RESEARCH

A combined 2-D and 3-D QSAR modeling, molecular docking study, design, and pharmacokinetic profiling of some arylimidamide-azole hybrids as superior *L. donovani* inhibitors

Fabian Audu Ugbe*[iD], Gideon Adamu Shallangwa, Adamu Uzairu and Ibrahim Abdulkadir

Abstract

Background: Leishmaniasis is one of the neglected tropical diseases which is prevalent in the tropical regions of the world most especially in Africa. It is caused by the *Leishmania* species and transmitted to humans majorly through the bite of the female sandfly. This study was carried out in support of the continuous search for new drug molecules effective enough for the treatment of leishmaniasis, and which have very limited side effects. This study was focused on a combined 2-D and 3-D QSAR modeling of thirty-six arylimidamide-azole hybrids, molecular docking study, design, and pharmacokinetic analysis of some selected and newly designed arylimidamide-azole analogs. The density functional theory (DFT) with B3LYP and 6-31G** basis set was employed for the geometry optimization of the various compounds. Genetic function approximation (GFA) and multi-linear regression (MLR) approaches were used for the 2-D QSAR model building, while the fractional factorial design (FFD) and uninformative variable elimination-partial least square (UVEPLS) were employed for building the 3-D QSAR model. Pyridoxal kinase (PdxK) receptor (PDB: 6K91) was the target protein of interest in this study.

Results: The built 2-D and 3-D QSAR models were found to satisfy the requirement of both internal and external validation tests as follows: 2-D QSAR; $R^2 = 0.9614$, $R^2_{adj} = 0.9513$, $Q^2_{cv} = 0.9350$, $R^2_{test} = 0.6766$ and $cRp^2 = 0.8779$, and for 3-D QSAR (UVEPLS at PC = 5); $R^2 = 0.9839$, $Q^2_{LOO} = 0.7539$ and $Q^2_{LTO} = 0.7367$. The CoMFA steric and electrostatic field contributions were 68.96% and 31.04%, respectively. All the designed analogs showed higher predicted activities than that of the template (**36**). Also, the new compounds showed higher binding affinities (MolDock scores) than that of the reference drug pentamidine (− 141.793 kcal/mol), with **36e** showing the highest negative MolDock score of − 208.595 kcal/mol. Additionally, these newly designed compounds were said to be orally bioavailable (excluding **36f** and **36g** that violated 2 of the Lipinski's provisions), with perfect intestinal absorption, less difficult to synthesize, AMES toxicity free, and showed strong interactions with the target.

Conclusions: The newly designed compounds especially **36e** have shown a marked pharmacological improvement over the template molecule and are therefore recommended for further practical evaluation as superior pyridoxal kinase inhibitors.

*Correspondence: ugbefabianaudu@gmail.com

Department of Chemistry, Faculty of Physical Sciences, Ahmadu Bello University, P.M.B. 1044, Zaria, Kaduna State, Nigeria

Ugbe *et al. Bulletin of the National Research Centre*   (2022) 46:189

Page 2 of 24

Background

Leishmania species is a group of protozoan parasites responsible for leishmaniasis, a neglected tropical disease known to affect a global population of up to 350 million people majorly in the tropical regions (Al-Tamimia et al. 2019; Baquedano et al. 2016). Three common clinical manifestations were earlier reported as cutaneous, mucocutaneous, and visceral leishmaniasis (VL) (Díaz et al. 2019). Among these forms, cutaneous is the most common, while VL is the most fatal (Keurulainen et al. 2018). Specific species causing VL include *Leishmania Donovani* (*L. donovani*) and *L. infantum,* transmitted by the insect vector (female sandfly) (Ugbe et al. 2022a, b). Only a limited number of drugs are readily available for the treatment of leishmaniasis such as pentamidine, amphotericin B, pentostam, paromomycin, and miltefosine, none of which is reported free from associated drug adverse effects, low efficacies, high cost, etc. (Fan et al. 2018). For example, amphotericin B has been linked to fever, nephrotoxicity, myocarditis, chill, and hypokalemia (Ghorbani and Farhoudi 2018). Induction of insulin-dependent diabetes mellitus has been the major drawback of pentamidine, while paromomycin is not free from nephrotoxicity and ototoxicity (Seifert 2011). Also, most organisms had in recent times shown constant resistance to some of these medicines, a case which has greatly threatened the effective treatment of leishmaniasis (Fan et al. 2018). Additionally, the disease has received less attention compared to other infections such as hepatitis, malaria, AIDS, diabetes, tuberculosis, and cancer. As such, it has become imperative to develop new medicines with attributes that overcome all of these shortcomings.

In the continuous search for new effective drug compounds, in silico approaches like the molecular docking studies, quantitative structure–activity relationship (QSAR), homology modeling, molecular dynamics, and pharmacokinetics analysis among others play a very crucial role because it is cost-effective, save time and proves effective than the crude methods (Adeniji et al. 2019; Ugbe et al. 2021). The connection between a compound's structural properties (independent variables) and its biological activities (dependent variable) is usually established by QSAR (Adeniji et al. 2019). Molecular docking on the other hand helps to investigate the binding of ligands with the active sites of the target receptor (Ibrahim et al. 2020). The pharmacokinetic study is a crucial step in the preclinical phase of new drug molecules, which helps to ascertain how such drug compounds affect the living organism when administered (Lawal et al. 2021; Ibrahim et al. 2021). The choice of drug molecules for oral bioavailability has been widely guided by the Lipinski rule of five (RO5) or the Pfizer rule which considers physicochemical properties such as molecular weight (MW), lipophilicity, hydrogen bond donors (HBDs), and hydrogen bond acceptors (HBAs) as necessary factors in predicting drug's likelihood of being orally bioavailable (Lipinski et al. 2001). Also important for evaluation as part of a pharmacokinetics study are absorption, distribution, metabolism, excretion, and toxicity (ADMET) (Ugbe et al. 2022a, b).

Pyridoxal kinase (PdxK) from *L. donovani* (PDB: 6K91) is an interesting enzyme previously reported to catalyze the phosphorylation of the 5′ hydroxyl group of pyridoxal to form pyridoxal-5′-phosphate, an active form of vitamin B6 (Are et al. 2020). PdxK is an important enzyme for parasite growth and key to host infection (Kumar et al. 2018). Chloroquine and primaquine which are well-known anti-malaria drugs were earlier reported to inhibit PdxK (Kimura et al. 2018). Therefore, PdxK serves as a promising protein target for designing new leishmanial inhibitors. Arylimidamide-azole compounds incorporate both arylimidamide and azole moiety in their structures, which were reported to possess the advantage of interacting with the targets of both classes of compounds and which also show improvement in the in vivo pharmacokinetics and/or pharmacodynamics of these classes of molecules (Abdelhameed et al. 2020).

Therefore, this study was focused on a combined 2-D and 3-D QSAR modeling for predicting the activities of some arylimidamide-azole hybrids, performing molecular docking studies, and design of new analogs, while subjecting same to pharmacokinetics analysis to evaluate their drug-likeness and ADMET properties.

Methods

Data acquisition

A series of thirty-six arylimidamide-azole hybrids with reported biological activities (IC$_{50}$ in µM) against *L. donovani* intracellular amastigotes were sourced from the literature (Abdelhameed 2018). The various bioactivity (IC$_{50}$) values were separately converted to pIC$_{50}$ using Eq. (1) (Ugbe et al. 2022a, b). The molecular structures of the various arylimidamide-azole hybrids

Ugbe *et al. Bulletin of the National Research Centre*   (2022) 46:189

Page 3 of 24

Table 1 Molecular structures of arylimidamide-azole hybrids, and pentamidine with their anti-leishmanial activities and MolDock scores

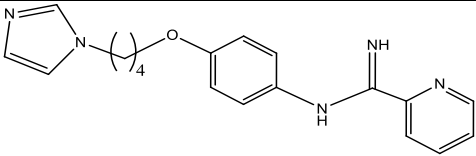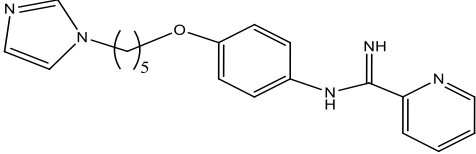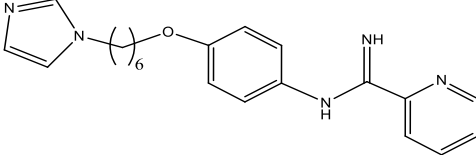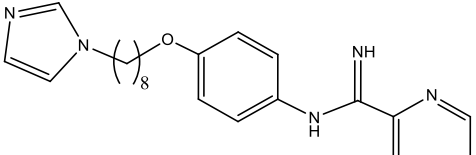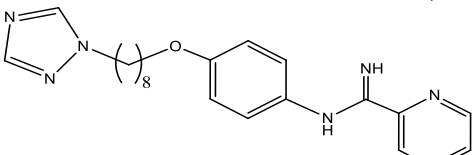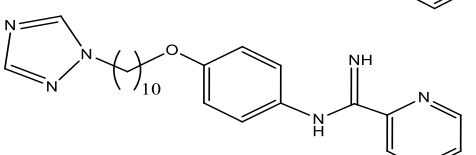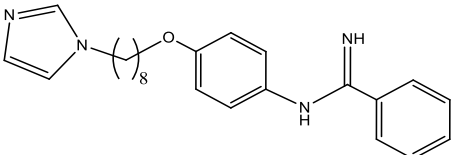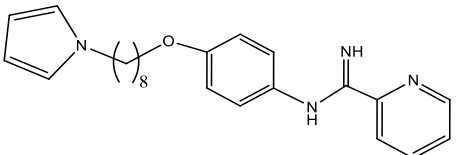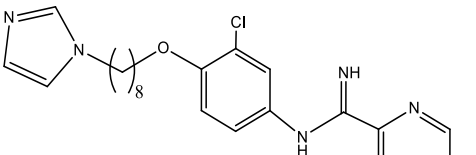| Comp. ID	Molecular structures	IC$_{50}$ (µM)	pIC$_{50}$	MolDock score
1		18	4.7447	−142.632
2		13	4.8861	−143.448
3		12	4.9208	−153.831
4		1.0	6.0000	−154.190
5		13	4.8861	−147.686
6		1.4	5.8539	−165.934
7		1.2	5.9208	−170.065
8		1.7	5.7696	−161.604
9		1.3	5.8861	−155.322

Ugbe *et al. Bulletin of the National Research Centre*   (2022) 46:189

Page 4 of 24
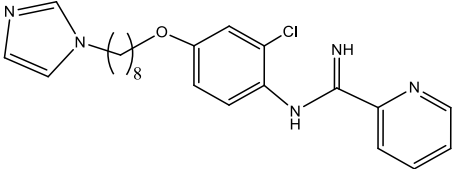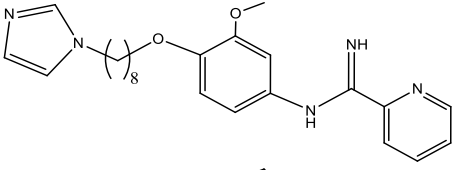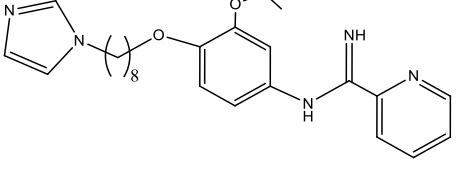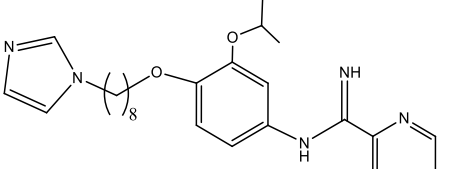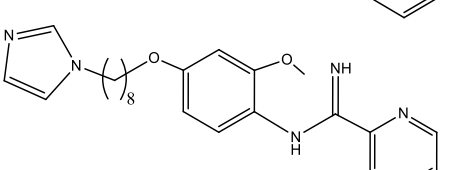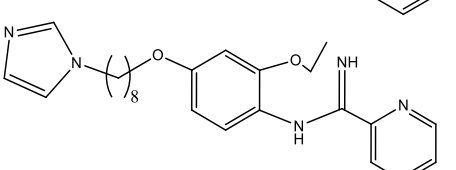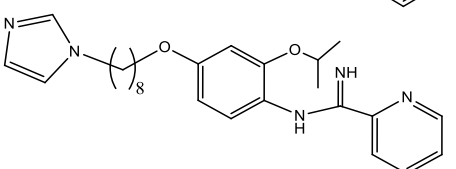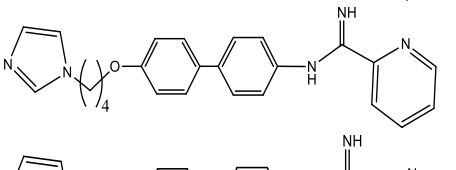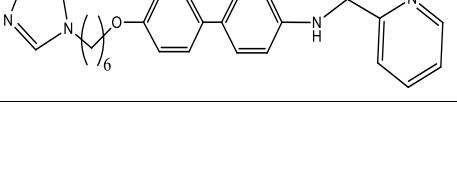
Table 1 (continued)

10		2.1	5.6778	−164.640
11		2.2	5.6576	−167.767
12		1.9	5.7212	−166.814
13		0.97	6.0132	−170.790
14		1.5	5.8239	−162.988
15		0.73	6.1367	−163.342
16		0.58	6.2366	−180.769
17		2.6	5.585	−168.282
18		0.99	6.0044	−173.908

