RESEARCH   **Open Access**

QSAR, QSTR, and molecular docking studies of the anti-proliferative activity of phenylpiperazine derivatives against DU145 prostate cancer cell lines

Fabian A. Ikwu[*] (iD), Gideon A. Shallangwa and Paul A. Mamza

Abstract

Background: Prostate cancer is the most common non-cutaneous cancer in males and accounts for about 4% of all cancer-related deaths in males annually. In silico methods provide faster, economical, and environmentally friendly alternatives to the traditional trial and error method of lead identification and optimization. This study, therefore, was aimed at building a robust QSAR and QSTR model to predict the anti-proliferate activity and toxicity of some phenylpiperazine compounds against the DU145 prostate cancer cell lines and normal prostate epithelial cells as well as carry out molecular docking studies between the compounds and the androgen receptor.

Results: Genetic Function Algorithm–Multilinear Regression approach was employed in building the QSAR and QSTR model. The QSAR model built had statistical parameters $R^2 = 0.7792$, $R^2_{adj.} = 0.7240$, $Q^2_{cv} = 0.6607$, and $R^2_{ext} = 0.6049$ and revealed the anti-proliferate activity to be strongly dependent on the molecular descriptors: VR3_Dzp, VE3_Dzi, Kier3, RHSA, and RDF55v. The QSTR model, on the other hand, had statistical parameters $R^2 = 0.8652$, $R^2_{adj.} = 0.8315$, $Q^2_{cv} = 0.7788$, and $R^2_{ext} = 0.6344$. The toxicity of the compounds was observed to be dependent on the descriptors MATS8c, MATS3s, ETA_EtaP_F, and RDF95m. The molecular descriptors in both models were poorly correlated ($R < 0.4$) and had variance inflation factors < 3. Molecular docking studies between the androgen receptor and compounds 25 and 32 revealed the compounds primarily formed hydrogen, halogen, and hydrophobic interactions with the receptor.

Conclusion: Findings from this study can be employed in in silico design of novel phenylpiperazine compounds. It can also be employed in predicting the toxicity and anti-proliferate activity of other phenylpiperazine compounds against DU145 prostate cancer cell lines.

Keywords: Quantitative structure activity relationship, Quantitative structure toxicity relationship, Molecular docking, Prostate cancer, Computational chemistry

1 Background

The emergence of new diseases coupled with the increasing resistance of existing diseases to therapies currently in use has ensured the continual need for novel medication and therapies. The need for novel therapies has had its toll on the environment—particularly on plants and micro-organisms [1, 2]. The use of the trial and error method of traditional drug design has led to wastage of precious natural resources, disruption of the ecosystem, and in some cases, environmental pollution. Additionally, obtaining a drug candidate using traditional methods usually take a long time [3]. Computational methods, however, proffer time-saving, cost-efficient, and cleaner alternatives. Computer-aided drug design (precisely ligand based design) employs tools that

* Correspondence: ikwufabian@gmail.com
Department of Chemistry, Faculty of Physical Sciences, Ahmadu Bello University, Zaria, Nigeria

model the activities of certain compounds, and afterward, the model built is used to predict the activities of other compounds having a similar pharmacophore [4, 5].

Cancer refers to a collection of over 200 different types of diseases in which certain cells grow abnormally and have the potential to invade or spread to other body parts, killing healthy cells in the process [6]. Prostate cancer is one of the most common cancers diagnosed in males. It is second only to skin cancer and more prevalent than lung cancer in the UK and the USA. Prostate cancer accounts for about 4% of all cancers suffered by men annually. An estimate of 1.3 million cases was reported worldwide in 2018. Prostate cancer is prevalent in older men (> 65 years) with 6 out of 10 men in the age group being diagnosed with the cancer. About 80% of all prostate cancer cases are reported in this age group [7–9]. The mortality rate of prostate cancer is not very high; the 5-year survival rate of prostate cancer is about 98%. However, prostate cancer is still one of the leading causes of cancer-related deaths in males, second only to lung cancer. It is the fifth leading cause of deaths in males worldwide. It also accounts for about 4% of cancer-related deaths, and over three hundred thousand deaths were attributed to the cancer in 2018 in the USA [7, 8].

