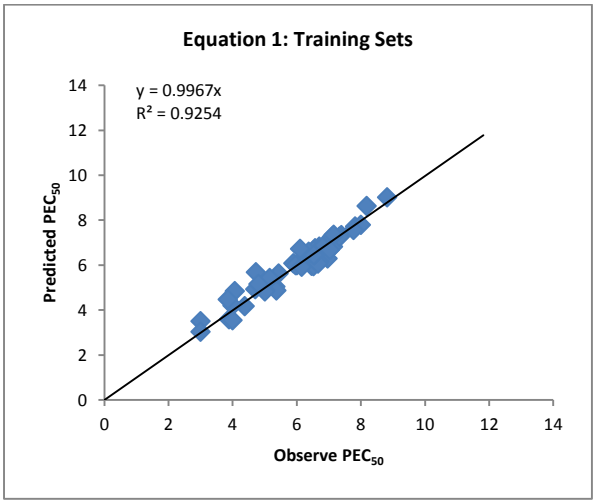

Fig. 3: Plot of observed versus predicted pEC_{50} values for the training sets

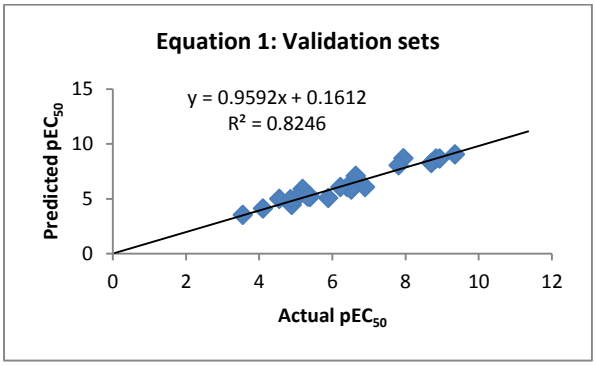

Fig. 4: Plot of observed versus predicted pEC_{50} values for the validation sets

As shown in Figs. 3, 4 and 5, the predicted pEC_{50} values were consistent with the observed values for both the validation and training sets. The model revealedacceptable predictability with R^2_{pred}= 0.824, RMSEvalues for the training and test sets are 0.353 and 0.432, respectively. In summary, the developed QSAR model showed satisfactory performance.

pEC50 = 5.315 -0.055 BCUTw-1h -4.151 SpMin7_Bhp -5.501 gmin -0.306 RDF65e +15.875 E1v

n(training set) = 49, R^2 = 0.930, Q^2_{LOO} = 0.911,RMSE = 0.353 (training set)

n(validation set) = 22, R^2_{ext}= 0.924, RMSE = 0.432 (validation set)

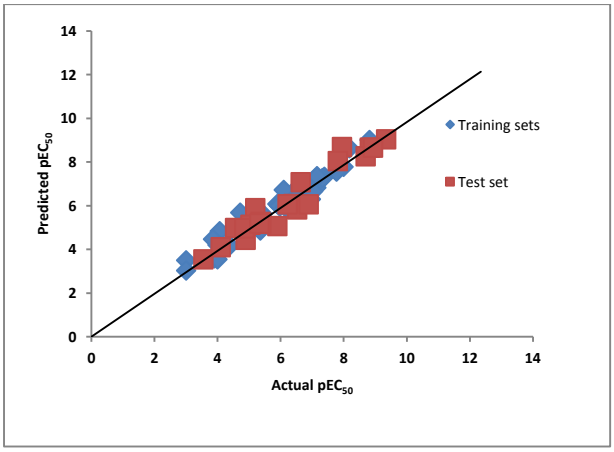

Fig. 5: Plot of observed versus predicted pEC_{50} values for the training and validation sets

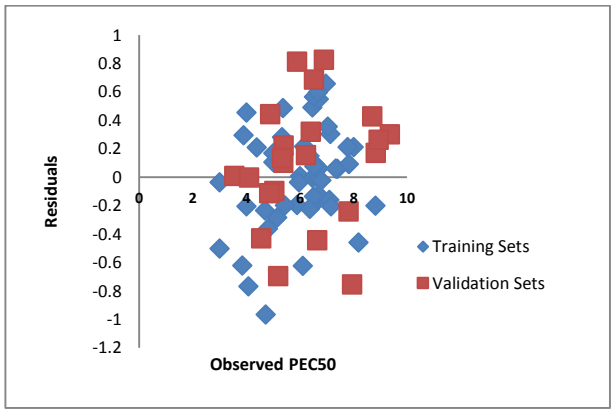

Fig. 6: The residuals vs. observed pEC50 values for the training and test sets

In this equation, R^2 is the squared correlation coefficient,Q^2_{LOO}, is the cross-validation coefficients for leave one out, and RMSE is the root mean square error. The predicted values for pEC50 for the compounds in the training and test sets using equation 1 were plotted against the experimental pEC50 values in Figs. 3, 4 and 5. Also, the plot of the residual for the predicted values of pEC50 for both the training and test sets against the observed pEC50 values are shown in Fig. 6. As can be seen the model did not show any proportional and systematic error, because the propagation of the residuals on both sides of zero is random. The Pearson's correlation matrixes for descriptors used in the model are shown in Table 5. The result from this Correlation matrix shows clearly that the correlation coefficients between each pair of descriptors is very low, thus, it can be inferred that there exist no significant inter-correlation among the descriptors used in building the model.