Ugbe *et al. Bulletin of the National Research Centre*   (2022) 46:189

Page 5 of 24
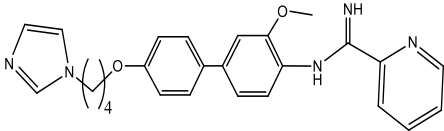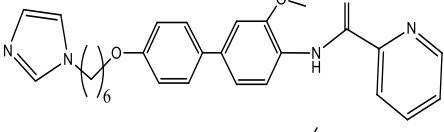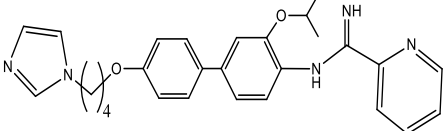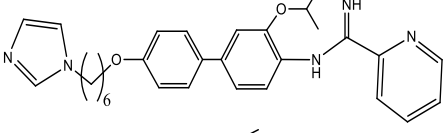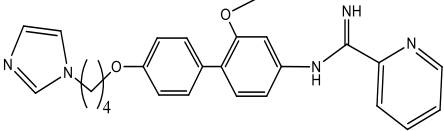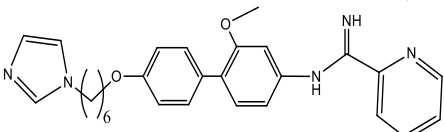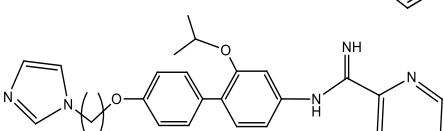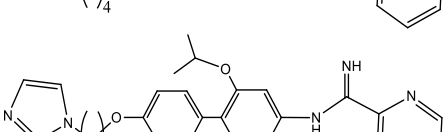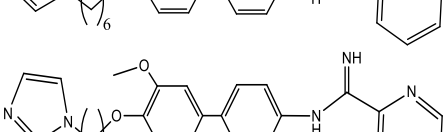
Table 1 (continued)

19		0.82	6.0862	−178.347
20		0.6	6.2218	−174.564
21		0.33	6.4815	−184.502
22		0.29	6.5376	−179.854
23		2.2	5.6576	−180.292
24		1.5	5.8239	−174.043
25		1.2	5.9208	−176.488
26		0.15	6.8239	−189.128
27		3	5.5229	−169.712

Table 1 (continued)

28		1.4	5.8539	−174.328
29		1.2	5.9208	−178.858
30		0.69	6.1612	−187.077
31		0.46	6.3372	−181.748
32		0.78	6.1079	−177.551
33		0.19	6.7212	−172.402
34		0.77	6.1135	−183.663
35		0.94	6.0269	−188.104
36		0.49	6.3098	−198.260
Pentamidine		−	−	−141.793

IC$_{50}$-half-maximal inhibitory concentration, pIC$_{50}$-negative log of IC$_{50}$

with their observed anti-leishmanial activities, including the molecular structure of the reference drug (pentamidine), are shown in Table 1.

$$pIC_{50} = -\log_{10}(IC_{50} \times 10^{-6}) \tag{1}$$

Ugbe *et al. Bulletin of the National Research Centre* (2022) 46:189

Page 7 of 24

Hardware and software

The hardware used was an HP laptop computer with the following specifications: Processor (Intel(R) Core(TM) i5-4210U CPU @ 1.70 GHz 2.40 GHz), Installed RAM (8.00 GB), System type (64-bit operating system, $\times$ 64-based processor), Edition (Windows 10 Home Single Language), Version 21H2.

Software used include ChemDraw Ultra v. 12.0.2, Spartan'14 v. 1.1.4, PaDEL-descriptor v. 2.20, and Drug Theoretics and Cheminformatics Laboratory (DTC-Lab)-based software. Others include Material Studio v. 8.0, Open3Dtools, Maestro v. 12.3, Biovia Discovery Studio Visualizer v. 16.1.0.15350, and Molegro Virtual Docker v. 6.0, a product of the A CLC bio company.

Molecular geometry minimization

The molecular structures of all the compounds were drawn using the ChemDraw Ultra, saved as MDL mol-file format, and thereafter imported separately onto the Spartan'14 Graphical User Interface while enabling the auto conversion of 2-D models to 3-D. The imported molecules were initially subjected to energy minimization and then saved in Spartan file format. The resulting structures were then fully optimized first by using molecular mechanics force field (MMFF) and thereafter density functional theory (DFT) with Becke's three-parameter read-Yang–Parr hybrid (B3LYP) option and utilizing the 6-31G** basis set. The optimized structures were then saved as SD files and PDB formats for subsequent use in descriptor generation and docking, respectively (Wang et al. 2020; Ugbe et al. 2021).

2-D QSAR model building

The 2-D QSAR model building began with the generation of molecular descriptors for all thirty-six compounds by feeding the various molecules in SD file format into the PaDEL-Descriptor software (Lawal et al. 2021). Data pretreatment using the DTC-Lab pretreatment tool (GUI 1.2) was carried out to remove uninformative descriptors from the generated descriptors pool (Adeniji et al. 2020). With the aid of the DTC-Lab data division tool which uses the Kennard Stone method, the resulting data were then partitioned into a training set and test set data in the 70:30 ratio, respectively (Kennard and Stone 1969). The Material studio software was used for the model building by employing the genetic function approximation (GFA) to obtain the optimum descriptor combination which is contained in the QSAR model equations built based on the multi-linear regression (MLR) approach. MLR establishes the relationship between the dependent variable (pIC_{50}) and the independent variables (molecular descriptors) (Arthur et al. 2020). The MLR equation takes the general form (Eq. 2) (Adawara et al. 2020):

$$Y = k_1 x_1 + k_2 x_2 + k_3 x_3 + \ldots C \tag{2}$$

where 'k's and 'x's are, respectively, regression coefficients and independent variables, Y is the dependent variable, and 'C' represents intercept or regression constant.

After the model building has been completed, the next important step was to subject the built models to an internal and external validation assessment. Some parameters computed from the Material studio are useful for the internal validation such as correlation coefficient (R^2), adjusted correlation coefficient (R^2_{adj}), and cross-validation coefficient (Q^2cv). The Y-randomization test was also necessary to show how well the model predicts the activities of the training set compounds (Adawara et al. 2020; Ugbe et al. 2021). To ascertain the built model's ability to predict the activities of the external test set compounds, a validation assessment was carried out externally. The predictive strength of the QSAR model depends heavily on the correlation coefficient (R^2 test) for the external test set (Isyaku et al. 2020). Furthermore, the Golbraikh and Tropsha acceptable QSAR model criteria were equally considered (Roy et al. 2013; Edache et al. 2020a, b; Ugbe et al. 2022a, b). Table 2 shows the selected equations and parameters used for the validation assessment.

Furthermore, the relative contribution of each descriptor in the built QSAR model is shown by the mean effect (ME) values, defined as (Eq. 9):

$$\text{ME} = \frac{B_j \sum_i^n D_j}{\sum_j^m (B_j \sum_i^n D_j)} \tag{9}$$

where β_j represents the coefficient j descriptor in the model, D_j equals each descriptor's value for each molecule in the training set, m represents the number of descriptors present in the model, and n is the total number of compounds in the training set (Adeniji et al. 2019).

The variance inflation factor (VIF) defined as (Eq. 10) shows the extent of inter-correlation between the descriptors.

$$\text{VIF} = \frac{1}{(1 - R^2)} \tag{10}$$

R^2 represents the coefficient of correlation of the multiple regressions between the variables in the model. VIF equal to 1 shows that no inter-correlation exists for each variable, and for VIF values within the range of 1–5, the related model is acceptable, and where VIF is above 10, the related model is unstable and therefore cannot be accepted (Abdullahi et al. 2019).

Also, evaluating the applicability domain (AD) of a QSAR model helps to ascertain the uncertainty in the

Ugbe *et al. Bulletin of the National Research Centre*   (2022) 46:189

Page 8 of 24

Table 2 Some selected equations and parameters for validation of the QSAR model

Parameter	Equation	Eq	Significance	Threshold value
Internal validation				
Friedman lack of fit (LOF)	$LOF = \dfrac{SEE}{(1-\frac{c+dxp}{M})^2}$ $SEE = \sqrt{\frac{(Y_{exp}-Y_{pred})^2}{N-P-1}}$	(3)	Allows for the best fitness score to be obtained	–
Correlation coefficient (R^2)	$R^2 = 1 - \left[\dfrac{\sum (\mathbf{Y}_{exp}-Y_{pred})^2}{\sum (Y_{exp}-\bar{Y}_{training})^2}\right]$	(4)	Measures the degree of fitness of the regression equation	≥ 0.6
Adjusted R^2	$R^2_{adj} = \dfrac{R^2-p(n-1)}{n-p+1}$	(5)	Ensures the model's stability and reliability	≥ 0.5
Cross-validation regression coefficient (Q^2cv)	$Q^2_{cv} = 1 - \left[\dfrac{\sum (Y_{pred}-Y_{exp})^2}{\sum (Y_{exp}-\bar{Y}_{training})^2}\right]$	(6)	Indicates a high internal predictive power	≥ 0.5
The coefficient of determination (cR^2_p) of Y-Randomization	$cR^2_p = RX[R^2 - (R_r)^2]^2$	(7)	This is for confirmation that the QSAR model built is strong and not created by chance	$cR^2_p > 0.50$
External validation				
Predicted R^2 (R^2 test)	$R^2_{test} = 1 - \dfrac{\sum (Ypred_{test}-Yexp_{test})^2}{\sum (Ypred_{test}-\bar{Y}_{training})^2}$	(8)	Measures the ability of the model to predict activity values of an external set of compounds	≥ 0.6
Golbraikh and Tropsha's acceptable model criteria	$\lvert r^2_o - r'^2_o \rvert$	–	Assess the robustness and stability of the model	< 0.3
	$[(r^2 - r^2_o)/r^2]$			< 0.1
	k			$0.85 \leq k \leq 1.15$

SEE = Standard error of estimation, c = number of terms in the model, d = user-defined smoothing parameter, p = total number of descriptors in the model, M = number of data in the training set, $\bar{Y}training$ = mean experimental activity of the training set, $Yexp$ = experimental activity in the training set, $Ypred$ = predicted activity in the training set, n = number of compounds in the training set., cR^2_p = Y-randomization coefficient, R = correlation coefficient for Y-randomization, Rr = average 'R' of random models. $Ypred_{test}$ = predicted activity of test set, $Yexp_{test}$ = experimental activity of test set, r^2 = square correlation coefficients of the plot of predicted activity versus experimental activity values. $r_o{}^2$ = square correlation coefficients of the plot of predicted activity versus experimental activity values at zero intercept. $r'_o{}^2$ = square correlation coefficients of the plot of experimental activity versus predicted activity at zero intercept. k = slope of the plot of predicted activity against experimental activity at zero intercept

prediction of compounds based on their similarity with the model building set. It is therefore a test of the reliability and robustness of the built QSAR model (Tropsha et al. 2003). The AD of the built model was described using the leverage approach. The leverage (h) of a given molecule is defined thus (Eq. 11):

$$h = X(X^T X)^{-1} X^T \qquad (11)$$

where $X = m \times k$ descriptor matrix of the training set compound, and X^T = transpose matrix of X. In order to check for influential molecule or outlier, it is necessary to define the warning leverage (h^*) because it provides the cutoff value (Eq. 12):

$$h^* = 3\frac{(j+1)}{m} \qquad (12)$$

where m = number of training set compounds, and j = number of descriptors in the model.

The plot of standardized residuals versus leverages (William's plot) was used to evaluate the area of significance within the model's chemical space. Therefore, molecules that are found within this area on the plot are said

to be the approved predicted compounds (Veerasamy et al. 2011; Adeniji et al. 2020).

3-D QSAR model building
Molecular alignment
The alignment of molecular structures plays a critical role in 3D-QSAR modeling (Al-Attraqchi et al. 2022) as it strongly determines the predictive accuracy and statistical quality of any given 3D-QSAR model (ElMchichi et al. 2020). Several alignment methods have been reported previously such as atom-based alignment, docking-based alignment, pharmacophore-based alignment, and co-crystallized conformer-based alignment among others (Zhang et al. 2020; Al-Attraqchi et al. 2022). In this study, the atom-based alignment was adopted using the Open3DAlign (O3A) software. The atom-based method attempts to match the atoms of the various structures to be aligned with those of the template structure, based on the atom's properties such as the partial charge. Compound **32** with the highest O3A_score was chosen as the template molecule, onto which the remaining thirty-five compounds were superimposed (aligned) and used to build the 3-D QSAR model.

Ugbe *et al. Bulletin of the National Research Centre*   (2022) 46:189

Page 9 of 24

Model development (CoMFA)

The aligned structures were used for building the 3-D QSAR model using the Open3DQSAR software (Edache et al. 2020a, b). 3-D QSAR can be comparative molecular field analysis (CoMFA) concerned with only steric and electrostatic fields' contributions, or comparative molecular similarity indices analysis (CoMSIA) which deals with the steric, electrostatic, hydrophobic, hydrogen bond donor, and hydrogen bond acceptor fields' contributions, among others (ElMchichi et al. 2020). Herein, the CoMFA model is studied. A dataset of 36 compounds was divided into a training set and a test set of 25 and 11 molecules, respectively, i.e., percentage ratio of 70:30. The steric and electrostatic molecular interaction fields (MIFs) analysis was carried out on the aligned compounds placed within a 3-D cubic lattice of grid size 1.5 Å and a 5.0 Å out gap (Tosco and Balle 2011). Variables pretreatment was carried out as follows: energy cutoff (30.0 kJ/mol), elimination of variables having constant or near-constant values, and standard deviation cutoff (level = 2.0) (Al-Attraqchi et al. 2022). The fractional factorial design (FFDSEL) and uninformative variable elimination-partial least square (UVE-PLS) were used to develop the statistical models and to generate the steric and electrostatic contour plots (Edache et al. 2022). The resulting models were then cross-validated using the leave one out (LOO), leave two out (LTO), and leave many out (LMO).