DU145 cell line is one of the three most used prostate cancer cell lines in prostate cancer research, the other two being PC3 and LNCaP cell lines. DU145 cells have moderate metastatic potential compared to PC3 cells and are androgen receptor positive [10]. The growth of prostate cancer cell lines has been reported to be via modification of the androgen receptor [11]. Thus, modification of the androgen receptor is a viable strategy for combating early-stage prostate cancer. This study built a QSAR and QSTR model to predict the activity and toxicity of some phenylpiperazine derivatives against DU145 prostate cancer cell lines and normal prostate epithelial cells. The study also investigated the interaction between the compounds and the androgen receptor via molecular docking studies.
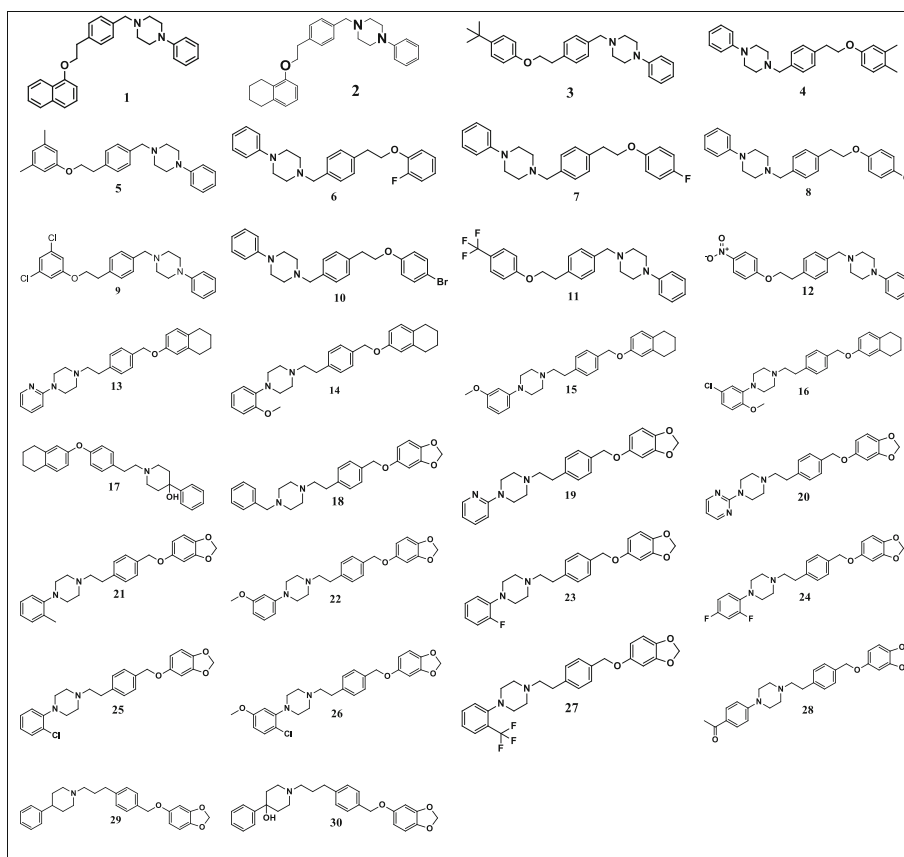
2 Methods
2.1 Dataset
Thirty-seven (37) phenylpiperazine derivatives reported by Chen et al., [12, 13] were employed in building a
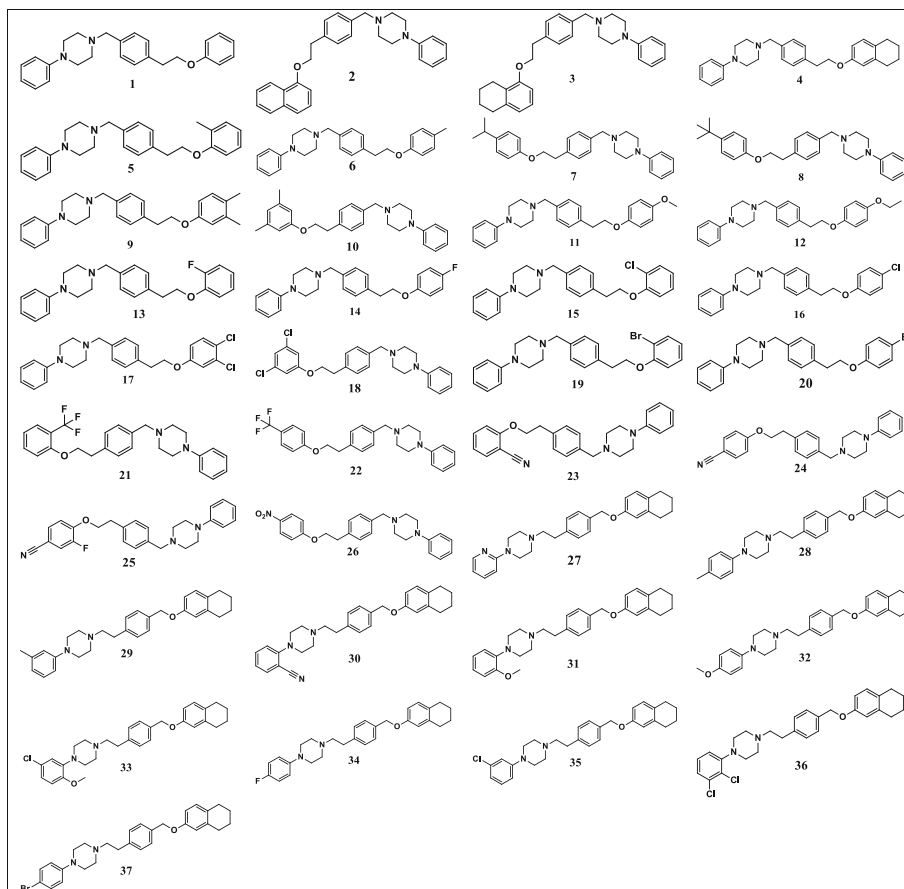
**Fig. 1** 2D structures compounds used in QSAR Study

Fig. 2 2D structures compounds used in QSTR Study

QSAR model that predicts their anti-proliferate activity against DU145 prostate cancer cell lines. Thirty (30) other derivatives having toxic (proliferative) activity against normal prostate epithelial cells were employed in building the QSTR model [12–14]. Figure 1 presents 2D structures of the compounds used to build the QSAR model while Fig. 2 presents those for the QSTR model. The anti-proliferative activity (IC_{50}) of the compounds ranged from 0.77 to 46.24 μM while toxicity ranged from 3.87 to 49.21 μM. The activity and toxicity of the compounds were converted to the logarithmic scale using the formula $pIC_{50} = -\log_{10}(IC_{50})$. This conversion reduced the skew in the activity and linearized the activity of the compounds [15]. In the logarithmic scale, the activity ranged from 4.3350 to 6.1135 while toxicity ranged from 4.3079 to 5.4123, respectively.

2.2 QSAR/QSTR model

A 2D structure of each molecule was drawn using the ChemDraw Ultra 12.0 software and then converted to their equivalent 3D structure using the Spartan 14 V1.4 software. The ground state equilibrium geometry of the compounds was afterward obtained via optimization using the B3LYP/6-31G* basis set of the density
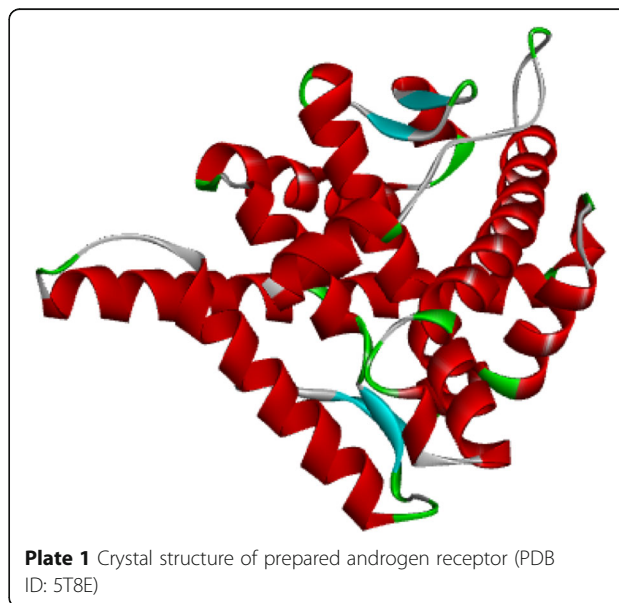
Plate 1 Crystal structure of prepared androgen receptor (PDB ID: 5T8E)

Table 1 Statistical parameters of model 1

Parameter	Model 1	Model 2	Model 3	Model 4
Friedman LOF	0.318397	0.380396	0.388804	0.390866
R-squared	0.779177	0.736178	0.730346	0.728916
Adjusted *R*-squared	0.723971	0.670223	0.662933	0.661146
Cross validated *R*-squared	0.660697	0.575643	0.553878	0.554898
Significant regression	Yes	Yes	Yes	Yes
Significance-of-regression *F*-value	14.11407	11.161743	10.833846	10.755597
Critical SOR *F*-value (95%)	2.732939	2.732939	2.732939	2.732939
Replicate points	0	0	0	0
Computed experimental error	0	0	0	0
Lack-of-fit points	20	20	20	20
Min expt. error for non-significant LOF (95%)	0.210178	0.229732	0.232257	0.232872

functional theory (DFT) in the Spartan 14 software [16, 17]. The molecular descriptors of the optimized molecules were calculated using the Pharmaceutical Data Exploration Laboratory (PaDEL) version 2.21 software. The dataset of molecular descriptors obtained was pretreated at a correlation cut-off of 0.8 in the DTC Lab Pretreatment software version 1.2 and divided into a training and test set using the Kennard-Stone algorithm in DTC Lab Dataset Division software version 1.2. The training set was transferred to the Accelery Material Studio Version 8.0 software where the model was built while the test set was subsequently employed for external validation of the built model [18, 19].

