Taibah University

Journal of Taibah University Medical Sciences

Original Article

Ligand based-design of potential schistosomiasis inhibitors through QSAR, homology modeling, molecular dynamics, pharmacokinetics, and DFT studies

Saudatu C. Ja'afaru, MChem [a,b,*], Adamu Uzairu, PhD [a], Anshuman Chandra, PhD [c], Muhammed S. Sallau, PhD [a], George I. Ndukwe, PhD [a], Muhammad T. Ibrahim, PhD [a] and Imteyaz Qamar, PhD [d]

[a] Department of Chemistry, Ahmadu Bello University Zaria, Nigeria
[b] Department of Chemistry, Aliko Dangote University of Science and Technology, Wudil, Kano, Nigeria
[c] School of Physical Sciences, JawaharLal Nehru University, New Delhi, India
[d] School of Biotechnology, Gautam Buddha University, Greater Noida, India

الملخص

أهداف البحث: داء البلهارسيات هو مرض استوائي مهمل وسبب رئيسي للوفيات في المناطق المتضررة. في الوقت الحاضر، لا يوجد لقاح عملي لمكافحة داء البلهارسيات، مما يضع الاعتماد الكبير على الاستخدام الواسع النطاق لعقار البرازيكوانتيل. يثير الاستخدام الشامل لعقار برازيكوانتيل مخاوف بشأن ظهور المقاومة للأدوية. ونتيجة لذلك، هناك حاجة إلى أهداف علاجية جديدة ومركبات محتملة لمكافحة داء البلهارسيات.

طريقة البحث: تم تحسين أربعة وعشرين مشتقا قويا من البرازيكوانتيل من خلال نظرية الكثافة الوظيفية عند مستوى "ب3ليب/6-31ج". تم إنشاء نماذج العلاقة الكمية بين الهيكل والنشاط، وتم التحقق من صحتها إحصائيا وتم اختيار المرشح الرئيسي بهدف تطوير المزيد من الخيارات العلاجية مع تحسين الفعالية تجاه داء البلهارسيات. تم تقييم الطاقات البيولوجية وطاقة الارتباط للمركبات المصممة. كما تم إجراء دراسات الديناميكية الجزيئية والتشابه الدوائي والامتصاص والتوزيع والتمثيل الغذائي والإفراز والسمية والكثافة الوظيفية على المركبات المصممة حديثا.

النتائج: تم إنشاء خمسة نماذج لعلاقة الهيكل بالنشاط الكمي، من بينها تفوق النموذج 1 بمعلمات التحقق المفضلة، وتم اختياره لتحديد مرشح رئيسي. المعلمات الإحصائية الأخرى للنموذج المختار كانت تتراوح قيم عامل تضخم التباين من

1.242 إلى 1.678 ومعامل التخليط البالغ 0.747. تم تصميم خمسة مركبات جديدة أظهرت أنشطة متوقعة محسنة (تتراوح من 5.081 إلى 7.022)، متجاوزة كلا من الرصاص والبرازيكوانتيل المتوقع البالغ 5.545. يكشف تحليل محاكاة الديناميكيات الجزيئية عن تقارب الارتباط العالي للمركبات المقترحة تجاه المستقبل المستهدف. تظهر تقييمات الامتصاص والتوزيع والتمثيل الغذائي والإفراز والسمية والتشابه مع الأدوية الالتزام بمعايير ليبينسكي، مما يشير إلى السلامة الدوائية والفموية. كما يشير تحليل نظرية الكثافة الوظيفية إلى مقاومة التغير الإلكتروني أثناء التفاعلات الكيميائية.

الاستنتاجات: تظهر المركبات المقترحة خصائص الأدوية المحتملة، مما يشير إلى مدى ملاءمتها لمزيد من البحث لتعزيز خيارات علاج داء البلهارسيات.