Molecular docking screening

A molecular docking simulation was conducted between the thirty-six arylimidamide-azole analogs, the reference drug (pentamidine), and the pyridoxal kinase (PdxK) receptor (PDB: 6K91) using Molegro Virtual Docker (MVD). PdxK in 3-D form was retrieved from the protein data bank and prepared on the Molegro Virtual Docker by removing water molecules, ligands, and cofactors associated with the protein structure. The software allows for the repair (rebuild) of all affected residues. The receptor's binding cavities were defined and that which has the largest volume and surface (volume: 398.336 and surface: 952.32) was adopted for the docking. All ligands were imported in PDB file format and prepared appropriately. The simulation was carried out using the parameter settings available in Table 3. The predicted ligand–protein interaction profiles were then visualized using the Biovia Discovery Studio Visualizer. A similar method was earlier reported elsewhere (Ibrahim et al. 2021; Abdullahi et al. 2022).

Table 3 Parameter settings used for the execution of the docking process

Setting	Chosen option
Scoring function	
Score	MolDock score
Grid resolution	0.30Å
Binding site	
Origin	Volume: 398.336; Surface: 971.52
Center	X: 20.73, Y: − 2.48, Z: 35.87
Radius	15
Search algorithm	
Algorithm	MolDock SE
Number of runs	10
Constrain poses to cavity	Yes
After docking: Energy minimization	Yes
After docking: Optimize H-bonds	Yes
Parameter setting	
Maximum iteration	1500
Maximum population size	50
Pose generation	
Energy threshold	100
Simplex evolution	
Maximum steps	300
Neighbor distance factor	1.00
Multiple poses	
The maximum number of poses returned	5
Enable energy threshold	0.00
Cluster similar poses	RMSD threshold: 1.00

Evaluation of pharmacokinetic properties

The prediction of pharmacokinetic properties plays a very critical role in the early stage of drug discovery. This is because only molecules which demonstrate good ADMET and drug-likeness properties reach the preclinical research phase (Ugbe et al. 2022a, b). Therefore, six arylimidamide-azole hybrids (**21**, **22**, **26**, **31**, **33,** and **36**) with the highest anti-leishmanial activities were subjected to drug-likeness and ADMET tests using two online web servers; http://www.swissadme.ch/index.php and http://biosig.unimelb.edu.au/pkcsm, respectively. Similar tests were performed on the newly designed compounds (**36a**–**36 g**) to ascertain their drug-likeness properties. Lipinski's RO5 also called the Pfizer rule is a well-established provision determining the oral bioavailability of a given compound (Lipinski et al. 2001; Lawal et al. 2021). Hence, the various compounds would be subjected to the RO5 criterion to ascertain their oral bioavailability.

Ugbe *et al. Bulletin of the National Research Centre*   (2022) 46:189

Page 10 of 24

Table 4 Selected descriptors used in the 2-D QSAR model

S/No	Descriptor	Description	Class	ME
1	AATS8p	Average Moreau–Broto autocorrelation of lag 8 weighted by polarizability	2-D	0.4986
2	ATSC3c	Centered Moreau–Broto autocorrelation of lag 3 weighted by Gasteiger charge	2-D	− 0.0316
3	VR1_Dzv	Randic-like eigenvector-based index from Barysz matrix weighted by vdw volume	2-D	− 0.0576
4	SpMin8_Bhv	Smallest eigenvalue n. 8 of Burden matrix weighted by van der Waals volume	2-D	0.6330
5	MDEC-22	Molecular distance edge between secondary C and Secondary C	2-D	− 0.0423

Key: ME—Mean effect

Fig. 1 Plot of predicted pIC_{50} against experimental pIC_{50}

Fig. 2 Plot of standardized residual against experimental pIC_{50}

Fig. 3 The plot of standardized residuals against leverages (William's plot)
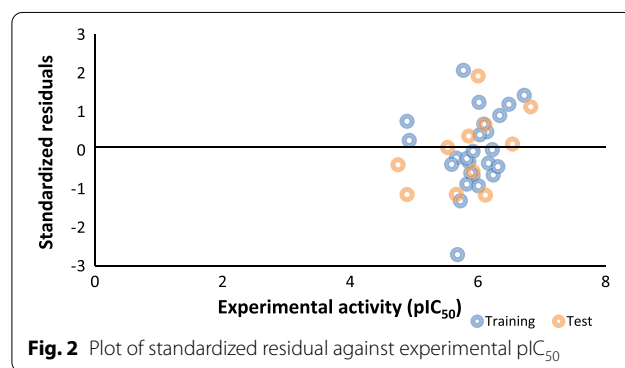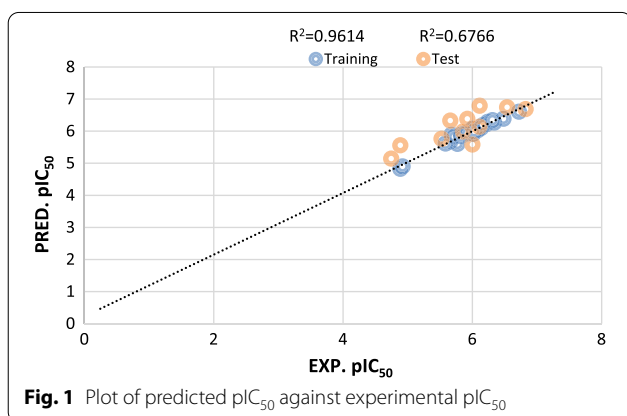
Ligand-based drug design (LBDD)

QSAR has over the years been utilized for ascertaining the relationships between the bioactivities of chemical entities and their physicochemical properties, to obtain a reliable model which could be used for predicting the antiproliferative activities of new chemical compounds. This is governed by the variation of structural properties across compounds which in turn is responsible for the difference in their biological activities. 3-D QSAR has emerged as a more sophisticated tool for designing new molecules than the classical QSAR study (2-D) because it exploits the 3-D properties of the ligands to predict their activities (Verma et al. 2010). In this study, seven compounds (36a–36 g) were designed via LBDD using the template compound (**36**) majorly by structural adjustments involving the addition and/or removal of certain chemical moieties or substituent groups based on the information provided by the steric and electrostatic fields' contour maps. The molecular structures of the new molecules were prepared according to the procedure earlier described under the 'molecular geometry minimization' section. Molecular docking was conducted between the newly designed compounds and the target receptor (PdxK), using MVD, while utilizing the Biovia Discovery Studio to visualize the resulting pharmacological interactions. The bioactivities of the new compounds were predicted by the built 2-D QSAR model equation.
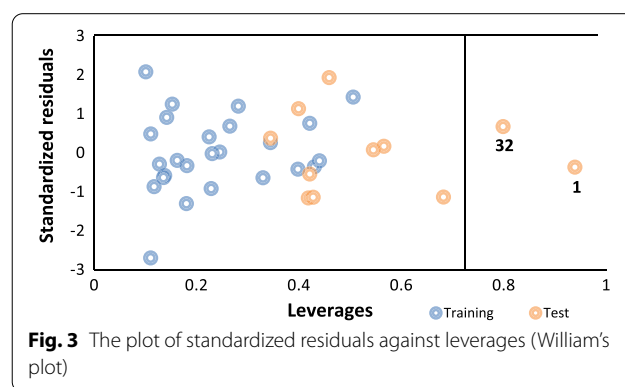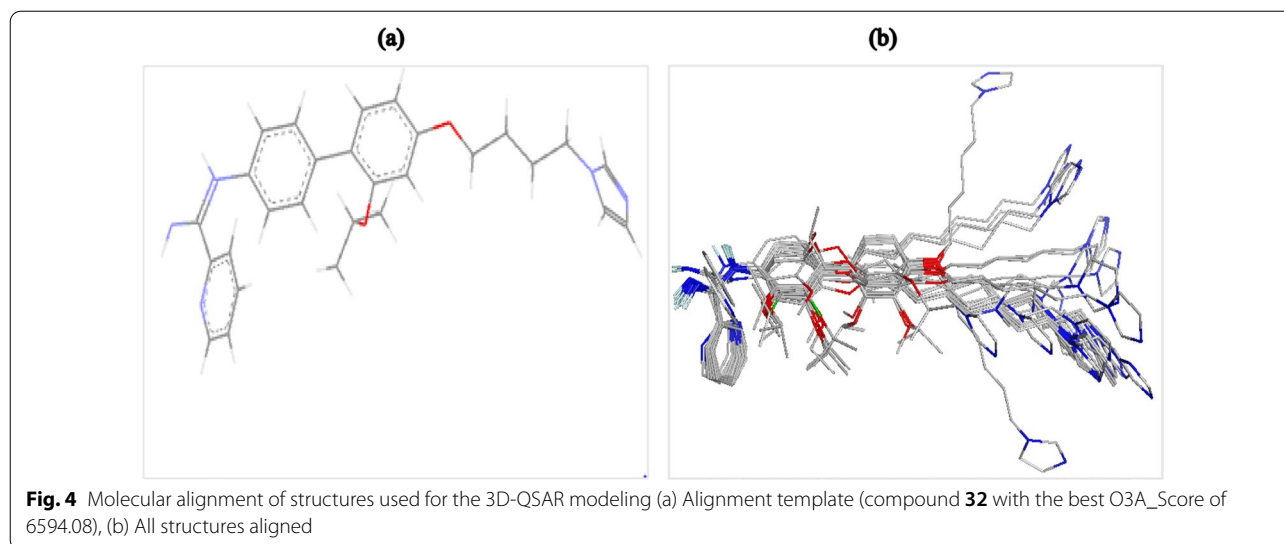
Results

2-D QSAR modeling

Theoretical modeling of thirty-six arylimidamide-azole derivatives was conducted to establish a quantitative relationship between their structures and their inhibitory activities. As a result, a five-descriptor QSAR model was built (Eq. 13) with the descriptors well described in Table 4.

Fig. 4 Molecular alignment of structures used for the 3D-QSAR modeling (a) Alignment template (compound **32** with the best O3A_Score of 6594.08), (b) All structures aligned

The outcome of the internal and external validation assessment conducted on the built model is available in Table 5. The computed descriptors, observed activities (pIC$_{50}$), and the predicted activities together with their residuals are presented in Table 6. Also, a plot of predicted activities versus experimental activities for the training set and test set is shown in Fig. 1, while Fig. 2 shows the plot of standardized residuals against experimental activities (pIC$_{50}$).

$$pIC_{50} = 3.616080840*AATS8p - 2.281588771*ATSC3c$$
$$- 0.000838864*VR1_{Dzv} + 4.105572067*SpMin8_{Bhv}$$
$$- 0.019073658*MDEC - 22 - 2.787505814$$
$$(13)$$

Furthermore, the result of Pearson's correlation statistical analyses performed on all five descriptors' values in the model building set is reported in Table 7. Another powerful validation method called Y-randomization or Y-scrambling

test was equally applied, and the result is presented in Table 8. Finally, Fig. 3 shows the plot of standardized residuals against the leverages otherwise known as William's plot determined to show the model's applicability domain space.

3-D *QSAR modeling (CoMFA)*

Molecular structural alignment represents a crucial factor in determining the predictive capacity of a built 3-D QSAR model. Figure 4a, b shows the molecular structure of the alignment template (compound **32**) and the aligned structures as obtained from the atom-based superimposition of the remaining structures on the template. Two models were developed: each from the FFD-SEL and UVEPLS approaches. Some significant statistical parameters calculated for both the 3-D models are presented in Table 9. Presented in Table 10 are the O3A scores, experimental pIC$_{50}$, and the predicted pIC$_{50}$ available from the two models, as well as their residual values.