2.3 Model validation
The validity, robustness, and predicting ability of a built model are ascertained by subjecting the model to certain validation tests. The internal consistency and validation of the model were ascertained using the coefficient of determination (R^2 and $R^2_{adj.}$) and cross-validated coefficient of determination (Q^2_{cv}). A robust model has values ≥ 0.7 for R^2 and ≥ 0.6 for Q^2_{cv} [20]. R^2 is a measure of the variation in the activity/toxicity of the molecules that can be explained by the model. Thus, a robust model should be able to explain at least 0.7 (70%) of the variation in the activity/toxicity of the compounds. Q^2_{cv} measures the degree to which the model can generalize

to another independent dataset of phenylpiperazine derivatives [21]. R^2, $R^2_{adj.}$, and Q^2_{cv} alongside other statistical parameters such as Friedman's lack of fit (LOF) and significance-of-regression *F*-value were automatically generated by the Material Studio software while the model was built. The reproducibility of the model on other independent datasets was evaluated by subjecting the model on an independent dataset. The external coefficient of determination ($R^2_{ext.}$) was determined using Eq. 1 [22].

$$R^2_{\text{ext.}} = 1 - \frac{\sum \left(Y_{\text{exp}_{\text{test}}} - Y_{\text{pred}_{\text{test}}} \right)^2}{\sum \left(Y_{\text{exp}_{\text{test}}} - \overline{Y}_{\text{exp}_{\text{test}}} \right)^2} \tag{1}$$

where Y_{exptest} is the experimental activity/toxicity of each test set molecule, Y_{predtest} is the predicted activity/toxicity of each test set molecule, and $\overline{Y}_{\text{exp}_{\text{train}}}$ is the mean activity/toxicity of the training set compounds. A model with good predicting power has $R^2_{\text{ext.}} \geq 0.6$ [20].

The inter-correlation between the molecular descriptors was also evaluated using Pearson's correlation and variance inflation factor (VIF) tests [18]. To ensure that each molecular descriptor made a unique contribution to the prediction of the activity/toxicity, the descriptors ought to be poorly correlated ($R < 0.4$) and have VIF values less than 10. The VIF of each molecular descriptor was calculated as $(1 - R^2)^{-1}$ [18, 22]. The mean effect

Table 2 Description and class of molecular descriptors in the built model

Name	Description	Class
VR3_Dzp	Logarithmic Randic-like eigenvector-based index from Barysz matrix/weighted by polarizabilities	2D
VE3_Dzi	Logarithmic coefficient sum of the last eigenvector from Barysz matrix/weighted by first ionization potential	2D
Kier3	Third kappa (κ) shape index	2D
RHSA	THSA/total molecular surface area	3D
RDF55v	Radial distribution function - 055/weighted by relative van der Waals volumes	3D

Table 3 Correlation, VIF, and mean effect of molecular descriptors

	VR3_Dzp	VE3_Dzi	Kier3	RHSA	RDF55v	VIF	Mean effect
VR3_Dzp	1					2.1067	1.6178
VE3_Dzi	− 0.40427	1				1.4974	− 0.4341
Kier3	0.397934	0.124831	1			1.5622	− 1.8079
RHSA	− 0.1099	− 0.02413	− 0.23632	1		1.0597	1.7549
RDF55v	0.355327	0.032944	− 0.05608	− 0.00731	1	1.3387	− 0.1307

of each molecular descriptor was also calculated using Eq. 2. The mean effect reveals the descriptors which have the highest positive (or negative) impact on the activity/toxicity [21].

$$\text{Mean effect} = \frac{\beta_j \sum_i^n D_j}{\sum_j^m \left(\beta_j \sum_i^n D_j \right)} \qquad (2)$$

where β_j is the jth descriptor's coefficient in the regression model, D_j is the value of the jth descriptor for each molecule in the training set, m is the number of molecular descriptors in the regression model, and n is the size of the training set.

William's plot of the applicability domain was also drawn using the leverage technique (Eq. 3) [19]. The applicability domain is the surface space in which the model makes reliable predictions for the activity/toxicity of the compounds [23]. Compounds in the dataset which fell within the applicability domain can be reliably employed in further computer aided compound design such as ligand based design.

$$H_j = x_j \left(X^T X \right)^{-1} x_j^T \qquad (3)$$

where H_j is the leverage of the jth compound, x_j is a $1 \times m$ row matrix of the m molecular descriptors of compound j, and X is an $m \times k$ matrix made up of m row descriptor values and k columns of training set values. The boundary of the applicability domain was evaluated as h* = 3(m + 1)/k.

2.4 Molecular docking

Molecular docking studies were carried out using Biovia's Discovery Studio 2016 client and the Autodock Vina integration of the PyRx-Python Prescription 0.8 software. The crystal structure of the androgen receptor (PBD code: 5T8E) was downloaded from the protein data bank [24]. The downloaded receptor was prepared on the Discovery Studio software. To prepare the receptor, water molecules, heteroatoms, and cofactors were removed from the receptor. Plate 1 presents a 3D crystal structure of the prepared receptor. 3D optimized structures of compounds 25 and 32 (Fig. 1) were prepared for