الكلمات المفتاحية: جلوتاثيون اس-ترانسفيراز؛ داء البلهارسيات؛ البلهارسيا المنسونية؛ تصميم الأدوية القائم على الليجند؛ محاكاة الالتحام الجزيئي

Abstract

Objectives: Schistosomiasis, a neglected tropical disease, is a leading cause of mortality in affected geographic areas. Currently, because no vaccine for schistosomiasis is available, control measures rely on widespread administration of the drug praziquantel (PZQ). The mass administration of PZQ has prompted concerns regarding the emergence of drug resistance. Therefore, new therapeutic targets and potential compounds are necessary to combat schistosomiasis.

Methods: Twenty-four potent derivatives of PZQ were optimized via density functional theory (DFT) at the B3LYP/6-31G* level. Quantitative structureactivity relationship (QSAR) models were generated and statistically

* Corresponding address: Department of Chemistry, Ahmadu Bello University, Zaria, Nigeria.
E-mail: sjchinade@yahoo.co.uk (S.C. Ja'afaru)
Peer review under responsibility of Taibah University.

validated, and a lead candidate was selected to develop therapeutic options with improved efficacy against schistosomiasis. The biological and binding energies of the designed compounds were evaluated. In addition, molecular dynamics; drug-likeness; absorption, distribution, metabolism, excretion, and toxicity (ADMET); and DFT studies were performed on the newly designed compounds.

Results: Five QSAR models were generated, among which model 1 had favorable validation parameters (R^2_{train}: 0.957, R^2_{adj}: 0.941, LOF: 0.101, Q^2cv: 0.906, and R^2_{test}: 0.783) and was chosen to identify a lead candidate. Other statistical parameters for the chosen model included variance inflation factor values ranging from 1.242 to 1.678, and a Y-scrambling coefficient (cRp^2) of 0.747. Five new compounds were designed with improved predicted activity (ranging from 5.081 to 7.022) surpassing those of both the lead compound and PZQ (predicted pEC_{50} of 5.545). Molecular dynamics simulation revealed high binding affinity of the proposed compounds toward the target receptor. ADMET and drug-likeness assessments indicated adherence to Lipinski's rule of five criteria, thereby suggesting pharmacological and oral safety. In addition, DFT analysis indicated resistance to electronic alteration during chemical reactions.

Conclusion: The proposed compounds exhibited potential drug characteristics, thus indicating their suitability for further investigation to enhance schistosomiasis treatment options.

Keywords: Glutathione S-transferase; Ligand based drug design; Molecular docking simulations; *Schistosoma mansoni*; Schistosomiasis

Introduction

Schistosomiasis, also known as bilharzia or snail fever, is a tropical disease caused by parasitic flatworms belonging to the *Schistosoma* genus. This disease remains a prevalent and debilitating parasitic ailment on a global scale, affecting more than 200 million individuals in more than 70 countries, predominantly in sub-Saharan Africa, the Middle East, and certain parts of South America and Southeast Asia.[1–4] Schistosomiasis is a major public health concern, as highlighted in a study conducted by Rinaldo and colleagues.[5] The substantial burden on affected communities may lead to chronic illness, decreased productivity, and economic hardship.[6] Furthermore, according to research performed by Aula and others, prolonged exposure to *Schistosoma* species can lead to serious complications including liver fibrosis; urinary tract infections; and vulnerability to other infections and conditions including intestinal cancer, human immunodeficiency virus (HIV), infertility, renal failure, and neuroschistosomiasis.[7–10] In addition, Bishop and colleagues have emphasized that the persistent nature of the disease poses substantial challenges to control efforts.[11] Current treatments rely on a single drug, praziquantel (PZQ), and are hindered by drug resistance and limited effectiveness against juvenile worms.[12–14] Therefore, the discovery of more potent inhibitors targeting *Schistosoma* species is critical.