Fig. 5 Correlation between predicted pIC$_{50}$ and experimental pIC$_{50}$ for training and test sets **a** FFDSEL, **b** UVEPLS

Ugbe *et al. Bulletin of the National Research Centre*   (2022) 46:189

Page 12 of 24
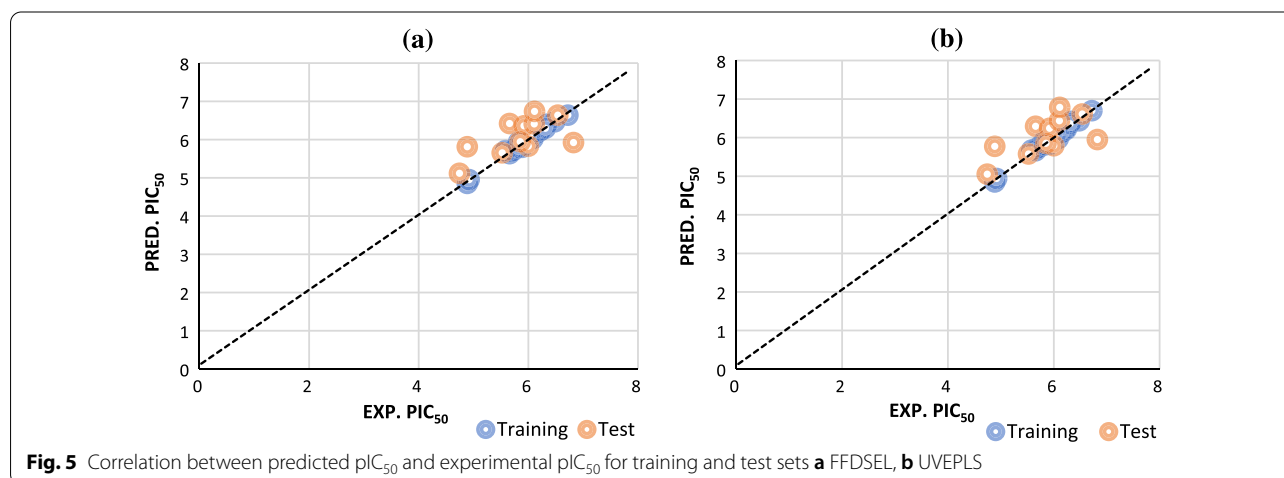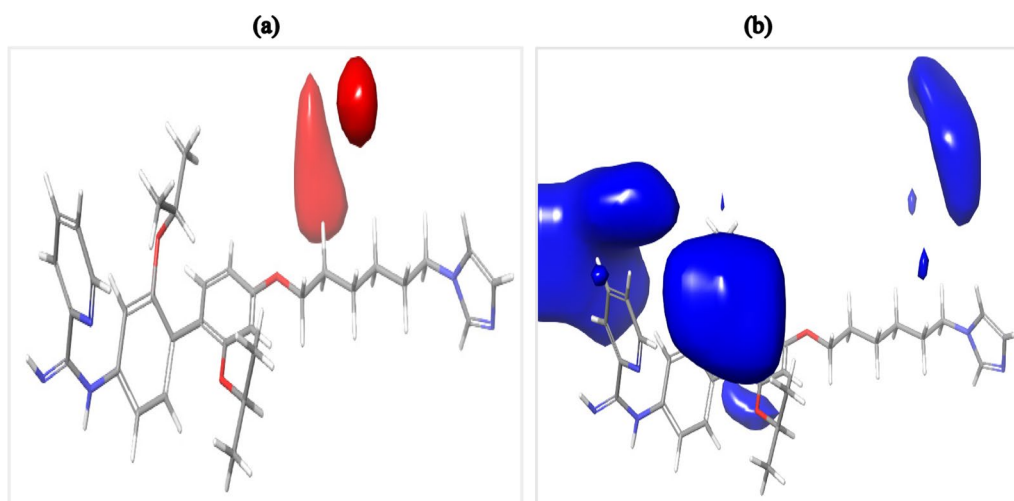
Fig. 6 Visualization of the CoMFA steric field contour maps of compound **36 a** Red contours represent regions of unfavorable steric bulk; **b** Blue contours represent regions of favorable steric bulk
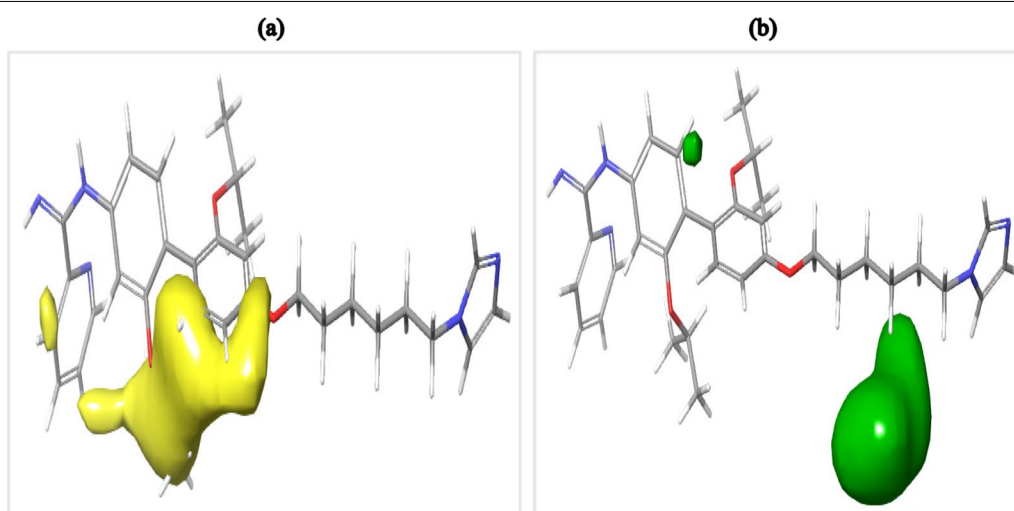
Fig. 7 Visualization of the CoMFA electrostatic field contour maps of compound **36 a** Yellow represents regions favored by high electron density or unfavorable to electron-withdrawing substituents; **b** Green represents regions of unfavorable high electron density or favorable to electron-withdrawing groups

Additionally, a plot showing the correlation between predicted and experimental activities for training and test sets was obtained for both models and is presented in Fig. 5a, b. The CoMFA model equation (FFDSEL) is summarized graphically as 3-D contour maps as shown in Figs. 6a, b and 7a, b.

Molecular docking study

The results (MolDock scores) of the molecular docking simulation conducted between the receptor (PdxK) and

the 36 arylimidamide-azole hybrids as well as the reference drug (pentamidine) using the Molegro Virtual Docker are earlier reported in Table 1. The predicted pharmacological interactions of compound **36** (template) and pentamidine, with the target receptor, are shown in Fig. 8.

Pharmacokinetics study

The results of the drug-likeness and ADMET properties investigation conducted on the selected

arylimidamide-azole analogues and the newly designed compounds are presented in Tables 11 and 12.

Ligand-based drug design (LBDD)

A total of seven compounds were designed using the ligand-based drug design approach. A combined 2-D and 3-D QSAR approach was adopted, where information provided by the 3-D QSAR (CoMFA) contour maps was used for the design, while the inhibitory activities of these new compounds were predicted using the 2D-QSAR model equation. Their molecular structures predicted activities and docking scores are presented in Table 13, while Table 14 shows their predicted pharmacological interaction profiles with PdxK. Also, the binding interactions of these compounds (excluding **36f** and **36 g**) with the active sites of PdxK as visualized via Biovia discovery studio are presented in Figs. 9, 10, 11. Their predicted drug-likeness and ADMET properties are earlier reported in Tables 11 and 12.

Discussion

2-D *QSAR modeling*

The values of VIF reported in Table 7 range from 1 to 5 for all 5 descriptors, indicating that the built QSAR model was statistically substantial and is therefore stable and acceptable. Each descriptor has an absolute t-statistics value of greater than 2, which shows that the selected descriptors were significant (Adeniji et al. 2018). Additionally, the calculated p values (Table 7) of these descriptors in the model at a 95% confidence level were less than 0.05. This is therefore in support of an alternative hypothesis that states that there is a relationship between the descriptors in the model equation and the compounds' inhibitory activities at $p < 0.05$.

The strength and direction of each descriptor's contributions in the built QSAR model are determined by the value of the mean effect (ME) (Table 4). The magnitude of ME signifies the extent to which the descriptor influences the molecule's inhibitory activity, while the sign of ME shows the direction of influence. As shown in Table 4, SpMin8_Bhv and AATS8p have the largest positive ME, which implies elevating their values will mean a corresponding increase in the molecule's inhibitory activities. ATSC3c, VR1_Dzv, and MDEC-22 on the other hand have negative ME values, signifying that an increase in these descriptors' values will lead to a decrease in the molecule's antiproliferative activities.

The application of GFA in conjunction with the MLR approach led to the development of the 2-D QSAR model with a total of five descriptors as shown in Eq. 13. The built model was adjudged to satisfy the requirement of a good QSAR model. The low residual values between the

Table 5 Validated parameters of the 2D-QSAR model

Validation parameters	Model	Threshold	Remarks
Training set			
Friedman LOF	0.037669	–	–
R-squared (R^2)	0.961448	≥ 0.6	Passed
Adjusted R-squared (R^2_{adj})	0.951303	≥ 0.5	Passed
Cross-validated R-squared (Q^2_{cv})	0.935022	≥ 0.5	Passed
$R^2 - Q^2_{cv}$	0.026426	≤ 0.3	Passed
Significant regression	Yes	–	–
Significance of regression F value	94.768088	–	–
Critical SOR F value (95%)	2.7617200	–	–
Replicate points	0	–	–
Computed experimental error	0.000000	–	–
Test set			
R-squared (R^2_{test}), i.e.,r^2	0.67655	≥ 0.6	Passed
Number of test set compounds ($N_{test\ set}$)	11	≥ 5	Passed
$\left\| r_o^2 - r_o'^2 \right\|$	0.13553	< 0.3	Passed
$[(r^2 - r_o^2)/r^2]$	0.00036	< 0.1	Passed
k	0.95605	$0.85 \leq k \leq 1.15$	Passed

experimental and predicted activity values for both training and test set compounds as shown in Table 6 indicate a high predictive strength of the built model. The R^2 values of 0.9614 for the training set and 0.6766 for the test set from Fig. 1 were similar to those obtained from GFA (0.9615 and 0.6766) and MLRplusValidation analysis (0.9615 and 0.6766) available in Table 5, which means that the R^2 values were computed accurately. Figure 1 shows how well the predicted activities correlated with the experimental activities as suggested by the grouping of points along the line of best fits, implying that the built model is very reliable and robust in predicting the activities of new chemical entities. A random distribution of standard residuals on both sides of zero as observed in Fig. 2 shows that the built model is not associated with any systematic error.

Furthermore, the result of the Pearson correlation analysis available in Table 7 shows that no pair of descriptors is significantly inter-correlated as suggested by the low values of the correlation coefficients. This was also observed by Adeniji et al. (2018).

The values of R^2 and Q^2 obtained from the y-randomization test (Table 8) were significantly low, and this implies that the developed model is stable, robust, and reliable. The Y-randomization coefficient, c R^2p (0.877748) greater than 0.50, signifies that the built model is powerful and not inferred by chance. The plot of standardized residuals against leverages (William's plot)

Table 6 Calculated descriptors, experimental pIC$_{50}$, predicted pIC$_{50}$, and residuals of the arylimidamide-azole hybrids

Comp ID	AATS8p	ATSC3c	VR1_Dzv	SpMin8_Bhv	MDEC-22	pIC$_{50}$	Pred. pIC$_{50}$	Residuals
1*	1.2199	0.10142	270.0957	1.058166	19.2253	4.7447	5.1436	− 0.3988
2	1.1374	0.10337	246.1391	1.059033	21.2696	4.8861	4.8252	0.0609
3	1.1717	0.10146	336.3404	1.074016	23.3487	4.9208	4.8999	0.0210
4*	1.1394	0.10088	274.3332	1.275013	27.5844	6.0000	5.5810	0.4190
5*	1.1366	0.09818	276.4469	1.259531	25.0924	4.8861	5.5593	− 0.6732
6	1.1171	0.09817	337.6577	1.386533	29.3725	5.8539	5.8772	− 0.0234
7	1.1436	− 0.0869	276.0232	1.274367	30.1033	5.9208	5.9725	− 0.0516
8	1.1421	0.08209	274.0999	1.279542	30.1964	5.7696	5.6024	0.1671
9	1.2460	0.13739	285.0252	1.274853	24.4658	5.8861	5.9330	− 0.0469
10	1.2228	0.11552	283.3093	1.274145	24.6116	5.6778	5.8948	− 0.2170
11	1.1349	0.21981	294.8321	1.357263	24.4658	5.6576	5.6734	− 0.0158
12	1.1172	0.19295	304.3565	1.407047	26.6076	5.7212	5.8262	− 0.1049
13	1.1034	0.15186	313.6883	1.409511	24.4658	6.0132	5.9129	0.1003
14	1.1247	0.16042	291.6092	1.386967	24.6116	5.8239	5.8938	− 0.0699
15	1.1465	0.13456	300.2214	1.414288	26.6427	6.1367	6.0978	0.0389
16	1.1650	0.09448	309.3355	1.414420	24.6116	6.2366	6.2877	− 0.0511
17	1.3544	0.11905	545.5559	1.157252	27.1705	5.5850	5.6139	− 0.0289
18	1.2919	0.11910	1400.000	1.373810	0.00000	6.0044	6.0781	− 0.0737
19	1.2847	0.25278	562.4740	1.383724	24.0822	6.0862	6.0312	0.0549
20	1.2443	0.25282	785.8467	1.399141	0.00000	6.2218	6.2202	0.0017
21	1.3170	0.18643	592.6294	1.410787	24.0822	6.4815	6.3852	0.0963
22*	1.2801	0.18647	625.0888	1.425862	0.00000	6.5376	6.7457	− 0.2081
23*	1.2981	0.09132	564.8354	1.354464	23.9122	5.6576	6.3293	− 0.6717
24	1.2529	0.09137	1670.000	1.389976	0.00000	5.8239	5.8405	− 0.0166
25*	1.2183	0.02712	583.0883	1.405821	23.9122	5.9208	6.3824	− 0.4616
26*	1.1929	0.02717	733.2270	1.422312	0.00000	6.8239	6.6885	0.1354
27*	1.2645	0.27261	662.1033	1.366036	23.7602	5.5229	5.7628	− 0.2399
28*	1.2511	0.20456	728.1258	1.408369	23.7602	5.8539	5.9881	− 0.1343
29	1.2271	0.27257	972.9505	1.390970	0.00000	5.9208	5.9225	− 0.0016
30	1.2229	0.20452	986.6988	1.424248	0.00000	6.1612	6.1878	− 0.0266
31	1.2649	0.05074	751.7122	1.382670	23.7266	6.3372	6.2643	0.0730
32*	1.2287	− 0.0132	1060.000	1.410908	23.7266	6.1079	6.1363	− 0.0284
33	1.2255	0.05078	810.6949	1.402543	00.0000	6.7212	6.6062	0.1150
34*	1.2017	− 0.0131	773.6185	1.425836	0.00000	6.1135	6.7928	− 0.6793
35	1.1972	0.01983	967.5941	1.408440	24.8039	6.0269	5.994	0.0328
36	1.1836	− 0.1077	1060.000	1.453531	24.8039	6.3098	6.3436	− 0.0338