Table 4 External validation of the built model

Molecule	$Y_{\text{exp.}}$	$Y_{\text{pred.}}$	$(Y_{\text{exp.}} - Y_{\text{pred.}})$	$(Y_{\text{exp.}} - Y_{\text{pred.}})^2$	$(Y_{\text{exp.}} - \overline{Y}_{\text{train.}})$	$(Y_{\text{exp.}} - \overline{Y}_{\text{train.}})^2$
1	5.0079	4.8446	0.1633	0.02667	0.0045	0.000020
6	4.9952	4.8272	0.1680	0.02822	− 0.0082	0.000067
7	5.1938	4.9735	0.2203	0.04853	0.1904	0.036252
8	4.5533	4.6312	− 0.0779	0.00607	− 0.4501	0.202590
16	5.0362	4.8410	0.1952	0.03810	0.0328	0.001076
17	5.2708	4.6867	0.5841	0.34117	0.2674	0.071503
26	5.0585	5.0349	0.0236	0.00056	0.0551	0.003036
29	6.0809	5.5371	0.5438	0.29572	1.0775	1.161006
31	5.6498	5.1371	0.5127	0.26286	0.6464	0.417833
35	6.0655	5.8044	0.2611	0.06817	1.0621	1.128056
37	4.6763	5.0227	− 0.3464	0.11999	− 0.3271	0.106994
	$\sum (Y_{\text{exp.}} - Y_{\text{pred}})^2 = 1.23606$			$\sum (Y_{\text{exp.}} - \overline{Y}_{\text{train}})^2 = 3.128433$		
	$\therefore R^2_{\text{ext}} = 1$ $- \frac{1.23606}{3.128433}$ $= 0.6049$					
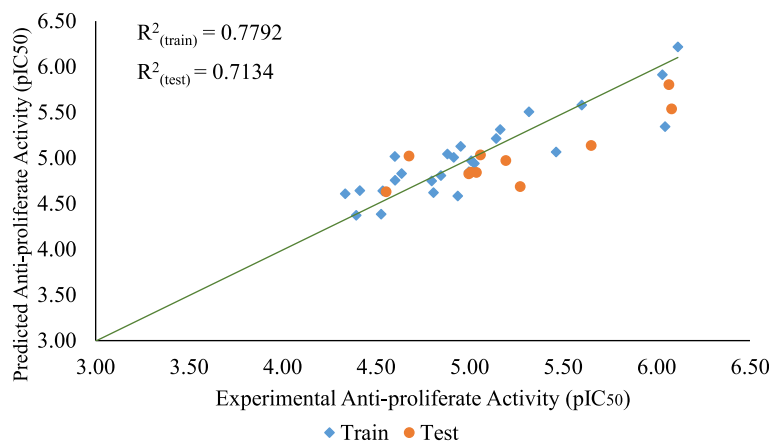
Fig. 3 Experimental activity against predicted activity

molecular docking by converting them to .pdb format on the Spartan 14 software [18, 22]. The binding affinity of the compound and the androgen receptor was determined using the PyRx software while the category and type of interaction between the compounds and the receptor were viewed on the Discovery Studio software.

3 Results

3.1 QSAR

The compounds presented in Fig. 1 were employed to build a QSAR model using the Genetic Function Algorithm–Multilinear Regression (GFA-MLR) method. Four models were built each consisting of five molecular descriptors. Model 1 was adopted as the best model because it had statistical parameters similar to those reported for a robust model [20]. The regression equation of model 1 is presented in Eq. 4. Its statistical parameters are presented in Table 1 while the definition and class of its molecular descriptors are presented in

Table 2. The regression equations of the other QSAR models built are presented in Supplementary Table S1. Table 3 presents the correlation between the molecular descriptors in the model as well as their VIF and mean effect values. The external validation of the model was investigated by subjecting the model to the test set. The external coefficient of determination ($R^2_{ext.}$) was calculated as presented in Table 4.

$$
\begin{aligned}
Model\ 1: pIC_{50} = &-0.298912928 \\
&\times VR3\_Dzp - 0.171389530 \\
&\times VE3\_Dzi + 0.961235807 \\
&\times Kier3 - 6.822313106 \\
&\times RHSA + 0.120746172 \\
&\times RDF55v + 8.757696717 \qquad (4)
\end{aligned}
$$

THSA sum of solvent accessible surface areas of atoms with absolute value of partial charges less than 0.2
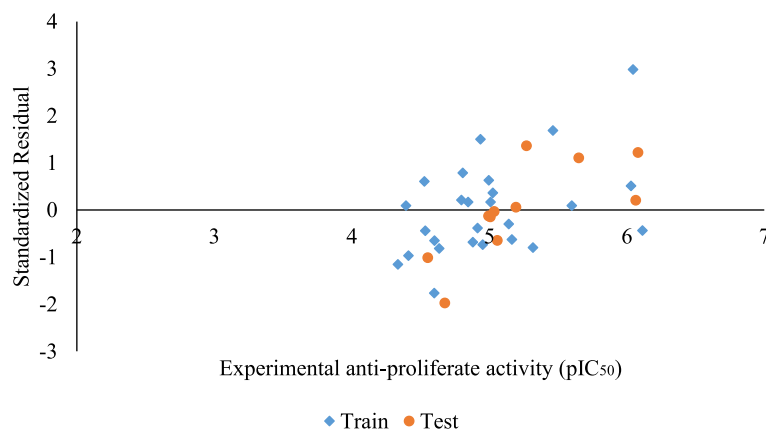
Fig. 4 Experimental activity against standardized residuals
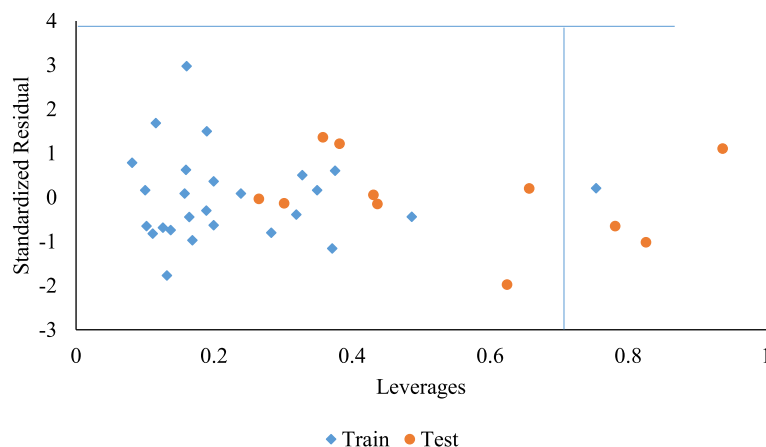
Fig. 5 William's plot of the applicability domain

NB, Y_{train} was evaluated to be 5.0034

Equation 1 was used to predict the anti-proliferate activity of the compounds in the training and test set. The predicted activity and residual of each compound in the training and test sets are presented in Supplementary Tables S2 and S3. Figure 3 presents a graph of the observed experimental activity against the predicted activity. Figure 4 is a graph of the observed anti-proliferate activity against the standardized residual, and Fig. 5 presents William's plot of the applicability domain of the built model.