Conclusion

Docking analysis showed that hydrogen bonding, hydrophobic interactions, Van der Waals, Pi-sigma, Amide-Pi stacked, Alkyl, Pi-Alkyl and $\pi - \pi$ interactions between compounds and AhR governed the binding affinities. GFA handled the physico-chemical descriptors effectively in the generation of QSAR models with significant statistical terms including external predictivity. Equation 1 from the five models generated was selected as representative equation and was able to explain more than 93%($R^2 = 0.9303$) of the total variance. The structural interpretation of the obtained models shows that the distribution of charge, the distant intramolecular, molecular size, shape profiles, polarizability and electropological interactions play important roles for modeling the toxicity. The developed QSAR model had good robustness, predictive ability andmechanism interpretability, which could be applied to predictthe binding affinity of some polychlorinated aromatic compounds.

References

Anitha K, Gopi G, GirishSenthil& Kumar P 2013. Molecular docking study on dipeptidyl peptidase-4 inhibitors. *Int. J. Res. Dev. Pharm. L. Sci.*, 2(5): 602-610

Azeddine A, Rachid H, Majdouline L, Mohammed B &Tahar L 2014. Binding affinities (AhR) of polychlorinated biphenyls (PCBs), dibenzo-p-dioxins (PCDDs) and dibenzofurans (PCDFs) study combining DFT and QSAR results. *IJARCSSE* vol. 4.

Borosky GL, LaaliKK 2005. A computational study of carbocations from oxidized metabolites of dibenzo[a,h]acridine and their fluorinated and methylated derivatives. *Chem. Res. Toxicol*., 18: 1876–1886.

Burbach KM, Poland A & Bradfield CA. 1992. Cloning of the Ah-receptor cDNA reveals a distinctive ligand-activated transcription factor. *P. Natl. Acad. Sci.*, 89(17): 8185-8189.

Colominas C, Orozco M, LuqueFJ, BorrellJI&Teixido J 1998. A priori prediction of substituent and solvent effects in the basicity of nitriles. *J. Org. Chem*., 63: 4947–4953.

Dong-Chan K & Jae-Ki R 2016. In silicoAnalysis and Molecular Docking Comparison of Curcumin and Bisdemethoxycurcumin on Transthyretin. *Am. J. Pharmacol. Sci.*, 4(2): 28-30.

Gallegos A, Robert D, Girones X &Carbo-Dorca R 2001. Structure–toxicity relationships of polycyclic aromatic hydrocarbons using molecular quantum similarity. *J. Comput. Aid. Mol. Des.*, 15: 67–80.

Gerhard Lammel& Rainer Lohmann 2012. Identifying the research needs in the global assessment of toxic compounds 10 years after the signature of the Stockholm Convention. *Envt. Sci. Pollut. Res.*, 19: 1873–1874.

Golbraikh A &Tropsha A 2002. Beware of q^2. *J. Molec. Graphics & Mod.*, 20(4): 269–276.

Hansch C, Hoekman D, Leo A, Weininger D & Selassie CD 2002. Chembioinformatics: Comparative QSAR at the interface between chemistry and biology. *Chem. Rev.*, 102: 783–812.

HestermannEV, StegemanJJ& Hahn ME 2000. Relative contributions of affinity and intrinsic efficacy to aryl hydrocarbon receptor ligand potency. *Toxicol. Appl. Pharmacol*., 168: 160–172.

Hilscherova K, Machala M, Kannan K, Blankenship AL &Giesy JP 2000. Cell bioassays for detection of aryl hydrocarbon (AhR) and estrogen receptor (ER) mediated activity in environmental samples. *Environ. Sci. Pollut.R.*, 7: 159–171.

http://pyrx.sourceforge.net; Scripps Institute.

http://www.rcsb.org/pdb; PDB ID: 1P97.

Huang ZJ, Edery I &Rosbash M 1993. PAS is a dimerization domain common to Drosophila period and several transcription factors. *Nature*, 364: 259-262.

Huifeng Wu Fei Li, Xuehua Li, Xiaoli Liu, Linbao Zhang, Liping You &Jianmin Zhao 2011. Docking and 3DQSAR studies on the Ah receptor binding affinities of polychlorinated biphenyls (PCBs), dibenzo-p-dioxins (PCDDs) and dibenzofurans (PCDFs). *Envtal. Toxico. &Pharmaco.* 3(2): 478–485.

Kewley RJ, Whitelaw ML & Chapman-Smith A, 2004. The mammalian basic helix-loop-helix/PAS family of transcriptional regulators. *Int. J. Biochem. Cell B*, 36(2): 189-204.