Table 7 Inter-correlation analysis and statistical parameters of descriptors forming the 2-D QSAR model

Descriptors	Inter-correlation					Statistical parameters		
	AATS8p	*ATSC3c*	*VR1_Dzv*	*SpMin8_Bhv*	*MDEC-22*	*VIF*	*t-Stat*	*p value*
AATS8p	1					1.7182	10.359	3E−09
ATSC3c	0.1707	1				1.7813	− 8.6015	5.6E−08
VR1_Dzv	0.5182	− 0.0772	1			4.8285	− 8.16	1.2E−07
SpMin8_Bhv	− 0.011	0.1074	0.3819	1		1.3996	20.003	3.2E−14
MDEC-22	− 0.38	− 0.3148	− 0.7419	− 0.2559	1	3.5129	− 6.2199	5.6E−06

VIF-Variance inflation factor

Table 8 Y-randomization test parameters

Iteration	R	R^2	Q^2
Random 1	0.498658	0.248660	− 0.17040
Random 2	0.115103	0.013249	− 0.55512
Random 3	0.318787	0.101625	− 0.88510
Random 4	0.572381	0.327620	− 0.17467
Random 5	0.430382	0.185229	− 0.26640
Random 6	0.364943	0.133183	− 0.59773
Random 7	0.235410	0.055418	− 0.45178
Random 8	0.486282	0.236470	− 0.27500
Random 9	0.466289	0.217426	− 0.18640
Random 10	0.513190	0.263364	− 0.27803
Random models parameters			
Average R	0.400143		
Average R^2	0.178224		
Average Q^2	− 0.38406		
cRp^2	0.877748		

cR_p^2-Y—randomization coefficient, R—correlation coefficient for Y—randomization, R^2—correlation coefficient, Q^2—cross-validated $R.^2$

Table 9 Statistical parameters of the built CoMFA models

Parameters	CoMFA (FFDSEL)	CoMFA (UVEPLS)
PC	5	5
R^2	0.9817	0.9839
SDEC	d0.0544	0.0509
F-test	203.39	232.69
Q^2_{LOO}	0.6806	0.7539
$SDEP_{LOO}$	0.2268	0.1991
Q^2_{LTO}	0.6660	0.7367
$SDEP_{LTO}$	0.2319	0.2059
Q^2_{LMO}	0.5956	0.6784
$SDEP_{LMO}$	0.2537	0.2262
Field contributions		
Steric	0.6819 (68.19%)	0.6896 (68.96%)
Electrostatic	0.3181 (31.81%)	0.3104 (31.04%)

PC—principal components, SDEP—standard error of prediction, *F*-test—Fischer's statistics, LOO—leave one out, LTO—leave two out, LMO—leave many out, Q^2—cross-validated correlation coefficient, $R.^2$—correlation coefficient, SDEC—standard error of correlation

(Fig. 3) shows that all the compounds were contained within the ± 3.0 square area of standardized cross-validated residual, which clearly shows that no outlier is present in the data set. However, compounds 1 and **32** have leverages greater than the computed warning leverage ($h^* = 0.72$) and were said to be influential molecules.

3-D QSAR modeling

The 3-D QSAR modeling began with the atom-based alignment of structures using compound **32** as the alignment template. The alignment process involves an early step that provided all the 36 compounds the opportunity of being selected as the alignment template based on the compound with the highest Open3DAlign (O3A) score. The O3A scores of the various compounds are included in Table 10. Compound **32** was found with the highest O3A score of 6594.08 and hence selected as the template upon which the remaining structures were superimposed. Two CoMFA models were built using the FFDSEL and UVEPLS approaches, with their computed statistical parameters tabulated in Table 9. The parameters were computed for five principal components (PCs) among which the fifth PC (PC = 5) performed relatively better. The statistical parameters available in Table 9 were those associated with PC 5. The predictive strength of the regression models on new datasets of compounds can be estimated by cross-validation (Grohmann and Schindler 2008). A cross-validated coefficient of correlation (Q^2) ≥ 0.50 indicates a good QSAR model. Here, three types of Q^2 were calculated; leave one out (LOO), leave two out (LTO), and leave many out (LMO), together with their associated standard error of prediction (SDEP). Although both models passed the cross-validation test (i.e., $Q^2 > 0.5$), relatively higher values of correlation coefficient (R^2), Q^2_{LOO}, Q^2_{LTO}, $Q^2_{LMO,}$ and the Fischer's statistics (F-test) were associated with the UVEPLS CoMFA model. Also, the calculated standard error of correlation (SDEC) and SDEP were lower for UVEPLS than for FFDSEL. A linear correlation between the CoMFA descriptors (independent variables) and the activity values (dependent variables) was established by the PLS analysis method. The lower residual values between the predicted and observed activity values (Table 10) show a strong predictive strength of both models. This was supported by the clustering of points along the lines of best fits in the plots of predicted pIC_{50} against the experimental pIC_{50} for both models (Fig. 5a, b).

The CoMFA QSAR equation is summarized graphically as a 3-D contour map, which shows the regions within the molecules' 3-D structural space where steric and electrostatic fields are associated with extreme values. The underlying principle behind CoMFA is that variations in the shape and strength of non-covalent interaction fields surrounding the molecules, such as steric or electrostatic fields, can be related to changes in binding affinities (Kakarla et al. 2016). Therefore, molecular fields are key factors in binding affinity.

The contour maps presented in Figs. 6, 7 were those generated from the UVEPLS CoMFA model for compound **36**. The steric and electrostatic field contributions are 68.19% and 31.81% for FFDSEL, and 68.96% and 31.04% for UVEPLS, respectively (Table 9), which has only a very insignificant difference between both models.

From the steric field contour maps available in Fig. 6a, b, the red contours represent the regions of unfavorable steric bulk, while blue contours represent the regions of favorable steric bulk. An identified region of unfavorable steric bulk in compound **36** is the number 2-position of the alkyl chain linking the biphenyl group and the imidazole group. This means that if the hydrogen in that position is replaced with a more bulky group like methyl, ethyl, isopropyl, etc., the compound's activity or binding affinity will decrease. On the other hand, more steric bulk favorable regions were identified, which include the alkoxy substituents on the biphenyl ring system and the 4-position of the pyridine ring system. This implies that increasing the steric bulk around those areas will enhance the inhibitory activity or binding affinity of the molecule. From the electrostatic field contour maps available in Fig. 7a, b, yellow contours represent regions favored by high electron density or unfavorable to electron-withdrawing substituents, while green contours represent regions of unfavorable high electron density or favorable to electron-withdrawing groups. Two regions of low electron density were identified to include the number 4-position of the alkyl chain linking the biphenyl group to the imidazole group, and the number 2-position of the phenyl ring which links to the amide group. These regions need not be too electron-dense, and hence, the introduction of electron-withdrawing groups at these positions will enhance the molecule's inhibitory activity or binding affinity. Regions of high electron density were identified around positions 2 and 3 of the phenyl ring which links to the alkyl chain via the oxygen, and positions 3 and 5

of the pyridine ring system. These regions need not be less electron-dense, and hence, electron-donating groups will keep these regions at a high electron density which in turn will enhance the molecule's inhibitory activity or binding affinity.

Molecular docking study

The binding affinity of protein–ligand interaction is necessary to describe how well drugs bind to the target receptor. The spontaneity of the binding process and how well drugs can fit into the receptor's binding cavities to form the most stable receptor is shown by the negative value of the binding affinity (Ugbe et al. 2022a, b). Here, the results (binding affinities) obtained from the docking study using Molegro Virtual Docker are expressed as MolDock scores in kcal/mol as included in Table 1. On average, the result shows that binding affinity correlates well with inhibitory activity. The screening identified compounds **1** and **36** as showing the least and highest binding affinities of -142.632 kcal/mol and -198.260 kcal/mol, respectively. Also, the predicted binding affinity (MolDock score $= -141.793$ kcal/mol) of the reference drug (pentamidine) was lower than those of the arylimidamide-azole hybrids, an indication that the various compounds bind better with the active sites of the protein (PdxK) than pentamidine. This claim can further be supported by comparing the binding interaction profile of 36 with that of pentamidine as presented in Fig. 8 and Table 14. More interactions were visible from the binding profile of compound **36** with
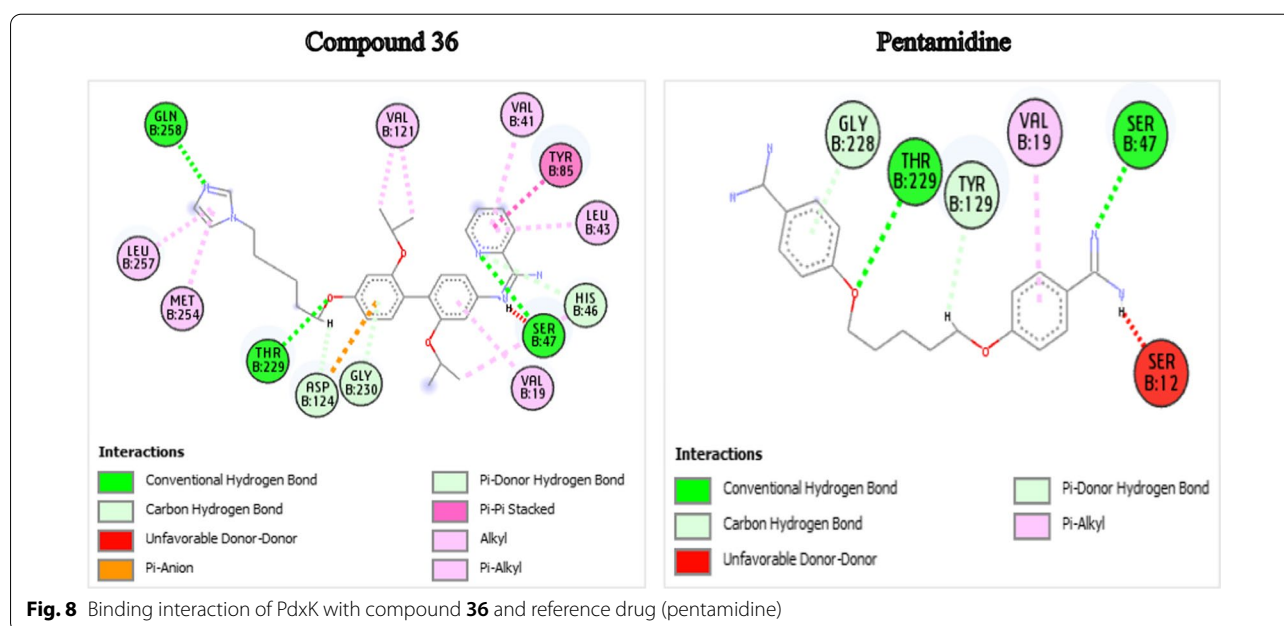

Fig. 8 Binding interaction of PdxK with compound **36** and reference drug (pentamidine)

Table 10 Open3DAlign scores, observed pIC$_{50}$, predicted pIC$_{50}$, and residuals for FFDSEL and UVEPLS CoMFA models

Comp. ID	pIC$_{50}$	O3A_Score	CoMFA (FFDSEL)		CoMFA (UVEPLS)	
			Pred. pIC$_{50}$	Residual	Pred. pIC$_{50}$	Residual
1*	4.745	5513.18	5.118	−0.373	5.052	−0.307
2	4.886	5615.18	4.852	0.034	4.850	0.036
3	4.921	5719.81	4.951	−0.030	4.938	−0.017
4*	6.000	6084.29	5.824	0.176	5.787	0.213
5*	4.886	6101.02	5.809	−0.923	5.773	−0.887
6	5.854	5927.51	5.824	0.030	5.851	0.003
7	5.921	5540.74	5.851	0.070	5.827	0.094
8	5.770	5693.62	5.780	−0.010	5.815	−0.045
9	5.886	5578.15	5.796	0.090	5.836	0.05
10	5.678	5906.85	5.683	−0.005	5.721	−0.043
11	5.658	5858.38	5.631	0.027	5.657	0.001
12	5.721	5580.25	5.709	0.012	5.742	−0.021
13	6.013	5788.80	5.968	0.045	5.971	0.042
14	5.824	5674.69	5.934	−0.110	5.924	−0.100
15	6.137	6165.71	6.149	−0.012	6.123	0.014
16	6.237	6197.31	6.249	−0.012	6.211	0.026
17	5.585	6323.30	5.707	−0.122	5.675	−0.090
18	6.004	6367.39	6.035	−0.031	6.023	−0.019
19	6.086	6399.99	6.019	0.067	5.973	0.113
20	6.222	6436.15	6.236	−0.014	6.217	0.005
21	6.481	6529.21	6.462	0.019	6.434	0.047
22*	6.538	6500.79	6.638	−0.100	6.612	−0.074
23*	5.658	6036.83	6.419	−0.761	6.293	−0.635
24	5.824	6180.15	5.792	0.032	5.809	0.015
25*	5.921	6013.88	6.357	−0.436	6.242	−0.321
26*	6.824	6174.96	5.922	0.902	5.950	0.874
27*	5.523	6049.77	5.651	−0.128	5.576	−0.053
28*	5.854	6068.11	5.944	−0.090	5.854	0.000
29	5.921	6104.10	5.912	0.009	5.918	0.003
30	6.161	5782.03	6.205	−0.044	6.137	0.024
31	6.337	6574.15	6.398	−0.061	6.422	−0.085
32*	6.108	6594.08	6.406	−0.298	6.435	−0.327
33	6.721	6455.52	6.637	0.084	6.699	0.022
34*	6.114	6579.92	6.734	−0.620	6.782	−0.668
35	6.027	5796.79	6.112	−0.085	6.050	−0.023
36	6.310	6137.54	6.292	0.018	6.359	−0.049

O3A—Open3DAlign

PdxK. For **36_PdxK,** three conventional hydrogen bonds were formed involving GLN-258 (interaction distance of 2.78 Å), THR-229 (1.95 Å), and SER-47 (2.42 Å), Two carbon-hydrogen (C-H) bonds were visible with HIS-46 and ASP-124, a π-donor hydrogen bond with GLY-230, π-anion electrostatic interaction with ASP-124, and several hydrophobic interactions. **Pentamidine_PdxK** on the other hand shows two conventional hydrogen bonds with THR-229 (2.53 Å) and SER-47 (2.52 Å), one C–H bond involving TYR-129, a π-donor hydrogen bond with GLY-228, and one π-alkyl hydrophobic interaction with VAL-19.