3.2 QSTR
The toxicity of the compounds against normal prostate epithelial cells was also modeled by building four GFA-MLR models. Each model consisted of four molecular descriptors. The first model had statistical parameters similar to those reported for a stable, robust model [20]. Equation 5 presents the regression equation of the model. Supplementary Table S4 presents the regression equation of the other QSTR models built. Table 5 presents the statistical parameters of the built models while Table 6 presents the description and class of molecular descriptors in model 1. Table 7 presents the results of correlation studies, VIF, and mean effect of the molecular descriptors. The model was used to predict the toxicity of the compounds in the training set (Table S5) and test set (Table S6). The external validation ($R^2_{ext.}$) of the model was also calculated (Table S7) and was equal to 0.6344.

$$
\begin{aligned}
Model\ 1 : pIC_{50} = &-5.247227835 \\
&\times MATS8c - 2.264116018 \\
&\times MATS3s + 3.344208710 \\
&\times ETA\_EtaP\_F - 0.082468094 \\
&\times RDF95m + 2.449569658 \quad (5)
\end{aligned}
$$

Figure 6 presents a graph of the experimental activity against the predicted toxicity of the compounds. Figure 7

Table 5 Statistical parameters of built models

Parameter	Model 1	Model 2	Model 3	Model 4
Friedman LOF	0.081427	0.161877	0.176692	0.178387
R-squared	0.865187	0.731993	0.707465	0.704658
Adjusted *R*-squared	0.831483	0.664991	0.634331	0.630823
Cross validated *R*-squared	0.778806	0.550227	0.438354	0.483458
Significant regression	Yes	Yes	Yes	Yes
Significance-of-regression *F*-value	25.67064	10.92497	9.673568	9.543635
Critical SOR *F*-value (95%)	3.055818	3.055818	3.055818	3.055818
Replicate points	0	0	0	0
Computed experimental error	0	0	0	0
Lack-of-fit points	16	16	16	16
Min expt. error for non-significant LOF (95%)	0.104609	0.147494	0.154096	0.154833

Table 6 Description and class of molecular descriptors in model 1

Descriptor	Description	Class
MATS8c	Moran autocorrelation - lag 8/weighted by charges	2D
MATS3s	Moran autocorrelation - lag 3/weighted by I-state	2D
ETA_EtaP_F	Functionality index EtaF relative to molecular size	2D
RDF95m	Radial distribution function - 095/weighted by relative mass	3D

is a graph of the standardized residuals against experimental toxicity while Fig. 8 presents the domain of applicability of the built model.

3.3 Molecular docking

Molecular docking studies were carried out to investigate the binding affinity and types of interactions between compounds 25 and 32 (Fig. 1) and the androgen receptor. The binding affinity, category, and type of interaction between each compound and the receptor are presented in Table 8. Plate 2 presents the 2D interaction between compound 25 and the receptor while Plate 3 presents the 2D interaction between compound 32 and the receptor.

4 Discussion

In silico methods are invaluable computer aided techniques employed in obtaining and optimizing potential drug leads and candidates. QSAR is a variant of the Quantitative Structure Property Relationship (QSPR) approach. QSAR models the activity of a set of compounds as a linear combination of certain molecular descriptors. Molecular descriptors are numbers that describe certain molecular properties of a compound [5, 25]. A robust QSAR model employs molecular descriptors which significantly affect the activity of the compounds. A QSAR model can be employed in predicting the activity of other similar compounds [18]. The built QSAR model had internal validation parameters ($R^2 = 0.7792$, $R^2_{adj.} = 0.7240$, $Q_{cv}^2 = 0.6607$) which are similar to those reported for a robust model [20]. The square of the coefficient of

determination (R^2 and $R^2_{adj.}$) is a measure of the variation in the activity of the compounds which can be explained by the model. The built model explained at least 70% (0.7) of the variation in the activity of the compounds. Q_{cv}^2 as earlier defined measures the degree to which the built model can generalize over another independent dataset of phenylpiperazine compounds [21]. The model built had at least a 66% (0.66) probability of generalizing over any dataset of phenylpiperazine compounds. The external coefficient ($R^2_{ext.}$) of 0.6049 obtained is similar to that reported for a model with high predicting capacity [20]. A robust model is characterized by poorly correlated molecular descriptors and VIF values less than 10.0 [18]. Table 3 presented the correlation between the molecular descriptors in the model as well as their VIF and mean effect values. The descriptors were observed to be poorly correlated ($R < 0.42$). The VIF values obtained (< 2.5) revealed that descriptors were poorly correlated and as such, each molecular descriptor can be considered to make a significant contribution to predicting the activity of the compounds. Mean effect studies (Table 3) revealed that the molecular descriptors RHSA and VR3_Dzp had the highest significant positive effect on the activity of the compounds while Kier3 and VE3_Dzi had the highest negative impact on the activity of the compounds. The coefficient of determination R^2_{train} and R^2_{test} obtained (Fig. 3) highlights the robustness and predicting power of the model while the random spread of the standardized residuals (Fig. 4) indicates the absence of systematic errors [19]. The domain of applicability of the model is the surface space where predictions made by the model can be reliably employed for further theoretical studies or experimental applications [23]. The domain of applicability (Fig. 5) revealed the presence of four outliers (compounds 2, 8, 26, and 31).

QSTR is also a variant of the QSPR approach, and it models the toxicity of compounds as a linear combination of molecular descriptors. QSTR models are built on the same premise as QSAR models. The QSTR model built explained at least 80% (R^2 and $R^2_{adj.} > 0.8$) of the variation in the toxicity of the

Table 7 Correlation, VIF, and mean effect of molecular descriptors

|	MATS8c	MATS3s	ETA_EtaP_F	RDF95m	VIF	Mean effect
MATS8c	1				1.985406	− 0.11295
MATS3s	− 0.67158	1			2.263527	0.012116
ETA_EtaP_F	− 0.13679	0.380611	1		1.217369	1.304967
RDF95m	0.117553	0.06594	− 0.02682	1	1.065157	− 0.20414
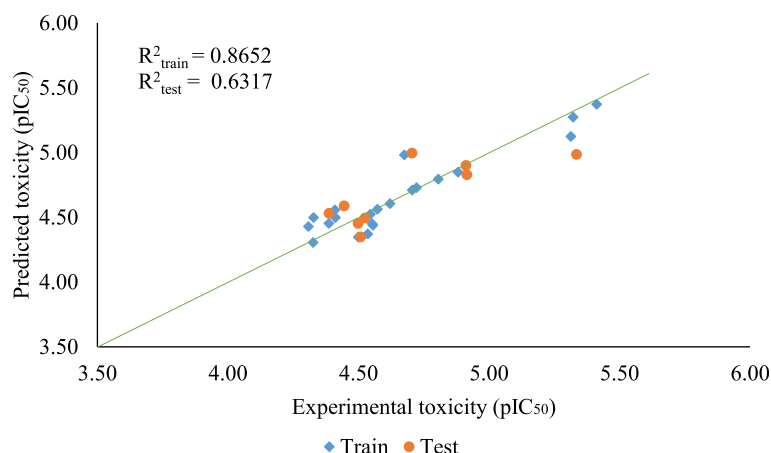
Fig. 6 Experimental toxicity against predicted toxicity