Glutathione S-transferase (GST) plays a critical role in the detoxification of harmful molecules and protecting parasites from oxidative stress.[15] Specifically, GST enzymes originating from *Schistosoma mansoni* play essential roles in the parasite's defense mechanisms against host immune responses and detoxification processes. *S. mansoni* GSTs (SmGST) bind and metabolize a diverse array of xenobiotic substances, including drugs, environmental toxins, and plant secondary metabolites.[16–18] Notably, SmGST has limited similarity to human GSTs, thus potentially enabling the development of selective inhibitors that target the parasite while preserving human GST enzymes, and enhancing drug safety. Numerous studies have emphasized the importance of phase I detoxification, and the central roles of glutathione-mediated detoxification and redox metabolism in the parasite.[19,20] Consequently, *Schistosoma* GSTs have potential not only as promising targets for drug development but also as candidates for the development of a schistosomiasis vaccine.

Park and co-workers have successfully synthesized a series of potent derivatives of PZQ and assessed their effectiveness against *S. mansoni*.[21] Their evaluation of worm movement before and after the addition of the derivatives demonstrated a dose-dependent decrease in mobility, thus indicating the promising anti-*Schistosoma* activity of the PZQ derivatives.[21] That study motivated further exploration of the PZQ derivatives' structural scaffolds against SmGST through quantitative structure–activity relationship (QSAR) analysis. Notably, traditional drug design faces challenges including long development times, expensive costs, and high attrition rates.[22–24] The limited understanding of biological processes and the difficulty in accessing a vast chemical space are further disadvantages. The evolution of computer-aided drug design has addressed these limitations by incorporating computational approaches[25–27] that take advantage of molecular modeling, virtual screening, and QSAR research, thus enabling faster, more cost-effective drug discovery with higher success rates.[28–30] Notably, QSAR is crucial in drug design, by facilitating efficient and cost-effective drug discovery by elucidating the relationships between compound structures and biological activity.[31–33] In this research, our objective was to create a QSAR model by using PZQ derivatives and to apply the model to identify a lead compound for creating new structural candidates with enhanced biological activity. Alongside the predictions, we incorporated molecular docking and molecular dynamics (MD) simulations to study protein–ligand interactions and stability under normal pharmacological conditions. We further assessed in silico ADMET and drug-likeness properties, and evaluated the DFT parameters for the newly designed analogs.

Materials and Methods

Dataset collection

A series of 24 PZQ derivatives, as potential SmGST inhibitors, were obtained from the literature.[21] With equation (1), the half-maximal effective concentration (EC_{50}) values, in units of μM, were converted to the corresponding logarithmic scale (pEC_{50}).[34] Table 1 illustrates the structural formula, actual biological activity (pEC_{50}), predicted activity, and residual value for all datasets.

$$pIC_{50} = -log\ IC_{50} \times 10^{-6} \tag{1}$$

Structure determination and optimization

The PZQ derivatives' two-dimensional molecular structures were drawn in ChemDraw software, according to the ACS Document 1996 guidelines for standardized chemical structures.[35,36] Subsequently, the two-dimensional structures were converted into three-dimensional format in Spartan 14 software, and their geometric energy was minimized with molecular mechanics forcefields.[37] To enhance geometrical accuracy, we subjected the minimized compounds to geometry optimization through DFT calculations, specifically using the B3LYP/6-31G* basis set, to obtain more reliable conformers.[38] The optimized conformers were saved in sdf format for the determination of molecular descriptors.