Pharmacokinetics properties prediction
Lipinski's RO5 for oral bioavailability has a broad application in the discovery of new drug molecules (Ugbe et al. 2022a, b). It asserts that a drug molecule may likely not be orally bioavailable when it has hydrogen bond donors

Table 11 Predicted drug-likeness properties of some selected arylimidamide-azole derivatives and the newly designed analogues

Comp ID	MW	TPSA	MLOGP	HBD	HBA	Rot. b	LOG S (ESOL)	SA	RO5 violation	Drug likeness
Six (6) arylimidamide-azole hybrids with the highest pIC$_{50}$										
21	469.58	85.05	2.78	2	5	12	− 5.54	3.60	0	YES
22	497.63	85.05	3.17	2	5	14	− 6.01	3.83	0	YES
26	497.63	85.05	3.17	2	5	14	− 6.01	3.80	0	YES
31	441.52	85.05	2.38	2	5	11	− 4.96	3.28	0	YES
33	469.58	85.05	2.78	2	5	13	− 5.42	3.51	0	YES
36	555.71	94.28	3.21	2	6	16	− 6.68	4.34	1	YES
Newly designed arylimidamide-azole analogs										
36a	583.76	94.28	3.57	2	6	16	− 7.06	4.59	1	YES
36b	551.68	94.28	3.07	2	6	14	− 6.68	2.67	1	YES
36c	581.75	94.28	3.50	2	6	15	− 7.05	2.78	1	YES
36d	579.73	94.28	3.44	2	6	14	− 7.06	3.02	1	YES
36e	630.22	94.28	4.13	2	6	15	− 7.96	3.01	1	YES
36f	664.66	94.28	4.30	2	6	15	− 8.36	3.25	2	NO
36 g	692.72	94.28	4.65	2	6	15	− 8.97	2.85	2	NO

MW—molecular weight, TPSA—topological polar surface area, ESOL—estimated solubility, HBD—hydrogen bond donors, HBA—hydrogen bond acceptors, RO5—Lipinski rule of five, SA—synthetic accessibility score

Table 12 Predicted ADMET properties of some selected arylimidamide-azole derivatives and the newly designed analogues

Comp ID	A				D		M				E	T	
	HIA	P-glycoprotein			Skin	BBB	CNS	CYP-2D6		CYP3A4		TCE	AMES	MRTD
	%	S	I	II	LogKp	LogBB	LogPS	S	I	S	I			
Six (6) arylimidamide-azole hybrids with the highest pIC$_{50}$														
21	92.56	YES	YES	YES	− 2.74	− 0.62	− 2.36	NO	YES	YES	YES	0.843	YES	0.276
22	90.83	YES	YES	YES	− 2.74	− 0.62	− 2.31	NO	YES	YES	YES	0.906	NO	0.193
26	90.68	YES	YES	YES	− 2.74	− 0.55	− 2.27	NO	YES	YES	YES	0.939	YES	0.282
31	92.52	YES	YES	YES	− 2.74	− 1.13	− 2.45	NO	NO	YES	YES	0.904	YES	0.406
33	90.80	YES	YES	YES	− 2.74	− 1.16	− 2.40	NO	YES	YES	YES	0.997	YES	0.323
36	100.0	YES	YES	YES	− 2.74	− 0.70	− 2.29	NO	YES	YES	YES	0.848	NO	0.132
Newly designed arylimidamide-azole analogs														
36a	100	YES	YES	YES	− 2.74	− 0.66	− 1.92	NO	YES	YES	YES	0.437	NO	0.073
36b	100	YES	YES	YES	− 2.74	− 0.70	− 2.13	NO	YES	YES	YES	0.529	NO	0.088
36c	100	YES	YES	YES	− 2.74	− 0.63	− 1.80	NO	YES	YES	YES	0.354	NO	0.072
36d	100	YES	YES	YES	− 2.74	− 0.60	− 1.74	NO	YES	YES	YES	0.144	NO	0.068
36e	100	YES	YES	YES	− 2.74	− 0.42	− 1.47	NO	NO	YES	YES	0.237	NO	0.098
36f	100	YES	YES	YES	− 2.74	− 0.61	− 1.44	NO	NO	YES	YES	0.290	NO	0.112
36 g	100	YES	YES	YES	− 2.74	− 0.55	− 1.26	NO	NO	YES	YES	0.338	NO	0.123

BBB—Blood–brain barrier, CNS—central nervous system, HIA—human intestinal absorption, Skin—skin permeability, LogBB—the logarithmic ratio of brain to plasma drug concentration, LogPS—blood–brain permeability-surface area product, CYP-34A—cytochrome p450 isoform, CYP-2D6—cytochrome p450 isoform, S—substrate, I—inhibitor, MRTD—maximum recommended tolerated dose, TCE—total clearance

(HBDs) of greater than 5, hydrogen bond acceptors (HBAs) > 10, molecular weight (MW) > 500, and lipophilicity (MLOGP > 4.15 or WLOGP > 5) (Lipinski et al. 2001). Whenever a molecule passed at least three of the four provisions of the RO5, it is said to obey Lipinski's rule

for oral bioavailability (Lawal et al. 2021). From Table 11, it was observed that all the selected arylimidamide-azole hybrids passed the drug-likeness test (Lipinski RO5). Five of the seven newly designed analogs pass the Lipinski criteria. Only **36f** and **36 g** failed the drug-likeness test

Ugbe *et al. Bulletin of the National Research Centre*   (2022) 46:189

Page 19 of 24

Table 13 Molecular structures, predicted pIC$_{50}$ and MolDock scores of the newly designed arylimidamide-azole analogs
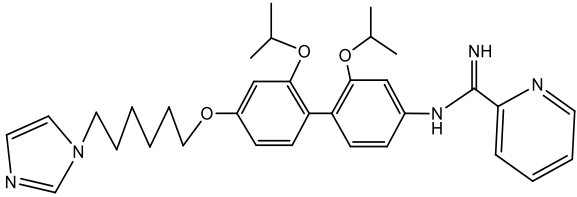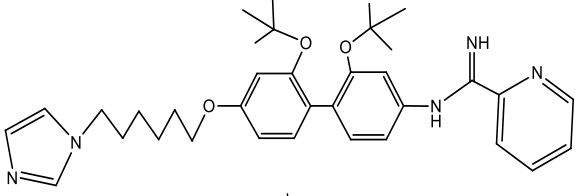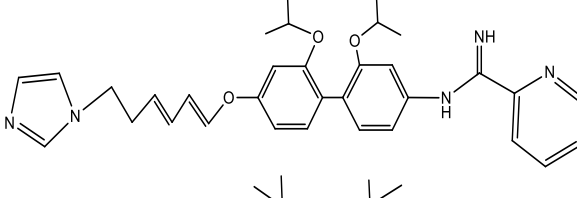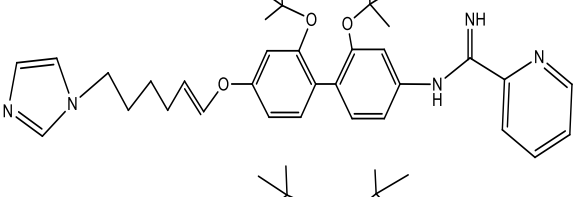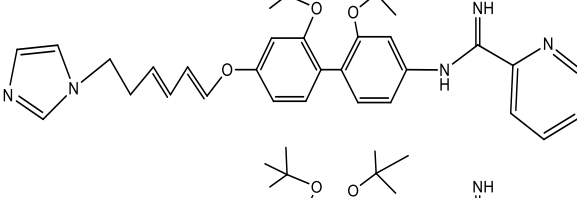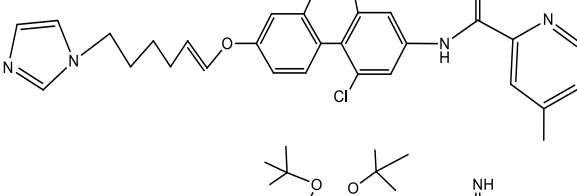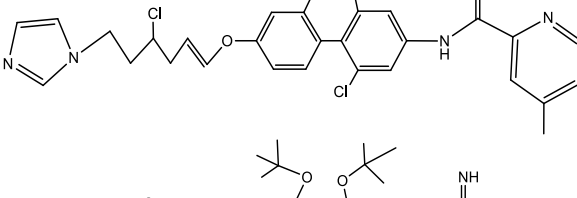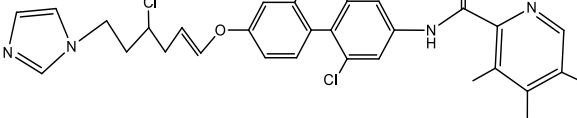
Comp. ID	Molecular structures	Pred. pIC$_{50}$	MolDock score
36		6.344	− 198.26
36a		6.700	− 206.468
36b		6.635	− 196.665
36c		6.859	− 205.004
36d		6.768	− 199.635
36e		7.375	− 208.595
36f		7.670	− 206.812
36g		6.860	− 187.155

Ugbe *et al. Bulletin of the National Research Centre*   (2022) 46:189

Page 20 of 24

Table 14 Predicted binding interaction profiles of the newly designed molecules with PdxK receptor

ID	Hydrogen bond Interactions			Electrostatic/hydrophobic interactions
	Amino acid	Type	Distance (Å)	
36	THR-229	Conventional	1.95	ASP-124 (π-anion), TYR-85 (π–π stacked), VAL-121 (alkyl), π-alkyl (LEU-43, VAL-41, VAL-19, MET-254, LEU-257) and unfavorable donor–donor clash with SER-47
	GLN-258	Conventional	2.78	
	SER-47	Conventional	2.42	
	HIS-46	C–H	2.95	
	ASP-124	C–H	2.69	
	GLY-230	π-donor	3.18	
36a	ARG-225	Conventional	2.31	ASP-124 (attractive charge), ASP-124 (π-anion), VAL-121 (alkyl), π-alkyl (LEU-257, LYS-187, ARG-127, TYR-129, VAL-19)
	LYS-187	C–H	2.62	
	ASP-124	C –H	2.14	
	GLY-228	π –donor	2.32	
	GLY-230	π –donor	2.79	
36b	HIS-222	Conventional	2.03	ASP-231 (attractive charge), ASP-231 (π-anion), HIS-46 (π–π stacked), VAL-121 (alkyl), π-alkyl (TYR-129, VAL-121, VAL-19, LEU-198, ILE-261)
	HIS-222	C–H	2.35, 2.96	
	TYR-129	C–H	3.09	
	LYS-187	C–H	2.09	
36c	SER-47	Conventional	2.39	π-anion (ASP-124, ASP-125), TYR-85 (π–π stacked), VAL-121 (alkyl), π-alkyl (LEU-43, VAL14, VAL-19, TYR-129, ARG-225)
	ASN-45	C–H	3.00	
	THR-227	C–H	2.45	
	TYR-226	C–H	1.75	
	ARG-225	C–H	2.41	
36d	GLY-48	C–H	2.63	ASP-124 (attractive charge), ASP-124 (π-anion), ASP-124 (π-cation), TYR-49 (π–π T-shaped), THR-229 (amide-π stacked), VAL-121 (alky), alkyl (ARG-278, TYR-129, LEU-257, LYS-187) and unfavorable steric bump with THR-229
	GLY-228	π-donor	3.09	
	GLY-230	π-donor	2.96	
36e	ASP-125	C–H	2.69	ASP-231 (attractive charge), HIS-46 (π-anion), ASP-124 (π-anion), HIS-46 (π–π stacked), VAL-121 (alkyl), π-alkyl (HIS-46, TYR-129, halogen (ASP-231)
	GLY-228	π-donor	3.32	
	GLY-230	π-donor	2.83	
Pentamidine	SER-47	Conventional	2.52	VAL-19 (π-alkyl) and unfavorable donor-donor clash with SER-12
	THR-229	Conventional	2.53	
	TYR-129	C-H	2.36	
	GLY-228	π-donor	2.39	

ARG—Arginine, ASN—asparagine, ASP—aspartic acid, GLN—glutamine, GLY—glycine, HIS—histidine, LYS—lysine, SER—serine, THR—threonine, TYR—tyrosine

because both violated two of the RO5 criterion: MW > 500 and MLOGP > 4.15. The reported values of topological polar surface area (TPSA) for all molecules are less than 140 Å². Also, the predicted values of the estimated water solubility (Log S) for the selected compounds showed that only **21**, **31,** and **33** are moderately water-soluble (-4 > Log S > -6), while **22**, **26**, **36** and the newly designed compounds are associated with poor solubility (-6 > Log S > -10). Solubility can, however, be improved by careful introduction of moieties that suggest so into the structural template. The values of the synthetic accessibility (SA) scores of the designed compounds except **36a** showed that they may be more easily synthesized in the laboratory compared to the design template (compound **36**).