model. The cross validated R-squared (Q^2_{cv} = 0.7788) and external coefficient of determination ($R^2_{ext.}$ = 0.6344) obtained were similar to those reported for a robust model [20]. The molecular descriptors were observed to be poorly correlated, having VIF values less than 2.0. The molecular descriptor ETA_EtaP_F was observed to have the highest positive impact on the toxicity of the compounds. Figure 6 showed the built QSTR model to have a high predicting ability, posing R^2_{train} and R^2_{test} values of 0.8652 and 0.6317. The model was also observed to have no systematic error (Fig. 7), and the domain of applicability plot (Fig. 8) revealed the presence of only one outlier (compound 4). The QSAR and QSTR models built can be employed in ligand based compound design to design novel phenylpiperazine compounds with better anti-proliferative activity and less toxicity.

Molecular docking is an in silico approach that investigates the binding interaction between a ligand and receptor. A receptor is a macromolecule, usually an enzyme, biological receptor, tissue, and so on. Molecular docking is primarily concerned with the binding affinity (or energy) and type of interaction between the receptor and ligand [26, 27]. Findings from molecular docking studies are employed in in silico compound design via structure based methods. The binding affinity and types of interactions between compounds 25 and 32 and the androgen receptor were presented in Table 8. Compound 32 was observed to form more non-bonding interactions with the androgen receptor, and this accounts for its relatively higher binding affinity (– 7.00 kcal/mol). It was observed to have interactions with threonine (THR755), arginine (ARG752) proline (PRO801), and phenylalanine (PHE754) amino acid residues of the receptor and compound 25 on the other hand for interactions with glutamic acid (GLU772),
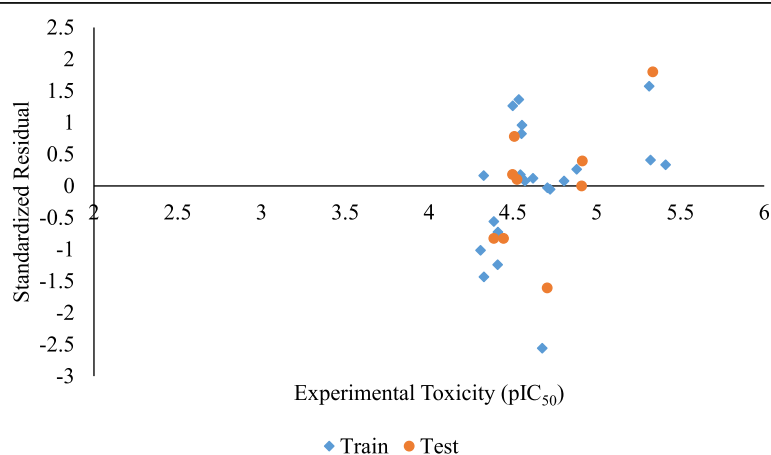

**Fig. 7** Standardized residuals against experimental toxicity
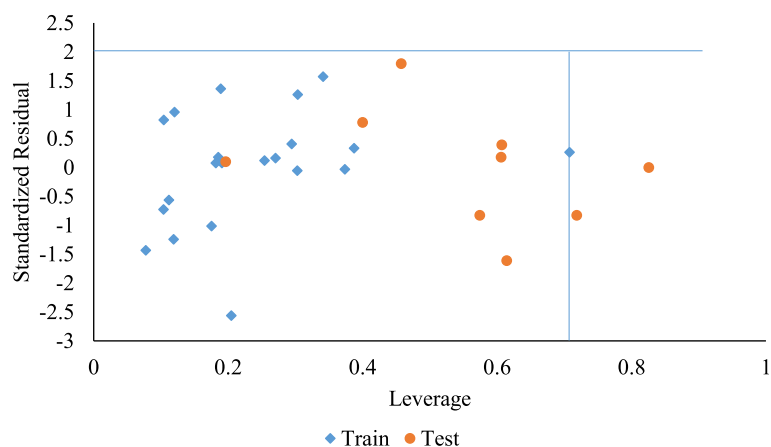
Fig. 8 William's plot of the applicability domain

Table 8 Binding affinity and interaction between compounds 25 and 32 and the androgen receptor

Molecule	Binding affinity (kcal/mol)	Interaction		Amino acid residue	Distance (Å)
		Category	Type		
25	− 6.40	Halogen	Halogen (fluorine)	GLU772	3.13077
		Hydrophobic	Pi-pi stacked	TYR781	3.79051
			Pi-alkyl	ARG779	4.90829
32	− 7.00	Hydrogen bond	Conventional hydrogen bond	THR755	2.21195
		Hydrophobic	Pi-sigma	THR755	3.55928
			Alkyl	ARG752, PRO801	4.57691, 5.43965
			Pi-alkyl	PHE754, ARG752	4.98372, 4.33078
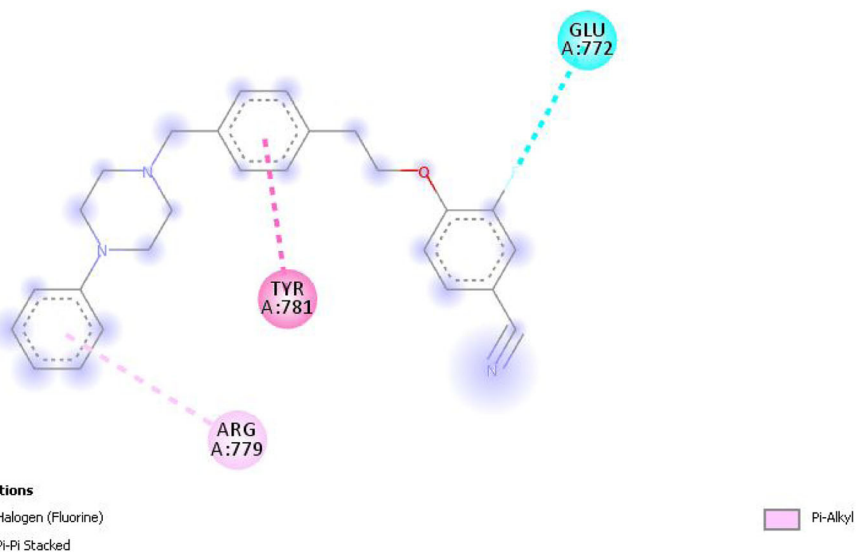
Plate 2 2D structure of the interaction of compound 25 and the androgen receptor (PDB ID: 5T8E)