Kumar S, Chang RL, Wood AW, Xie JG, Huang MT, Cui XX, Kole PL, Sikka HC, Balani SK, Conney AH &JerinaDM 2001. Tumorigenicity of racemic and optically pure bay region diol epoxides and other derivatives of the nitrogen heterocycledibenz[a,h]acridine on mouse skin. *Carcinogenesis*, 22: 951–955.

Landers JP &Bunce NJ 1991. The Ah receptor and the mechanism of dioxin toxicity. *Biochem. J.*, 276, 273–287.

Lei Q, Liu H, Peng Y & Xiao P 2015. In silico target fishing and pharmacological profiling for the isoquinoline alkaloids of Macleayacordata (Bo LuoHui). *Chinese Medicine*, 10: 37.

LucierGWPortier CJ & Gallo MA 1993. Receptor mechanisms and dose–response models for the effects of dioxins. *Environ. Health Perspect*. 101: 36–44.

MárciaSP 2004. Polychlorinated Dibenzo-P-Dioxins (PCDD), dibenzofurans (PCDF) and polychlorinated biphenyls (PCB): main sources, environmental behaviour and risk to man and biota; *Quim. Nova*., 27(6) 934-943.

Martin Šala, Alexandru T, BalabanMarjanVeber&MatevžPompe, 2016. QSAR Models for Estimating Aryl Hydrocarbon Receptor Binding Affinity of Polychlorobiphenyls, Polychlorodibenzodioxins, and Polychlorodibenzofurans, *MATCH Commun. Math. Comput. Chem.*,75: 559-582.

Maryam I, Atefeh S &Asghar D 2015. Molecular docking analysis and molecular dynamics simulation study of ameltolide analogous as a sodium channel blocker. *Turkish J. Chem.*, 39: 306 – 316.

NebertDW, Puga A &Vasiliou V 1993. Role of the Ah receptor and the dioxin-inducible [Ah] gene battery in toxicity, cancer, and signal-transduction. *Ann. N.Y. Acad. Sci.*, 685: 624–640.

Ohura T, Morita M, Kuruto-Niwa R, Amagai T, Sakakibara H &Shimoi K 2010. Differential action of chlorinated polycyclic aromatic hydrocarbons on aryl hydrocarbon receptor-mediated signaling in breast cancer cells. *Environ. Toxicol*., 25: 180–187.

PerdewGH 1988. Association of the Ah receptor with the 90-kDa heat shock protein. *J. Biol. Chem.*, 263(27): 13802-5.

Politzer P, Abrahmsen L &Sjoberg P 1984. Effects of amino and nitro substituents upon the electrostatic potential of an aromatic ring. *J. Am. Chem. Soc.*, 106: 855–860.

Ravinchandran V, Rajak H, Jain A, Sivadasan S, Varghese CP& Kishore-Agrawal R 2011. Validation of QSAR models-strategies and importance. *Int. J. Drug Design & Discovery*, 2: 511–519.

Rohmah RN, Hardiyanti F &Fatchiyah F 2015. Inhibition on JAK-STAT3 Signaling Transduction Cascade Is Taken by Bioactive Peptide Alpha-S2 Casein protein from goat Ethawah breed milk. Actainformaticamedica: AIM: *J. Soc. Medical Informatics of Bosnia & Herzegovina: CasopisDrustvazaMedicinskuInformatikuBiH*., 23: 233-238.

SambasivaraoSV, Roberts J, BharadwajVS, Slingsby JG, Rohleder C & Mallory C 2014. Acetylcholine promotes binding of alpha-conotoxin MII at alpha3beta2 nicotinic acetylcholine receptors. *Chembiochem: Eur. J. Chem. Bio*., 15: 413-24.

Schmidt JV & Bradfield CA 1996. Ah receptor signaling pathways. *Annu. Rev. Cell Devt. Bio.*, 12(1): 55-89.

Trott O & Olson AJ 2010. AutoDockVina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Computational Chem*., 31: 455-461.

Usha T, Middha SK, Goyal AK, Karthik M, Manoj D &Faizan S *et al* 2014. Molecular docking studies of anti-cancerous candidates in *Hippophaerhamnoides* and *Hippophaesalicifolia*. *J. Biomed. Res.*, 28: 406-415.

Wavefunction2013. Inc. Spartan'14, version 1.1.2: Irvine, California, USA.

Whitlock Jr JP 1993. Mechanistic aspects of dioxin action. *Chem. Res. Toxicol*., 6(6): 754-763.

www.cambridgesoft.Corn

Xue WL &Warshawsky D 2005. Metabolic activation of polycyclic and heterocyclic aromatic hydrocarbons and DNA damage: a review. *Toxicol. Appl. Pharm*. 206, 73–93.

Yap Chun Wei 2011. Inc. PaDEL-Descriptor, version 2.18: A software to calculate molecular descriptors and fingerprints. http://padel.nus.edu.sg.