The predicted ADMET properties available in Table 12 show a very high human intestinal absorption for all tested compounds with **36** and the new compounds reaching 100%. All the tested molecules are substrates and inhibitors of p-glycoprotein, indicating that these molecules may easily mediate to reach their target sites.

Fig. 9 Binding interaction of PdxK with **36a** and **36b**

Fig. 10 Binding interaction of PdxK with **36c** and **36d**
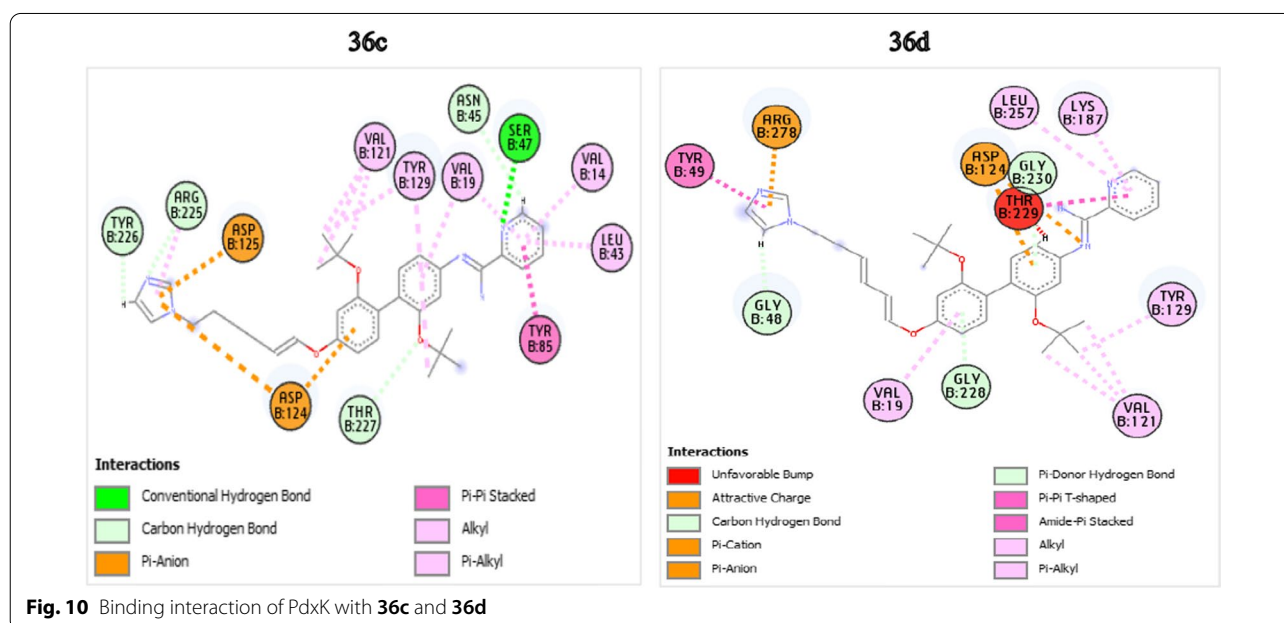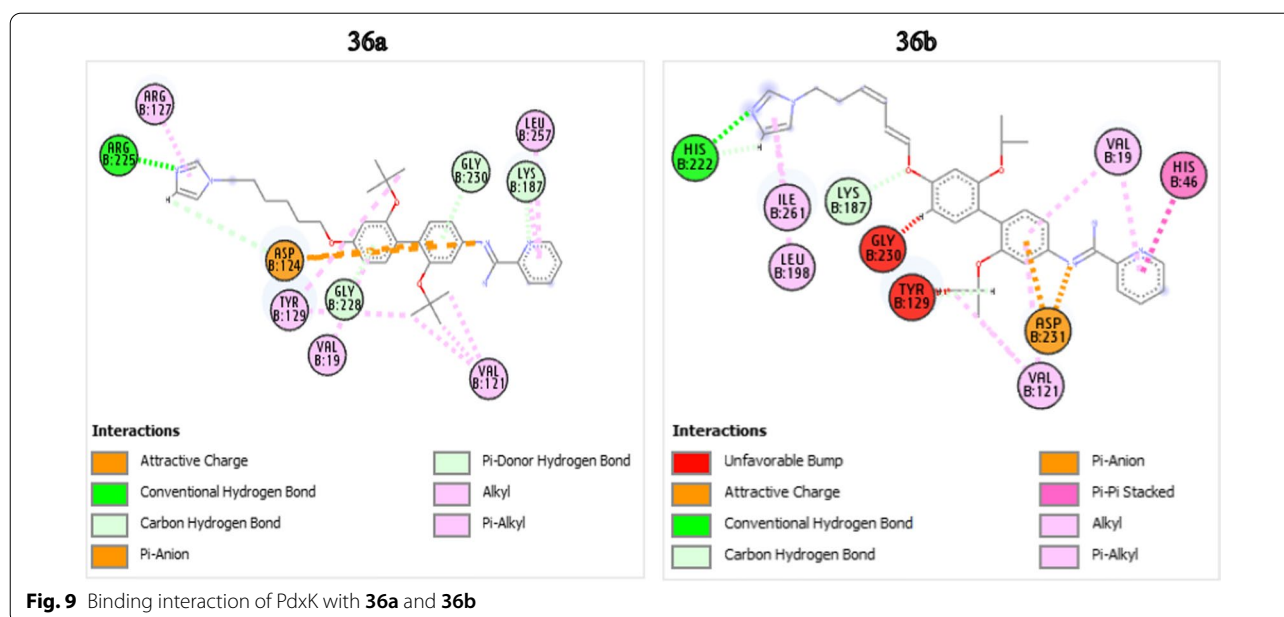
Skin permeability is a key factor in transdermal drug delivery development. Values of skin permeation constant LogKp > -2.50 connotes low skin permeability. As a result, the various compounds tested showed LogKp values < -2.50, indicating good skin permeability. Drug molecule penetration through the blood–brain barrier (BBB) and central nervous system (CNS) comes with certain criteria. For a drug molecule to be able to pass through the BBB and CNS readily, the logarithmic ratio of brain to plasma drug concentration (logBB) must be > 0.3 and the blood–brain permeability-surface area product (logPS) be > -2, respectively. As a result, all the selected arylimidamide-azole hybrids do not readily penetrate the BBB and the CNS. On the other hand, all the newly designed compounds are poorly distributed to the brain but are said to readily penetrate the CNS except **36b**.

Furthermore, some group of enzymes called cytochrome P450 enzymes are important in the body to facilitate drug metabolism and to help in their excretion. The two major isoforms enhancing drug

Fig. 11 Binding interaction of **36e** with PdxK
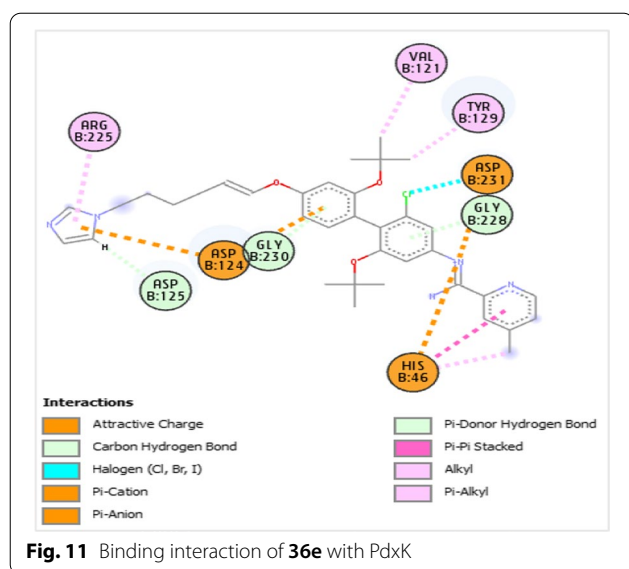
metabolism, CYP-34A and CYP-2D6, were tested. None of all the tested molecules is a substrate of CYP-2D6. However, all molecules are both substrates and inhibitors of CYP3A4, an indication of a well-moderated metabolic process. The extent of drug removal from the body is determined by the drug's total clearance (TCE). The range of values of TCE for all tested molecules is good. Compounds **22**, **36,** and all the newly designed molecules have no AMES toxicity, implying that they are non-mutagenic and cannot act as a carcinogen. Also available in Table 12 is the maximum recommended tolerated dose (MRTD) predicted for the various molecules. MRTD value of ≤ 0.477 log (mg/kg/day) is considered low, while a value > 0.477 log (mg/kg/day) is considered high. The overall drug-likeness and ADMET properties of the design template (**36**) and the newly designed compounds (except for **36f** and **36 g**) showed good pharmacokinetic profiles. Therefore, these molecules could be considered potential drug candidates for the treatment of leishmaniasis.

Ligand-based drug design (LBDD)
As stated earlier, the information encoded in the 3-D QSAR contour maps of the design template (**36**) was utilized in designing seven new arylimidamide-azole derivatives (36a–36 g). **36a** was obtained by replacing the isopropoxy groups of the template with a bulky tert-butoxy group to increase the steric bulk as observed from the contour maps. 36b was designed by reducing the steric bulk around the 2-position of the alkyl (hexyl) chain with the introduction of two double bonds (conjugated) to take off four hydrogens around that position. **36c** reduces the two double bonds in **36b** to one. **36d**

replaces the isopropoxy groups in **36c** with tert-butoxy groups. **36e** introduces the methyl group (a more bulky group than hydrogen) at the para-position of the pyridine ring system and also takes into account the electrostatic field effect by introducing the chloro group (electron-withdrawing group) at that position. **36f** further introduced the chloro group at the 4-position of the hexyl chain to deplete the electron density in that region. Finally, **36 g** further introduced two methyl groups at positions 3 and 5of the pyridine ring system.

The results of activity prediction and molecular docking conducted on the newly designed compounds (Table 13) showed that all the compounds have higher predicted inhibitory activities than the template (**36**) with only **36b** and **36 g** showing lower MolDock scores than **36**. Compound **36e** showed the highest MolDock score of -208. 595 kcal/mol and the second-highest predicted pIC_{50} of 7.375 only behind **36f** with a predicted pIC_{50} value of 7.670. Also, **36f** and **36 g** failed the drug-likeness test as both violated 2 out of 4 provisions of the Lipinski RO5. Hence, the interaction profile of compound **36e** will be discussed further. One C-H bond involving ASP-125 at a distance of 2.69Å and two π-donor hydrogen bonds with GLY-228 (3.32Å) and GLY-230 (2.83Å) were visible in the binding interaction profile of **36e** with PdxK. Electrostatic interaction types are attractive charge with ASP-231 and π-anion with HIS-46 and ASP-124. Others are hydrophobic interactions such as π–π stacked with HIS-46, alkyl with VAL-121, π-alkyl with HIS-46 and TYR-12, and halogen interaction with ASP-231. Interestingly, unlike the template and reference drug (pentamidine), most of the newly designed compounds have no unfavorable clashes with some amino acid residues, indicating a possible elimination or lowering of side effects that could be associated with the newly designed hybrids. Therefore, compound **36e** and the other newly designed analogs (except **36f** and **36 g**) have shown good pharmacological and pharmacokinetic properties and are hence worthy of selection for further practical evaluation in the laboratory as pyridoxal kinase inhibitors.

Conclusions
In this study, a five-descriptor 2-D QSAR model and a 3-D QSAR (CoMFA) model were developed with 36 arylimidamide-azole hybrids, both of which were found to satisfy the requirement for internal and external validation tests. The anti-leishmanial activities of the various compounds were well predicted by both models. A combined 2-D and 3-D QSAR approach was utilized to design and predict the inhibitory activities of seven new arylimidamide-azole analogs using compound **36** as the template. The molecular docking screening conducted

Ugbe *et al. Bulletin of the National Research Centre* (2022) 46:189

Page 23 of 24

between the 36 arylimidamide-azole compounds and the target receptor, pyridoxal kinase (PdxK), revealed compounds **1** and **36** as having the least and highest Mol-Dock scores of -142.632 kcal/mol and -198.260 kcal/mol, respectively, both higher than MolDock score of -141.793 kcal/mol reported for the reference drug pentamidine. The new compounds bind excellently into PdxK's cavities with binding affinities (MolDock scores) in the order: **36e** (-208.595 kcal/mol) > **36f** > **36a** > **36c** > **36d** > **36** (Template) > **36b** > **36 g** (-187.155 kcal/mol), while the predicted pIC_{50} follows the order: **36f** (7.670) > **36e** > **36 g** > **36c** > **36d** > **36a** > **36b** > Template (**36**) (6.344). Also, the newly designed analogs showed a good pharmacokinetic profile and obeyed Lipinski's RO5 for oral bioavailability except for **36f** and **36 g.** Special emphasis on **36e** because it appears to be the most consistent with the various employed validation protocols, being that it possessed the highest binding score, second-highest predicted pIC_{50}, showed excellent pharmacokinetic properties and binds adequately with the target protein. Therefore, this work has provided useful information on the arylimidamide-azole hybrids as antileishmanial agents.