Plate 3 2D structure of the interaction of compound 32 and the androgen receptor (PDB ID: 5T8E)
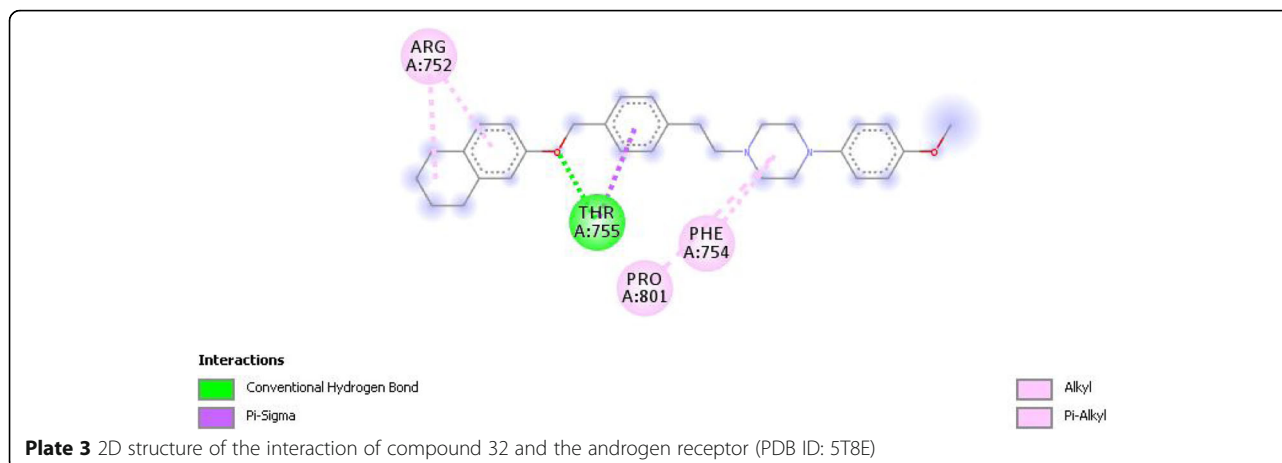
tyrosine (TYR781), and arginine (ARG779) protein residues. Furthermore, it was observed that the hydrogen bond interaction (bond distance = 2.21195 Å) had a greater stabilizing effect compared to the halogen interaction (bond distance = 3.13077 Å). Shorter bonds hold atoms closer to each other and as such are stronger than longer ones. The findings from the molecular docking studies can be employed in the in silico design of novel phenylpiperazine compounds via structure based design.

The compounds employed in this study have been reported to show in vitro cytotoxic activity against the DU145 prostate cancer cell lines [12, 13]. The regression models built revealed the molecular descriptors which significantly affect the cytotoxicity of the compounds. Furthermore, the models provide a platform upon which novel phenylpiperazine compounds can be designed via the ligand based design approach.

5 Conclusion

This study employed variants of the Quantitative Structure Property Relationship approach for the in silico studies of some phenylpiperazine compounds. The variants employed were QSAR and QSTR techniques. The study built two robust models to predict the toxicity and anti-proliferative activity of the phenylpiperazine compounds against normal prostate epithelial cells and the DU145 prostate cancer cell lines. Both models were built using the Multilinear Regression–Genetic Function Algorithm method available on the Biovia's Material Studio version 8.0 software. Both models had statistical parameters: $R^2 > 0.7$, $R^2_{adj.} > 0.7$, $Q^2_{cv} > 0.6$, $R^2_{ext} > 0.6$, $R^2_{train} > 0.7$, and $R^2_{test} > 0.7$ which were similar to those reported for robust models. The activity of the compounds was revealed to be strongly dependent on the molecular descriptors VR3_Dzp, VE3_Dzi, Kier3, RHSA, and RDF55v. The toxicity of the compounds, on the other hand, was observed to be strongly dependent on the molecular descriptors

MATS8c, MATS3s, ETA_EtaP_F, and RDF95m. Molecular docking studies were also carried out between compounds 25 and 32 and the androgen receptor. Molecular docking studies revealed that compound 25 formed halogen and hydrophobic interactions with a binding affinity of − 6.40 kcal/mol while compound 32 form hydrogen and hydrophobic bond interaction with a binding affinity of − 7.00 kcal/mol with the receptor. This study provides insight on the activity and toxicity of phenylpiperazine compounds and also provides information that can be employed in in silico design of other phenylpiperazine compounds via ligand based and/or structure-based design methodology.

6 Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s43088-020-00054-y.

Additional file 1: Table S1. Regression equations of other QSAR models built. **Table S2.** Predicted activity and residual of Molecules in the QSAR Train set. **Table S3.** Predicted activity and residual of Molecules in the QSAR Test set. **Table S4.** Regression equations of other QSTR models built. **Table S5:** Predicted activity and residual of Molecules in the QSTR Train set. **Table S6.** Predicted activity and residual of Molecules in the QSTR Test set. **Table S7.** External Validation of QSTR Model

Abbreviations
GFA: Genetic Function Approximation; PaDEL: Pharmaceutical Data Exploration Laboratory; QSAR: Quantitative Structure Activity Relationship; QSTR: Quantitative Structure Toxicity Relationship; QSPR: Quantitative Structure Property Relationship

Authors' contributions
IAF: conceptualization, methodology, investigation, data curation, writing—original draft preparation. SAG: conceptualization, writing—review and editing, supervision. MAP: conceptualization, writing—review and editing, supervision. All authors also read and approved the final manuscript.