Abbreviations
AD: Applicability domain; ADMET: Absorption, distribution, metabolism, excretion, and toxicity; ARG: Arginine; ASN: Asparagine; ASP: Aspartic acid; B3LYP: Becke's three-parameter read-Yang–Parr hybrid; BBB: Blood–brain barrier; CNS: Central nervous system; CoMFA: Molecular field analysis; CoMSIA: Comparative molecular similarity indices analysis; CYP-34A/CYP-2D6: Cytochrome p450 isoforms; DFT: Density functional theory; DTC: Drug Theoretics and Cheminformatics; ESOL: Estimated solubility; FFD: Fractional factorial design; F-test: Fischer's statistics; GFA: Genetic function approximation; GLN: Glutamine; GLY: Glycine; h*: Warning leverage; HBA: Hydrogen bond acceptor; HBD: Hydrogen bond donor; HIA: Human intestinal absorption; HIS: Histidine; IC_{50}: Half-maximal inhibitory concentration; LBDD: Ligand-based drug design; LMO: Leave many out; LOF: Friedman lack of fit; LogBB: Logarithmic ratio of brain to plasma drug concentration; LogPS: Blood–brain permeability-surface area product; LOO: Leave one out; LTO: Leave two out; LYS: Lysine; ME: Mean effect; MIF: Molecular interaction field; MLR: Multi-linear regression; MMFF: Molecular mechanics force field; MRTD: Maximum recommended tolerated dose; MW: Molecular weight; O3A: Open3DAlign; PaDEL: Pharmaceutical Data Exploration Laboratory; PC: Principal component; PDB: Protein data bank; PdxK: Pyridoxal kinase receptor; pIC_{50}: Negative log of IC_{50}; QSAR: Quantitative structure–activity relationship; RO5: Rule of five; SA: Synthetic accessibility; SDEC: Standard error of correlation; SDEP: Standard error of prediction; SEE: Standard error of estimation; SER: Serine; TCE: Total clearance; THR: Threonine; TPSA: Topological polar surface area; TYR: Tyrosine; UVEPLS: Uninformative variable elimination-partial least square; VIF: Variance inflation factor; VL: Visceral leishmaniasis.

Author contributions
GAS, AU, and FAU conceived and designed the study. FAU carried out the study and drafted the manuscript. IA conducted the technical editing. All authors read and approved the final manuscript.

References
Abdelhameed A (2018) The design and synthesis of cyanines and arylimidamide azole hybrids as antilesihmanaial agents, Ph.D. dissertation The Ohio State University. https://gradsch.osu.edu/calendar/events/design-and-synthesis-cyanines-and-arylimidamide-azole-hybrids-antileishmanial-agents: Accessed on 11/02/2022

Abdelhameed A, Feng M, Joice C, Zywot EM, Jin Y, Rosa CL, Liao X, Meeds HL, Kim Y, Li J, McElroy CA, Wang MZ, Werbovetz KA (2020) Synthesis and antileishmanial evaluation of arylimidamide—azole hybrids containing a phenoxy alkyl linker. ACS Infect Dis. https://doi.org/10.1021/acsinfecdis.0c00855

Abdullahi A, Shallangwa GA, Ibrahim MT, Bello AU, Arthur DE, Uzairu A, Mamza P (2019) QSAR studies on some C14-urea tetrandrine compounds as potent anti-cancer agents against Leukemia cell line (K562). JOTCSA 5(3):1391–1402

Abdullahi SA, Uzairu A, Shallangwa GA, Uba S, Umar AB (2022) In-silico activity prediction, structure-based drug design, molecular docking and pharmacokinetic studies of selected quinazoline derivatives for their antiproliferative activity against triple negative breast cancer (MDA-MB231) cell line. Bull Natl Res Centre 46:2

Adawara SN, Shallangwa GA, Mamza PA, Ibrahim A (2020) Molecular docking and QSAR theoretical model for prediction of phthalazinone derivatives as new class of potent dengue virus inhibitors. Beni-Suef Univ J Basic Appl Sci 9(50):1–17

Adeniji SE, Uba S, Uzairu A (2018) QSAR modeling and molecular docking analysis of some active compounds against mycobacterium tuberculosis receptor (Mtb CYP121). J Path Hindawi 2018:24–64

Adeniji SE, Uba S, Uzairu A (2019) Activity modeling of some potent inhibitors against mycobacterium tuberculosis using genetic function approximation approach. Adıyaman Univ J Sci 9(1):77–98

Adeniji SE, Arthur DE, Abdullahi M, Abdullahi A, Ugbe FA (2020) Computer-aided modeling of triazole analogues, docking studies of the compounds on DNA gyrase enzyme and design of new hypothetical compounds with efficient activities. J Biomol Struct Dyn

Al-Attraqchi OH, Mordi MN (2022) 2D- and 3D-QSAR, molecular docking, and virtual screening of pyrido [2, 3-d] pyrimidin-7-one-based CDK4 inhibitors. J Appl Pharmaceut Sci 12(01):165–175

Al-Tamimia AS, Mitxeltorenab ME, Sanmartínb C, Jiménez-Ruizc A, Syrjänend S, Parkkilae S, Sellerif S, Cartaf F, Angelif A, Supuranf T (2019) Discovery of new organoselenium compounds as antileishmanial agents. Bioorg Chem 86:339–345

Are S, Gatreddi S, Jakkula P, Qureshi IA (2020) Structural attributes and substrate specificity of pyridoxal kinase from Leishmania donovani. Int J Biol Macromol 152:812–827

Arthur DE, Uzairu A, Mamza P, Abechi SE, Shallangwa GA (2020) Activity and toxicity modeling of some NCI selected compounds against leukemia P388ADR cell line using genetic algorithm-multiple linear regressions. J King Saud Univ Science 32(1):324–331

Baquedano Y, Alcolea V, Toro MÁ, Gutiérrez KJ, Nguewa P, Font M, Moreno E, Espuelas S, Jiménez-Ruiz A, Palop JA, Plano D, Sanmartín C (2016) Novel

Ugbe *et al. Bulletin of the National Research Centre*   (2022) 46:189

Page 24 of 24

heteroaryl selenocyanates and diselenides as potent antileishmanial agents. Antimicrob Agents Chemother 60:3802–3812

Díaz M, de Lucio H, Moreno E, Espuelas S, Aydillo C, Jiménez-Ruiz A, Toro MÁ, Gutiérrez KJ, Martínez-Merino V, Cornejo A, Palop JA, Sanmartín C, Plano D (2019) Synthesis and leishmanicidal activity of novel urea, thiourea, and selenourea derivatives of diselenides. Antimicrob Agents Chemother 63:e02200-e2218. https://doi.org/10.1128/AAC.02200-18

Edache EI, Samuel H, Sulyman YI, Arinze O, Ayine OI (2020a) QSAR and molecular docking analysis of substituted tetraketone and benzyl-benzoate analogs as anti-tyrosine: a novel approach to anti-tyrosine kinase drug design and discovery. Chem Res J 5(6):79–100

Edache EI, Uzairu A, Mamza PA, Shallangwa GA (2020b) A comparative QSAR analysis, 3D-QSAR, molecular docking and molecular design of iminoguanidine-based inhibitors of HemO: A rational approach to antibacterial drug design. J Drugs Pharmaceut Sci

Edache EI, Uzairu A, Mamza PA, Shallangwa GA (2022) Computational modeling and analysis of the theoretical structure of thiazolino 2-pyridone amide inhibitors for Yersinia pseudo-tuberculosis and Chlamydia trachomatis Infectivity. Bull Sci Res 4(1):14–39

ElMchichi L, Belhassan A, Lakhlifi A, Bouachrine M (2020) 3D-QSAR study of the chalcone derivatives as anticancer agents. Hindawi J Chem. https://doi.org/10.1155/2020/5268985

Fan Y, Lu Y, Chen X, Tekwani B, Li X, Shen Y (2018) Anti-Leishmanial and cytotoxic activities of a series of maleimides: synthesis, biological evaluation and structure-activity relationship. Molecules 23:2878

Ghorbani M, Farhoudi R (2018) Leishmaniasis in humans: drug or vaccine therapy? Drug Des Devel Ther 12:25–40

Grohmann R, Schindler T (2008) Toward robust QSPR models: synergistic utilization of robust regression and variable elimination. J Comput Chem 29(6):847–860

Ibrahim MT, Uzairu A, Shallangwa GA, Uba S (2020) Lead identification of some anti-cancer agents with prominent activity against Non-small Cell Lung Cancer (NSCLC) and structure-based design. Chem Afr 3:1023–1044

Ibrahim MT, Uzairu A, Uba S, Shallangwa GA (2021) Design of more potent quinazoline derivative EGFRWT inhibitors for the treatment of NSCLC: a computational approach. J Pharmaceut Sci 7:140

Isyaku Y, Uzairu A, Uba S, Ibrahim MT, Umar AB (2020) QSAR, molecular docking, and design of novel 4-(N, N-diarylmethyl amines) Furan-2(5H)-one derivatives as insecticides against Aphis craccivora. Bull Natl Res Centre 44:44

Kakarla P, Inupakutika M, Devireddy AR, Gunda SK, Willmon TM, Ranjana KC, Shrestha U, Ranaweera I, Hernandez AJ, Barr S, Varela MF (2016) 3D-QSAR and contour map analysis of tariquidar analogues as Multidrug Resistance Protein-1 (MRP-1) inhibitors. Int J Pharm Sci Res 7(2):554–572

Kennard RW, Stone LA (1969) Computer aided design of experiments. Technometrics 11:137–148

Keurulainen L, Siiskonen A, Nasereddin A, Kopelyanskiy D, Sacerdoti-Sierra N, Leino TO, Tammela P, Yli-Kauhaluoma J, Jaffe CL, Kiuru P (2018) Synthesis and biological evaluation of 2-arylbenzimidazoles targeting Leishmania donovani. Bioorg Med Chem Lett

Kumar V, Sharma M, Rakesh BR, Malik CK, Neelagiri S, Neerupudi KB, Garg P, Singh S (2018) Pyridoxal kinase: a vitamin B6 salvage pathway enzyme from Leishmania donovani. Int J Biol Macromol 119:320–334

Lawal HA, Uzairu A, Uba S, (2021) QSAR, molecular docking studies, ligand-based design and pharmacokinetic analysis on Maternal Embryonic Leucine Zipper Kinase (MELK) inhibitors as potential anti-triple-negative breast cancer (MDA-MB-231cell line) drug compounds. Bull Natl Res Centre 45(90)

Lipinski CA, Lombardo F, Dominy BW, Feeney PJ (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46:3–26

Roy K, Chakraborty P, Mitra I, Ojha PK, Kar S, Das RN (2013) Some case studies on application of "rm2" metrics for judging quality of quantitative structure–activity relationship predictions: Emphasis on scaling of response data. J Comput Chem 34(12):1071–1082

Seifert K (2011) Structures, targets and recent approaches in anti-leishmanial drug discovery and development. Open Med Chem J 5:31–39

Tosco P, Balle T (2011) Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields. J Mol Model 17(1):201–208

Ugbe FA, Shallangwa GA, Uzairu A, Abdulkadir I (2021) Activity modeling, molecular docking and pharmacokinetic studies of some boron-pleuromutilins as anti-wolbachia agents with potential for treatment of filarial diseases. Chem Data Collect 36:100783

Ugbe FA, Shallangwa GA, Uzairu A, Abdulkadir I (2022a) Theoretical modeling and design of some pyrazolopyrimidine derivatives as Wolbachia inhibitors, targeting lymphatic flariasis and onchocerciasis. In Silico Pharmacol 10:8. https://doi.org/10.1007/s40203-022-00123-3

Ugbe FA, Shallangwa GA, Uzairu A, Abdulkadir I (2022b) Theoretical activity prediction, structure-based design, molecular docking and pharmacokinetic studies of some maleimides against Leishmania donovani for the treatment of leishmaniasis. Bull Natl Res Centre 46:92. https://doi.org/10.1186/s42269-022-00779-z

Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK (2011) Validation of QSAR models-strategies and importance. Int J Drug Des Discov 3:511–519

Verma J, Khedkar V, Coutinho E (2010) 3D-QSAR in drug design–a review. Curr Top Med Chem 10(1):95–115

Wang X, Dong H, Qin Q (2020) QSAR models on aminopyrazole-substituted resorcylate compounds as Hsp90 inhibitors. J Comput Sci Eng 48:1146–1156

Zhang X, Yan J, Wang H, Wang Y, Wang J, Zhao D (2020) Molecular docking, 3D-QSAR, and molecular dynamics simulations of thieno [3, 2-b] pyrrole derivatives against anticancer targets of KDM1A/LSD1. J Biomol Struct Dyn. https://doi.org/10.1080/07391102.2020.1726819

Publisher's Note