References
1. Cheesman MJ, Ilanko A, Blonk B, Cock IE (2017) Developing new antimicrobial therapies: are synergistic combinations of plant extracts/compounds with conventional antibiotics the solution? Pharmacognsy Rev 11(22):57–72. https://doi.org/10.4103/phrev.phrev_21_17
2. Sofowora A, Ogunbodede E, Onayade A (2013) The role and place of medicinal plants in the strategies for disease prevention. Afr J Tradit Complement Altern Med 10(5):210–229. https://doi.org/10.4314/ajtcam.v10i5.2
3. Reynolds, T., Wessel, M., Konagurthu, S. and Crew, M. (2016). Computational methods – formulation development: an innovative, simulation – based approach. Drug Development and Delivery, September 2016. https://drug-dev.com/computational-methods-formulation-development-an-innovative-simulation-based-approach/ (Accessed 26 Feb 2020)
4. Batool, M., Ahmad, B. and Choi, S. (2019). A structure-based drug discovery paradigm. Int J Mol Sci, 20(11), 2783. https://doi.org/https://doi.org/10.3390/ijms20112783
5. Sliwoski, G., Kothiwale, S., Meiler, J. and Lowe, Jr. E. W. (2014). Computational methods in drug discovery. Pharmacol Rev, 66(1): 334–395. https://dx.doi.org/https://doi.org/10.1124/pr.112.007336
6. White, L., Duncan, G. and Baumle, W. (2013). Medical surgical nursing: an integrated approach, 3rd Ed. Australia: Delmar Cengage Learning. ISBN-13: 978-1-4354-8802-1. p.134
7. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R., Torre, L. and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin, 68(6), 394 – 424. https://doi.org/https://doi.org/10.3322/caac.21492
8. Rawla, P. (2019). Epidemiology of prostate cancer. World J Oncol, 10(2): 63–89. https://dx.doi.org/https://doi.org/10.14740/wjon1191
9. Taitt, H. E. (2018). Global trends and prostate cancer: a review of incidence, detection, and mortality as influenced by race, ethnicity, and geographic location. Am J Mens Health, 12(6): 1807–1823. https://dx.doi.org/https://doi.org/10.1177/1557988318798279
10. Pulukuri SM, Gondi CS, Lakka SS, Jutla A, Estes N, Gujrati M, Rao JS (2005) RNA interference-directed knockdown of urokinase plasminogen activator and urokinase plasminogen activator receptor inhibits prostate cancer cell invasion, survival, and tumorigenicity in vivo. J Biol Chem 280(43):36529–36540. https://doi.org/10.1074/jbc.M503111200
11. Recouvreux, M. V., Wu, J. B., Gao, A. C., Zonis, S., Chesnokova, V., Bhowmick, N. Chung, L. W. and Melmed, S. (2017). Androgen receptor regulation of local growth hormone in prostate cancer cells. Endocrinology, 158(7), https://doi.org/https://doi.org/10.1210/en.2016-1939
12. Chen, H., Wang, C., Sun, T., Zhou, Z., Niu, J., Tian, X. and Yuan, M. (2018a). Synthesis, biological evaluation and SAR of naftopidil-based arylpiperazine derivatives. Bioorganic Med Chem Lett, 28, 1534-1539 https://doi.org/https://doi.org/10.1016/j.bmcl.2018.03.070
13. Chen, H., Yu, Y., Tian, X., et al. (2018b). Synthesis and biological evaluation of arylpiperazine derivatives as potential anti-prostate cancer agents. Bioorganic Med Chem https://doi.org/https://doi.org/10.1016/j.bmc.2018.11.029
14. Chen H, Liang X, Xu F, Xu B, He X, Huang B, Yuan M (2014) Synthesis and cytotoxic activity evaluation of novel arylpiperazine derivatives on human prostate cancer cell lines. Molecules 19:12048–12064. https://doi.org/10.3390/molecules190812048
15. Rajput, A., Kumar, A. and Kumar, M. (2019). Computational identification of inhibitors using QSAR approach against Nipah virus. Front Pharmacol, https://doi.org/https://doi.org/10.3389/fphar.2019.00071
16. Becke, A. D. (1993). Becke's three parameter hybrid method using the LYP correlation functional. J Chem Phys, 98, 5648 - 5652. https://doi.org/https://doi.org/10.1063/1.464913
17. Farag, A. M. and Fahimb, A. M. (2019). Synthesis, biological evaluation and DFT calculation of novel pyrazole and pyrimidine derivatives. J Mol Struct, 1179. https://doi.org/https://doi.org/10.1016/j.molstruc.2018.11.008
18. Adeniji, S. E., Uba, S., Uzairu, A. and Arthur, D. E. (2019) A derived QSAR model for predicting some compounds as potent antagonist against mycobacterium tuberculosis: a theoretical approach. Adv Prev Med https://doi.org/https://doi.org/10.1155/2019/5173786
19. Isyaku, Y., Uzairu, A. and Uba, S. (2019). QSAR study of 2-substituted phenyl-2-oxo-, 2-hydroxyl- and 2-acylloxyethylsulfonamides as fungicides. J Eng Exact Sci, 5(3). https://doi.org/https://doi.org/10.18540/jcecvl5iss3pp0283-0290
20. Tropsha, A., 2010. Best practices for QSAR model development, validation and exploitation. Mol Inform, 29 (6-7), 476–488. https://doi.org/https://doi.org/10.1002/minf.201000061
21. Ikwu, F. A., Shallangwa, G. A., Paul, P. A. and Uzairu, A. (2020). In silico studies of piperazine derivatives as potent anti-proliferative agents against PC-3 prostate cancer cell lines. Heliyon, 6. https://doi.org/https://doi.org/10.1016/j.heliyon.2020.e03273
22. Abdullahi, M., Uzairu, A., Shallangwa, G. A., Paul, A. P., Arthur, D. E. and Ibrahim, M. T. (2019). An insilico modelling study on some C14-urea-Tetrandrine derivatives as potent anti-cancer against prostate (PC3) cell line. J King Saud Univ https://doi.org/https://doi.org/10.1016/j.jksus.2019.01.008
23. Netzeva, T.I., Worth, A., Aldenberg, T., Benigni, R., et al., (2005). Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM workshop 52. Altern LabAnim 33 (2), 155–173. https://doi.org/https://doi.org/10.1177/026119290503300209
24. Asano M, Hitaka T, Imada T, Yamada M, Morimoto M et al (2017) Synthesis and biological evaluation of novel selective androgen receptor modulators (SARMs). Part II: optimization of 4-(pyrrolidin-1-yl)benzonitrile derivatives. Bioorg Med Chem Lett 27:1897–1901. https://doi.org/10.1016/j.bmcl.2017.03.038
25. Fukunishi, Y., Yamasaki, S., Yasumatsu, I., Takeuchi, K., Kurosawa, T. and Nakamura, H. (2017). Quantitative structure-activity relationship (QSAR) models for docking score correction. Mol Inform, 36(1-2): 1600013. https://dx.doi.org/https://doi.org/10.1002/minf.201600013
26. Nnyigide, O. S., Lee, S. and Hyun, K. (2019). In silico characterization of the binding modes of surfactants with bovine serum albumin. Sci Rep, 9, 10643. https://dx.doi.org/https://doi.org/10.1038/s41598-019-47135-2
27. Ramsay, R., Popovic-Nikolic, M., Nikolic, K., Uliassi, E. and Bolognesi, M. (2018). A perspective on multi-target drug discovery and design for complex diseases. Clin Transl Med, 7, 3. https://dx.doi.org/https://doi.org/10.1186/s40169-017-0181-2

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.