Determination of descriptors, pretreatment, and splitting of the dataset

We used the PaDEL descriptor tool kit was used to compute a variety of descriptors, including topological, autocorrelation, fragment count, and geometrical descriptors.[39,40] To enhance the accuracy of these descriptors, we performed a manual pretreatment process to remove redundant descriptors.[41] Furthermore, we used the Kennard-Stone data preprocessing tool and a standard variance threshold of 0.001 to remove constant descriptors, and a coefficient threshold of 0.8 to eliminate highly correlated descriptors. Additionally, we applied the Kennard–Stone data partitioning tool to divide the preprocessed data into two subsets: a model training set comprising 70% of the data and an external validation test set containing the remaining 30% of the data.[42–44]

QSAR model generation and validation

The QSAR models were generated in Material Studio software, and the genetic function approximation (GFA) algorithm was applied to select the most important descriptors from the compounds in the training set.[45–47] In this GFA regression, the biological activity (pEC_{50}) values were considered dependent variables, and the descriptor values were considered independent variables. To ensure successful convergence of the model, we set the following specific parameters: the population sample and maximum generation were fixed at 10,000 and 1500, respectively; the number of top equations returned was set to 5; the mutation probability was set to 0.1, and the user-defined smoothing parameter was maintained at the default value of 0.5. The evaluation of the constructed model was performed with the internal correlation coefficient (R^2).[48] In QSAR modeling, the modeling set R^2 is a commonly used factor for internal assessment that indicates how well the model explains the variation observed in the overall dataset.[49] A higher internal R^2 value (closer to unity) indicates a more reliable model.[50] Equation (2) was used to compute this correlation coefficient.

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

where Y_{obs}, Y_{pred}, and $\bar{Y}$ are the experimental, predicted, and average training set activity, respectively. However, relying solely on the internal R^2 value is insufficient to evaluate the model's robustness. Therefore, we calculated additional metrics, including the adjusted R^2 (R^2_{adj}), lack of fit (LOF), and cross-validation coefficient (Q^2_{cv}). The R^2 value was adjusted (R^2_{adj}) to generate a more reliable and stable model, as demonstrated in equation (3) below

$$R^2_{adj} = 1 - \frac{(1 - R^2)(n - 1)}{n - m - 1} \tag{3}$$

where m is the number of descriptors used in the MLR model, and n denotes the number of training set compounds.

The LOF, which measures the discrepancy between the experimental data and the model's predictions, is essential for assessing the model's quality, detecting overfitting, optimizing the model, and gaining insights into data distribution. A low LOF value obtained with equation (4) indicated a well-fitted model with enhanced predictive capability.[51]

$$LOF = \frac{SSE}{\left(1 - \frac{a + bp}{M}\right)^2} \tag{4}$$

where SSE represents the sum of squares of errors in the model, *a* is the number of terms in the model, b is a user-defined parameter, p is the total sum of descriptors in all model terms (excluding the constant term), and M is the number of compounds in the modeling set.

The cross-validation coefficient (Q^2cv) metric quantifies the predictive fitness of a model, thus providing a precise estimation of its performance in making predictions. Equation (5) was used to calculate this coefficient.

$$Q^2cv = 1 - \frac{\sum (Y_{pred} - Y_{exp})^2}{\sum (Y_{exp} - \bar{Y}_{training})^2} \tag{5}$$

To assess the effectiveness and reliability of the chosen model, we performed external evaluation on the developed QSAR models by using the test set compounds and calculating the external R^2 (R^2_{test}) to select the best model. For a reliable external regression equation, the R^2_{test} value should be close to 1.[51] The selected model's robustness was further examined through various parameters, including variance inflation factor (VIF), which indicates multicollinearity

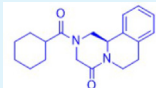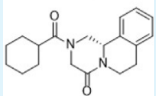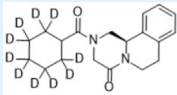**Table 1: Molecular structures, stereochemistry, and biological activity of praziquantel derivatives.**
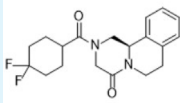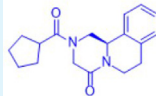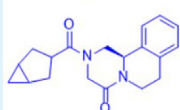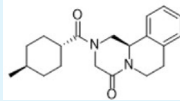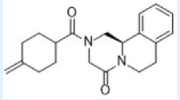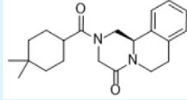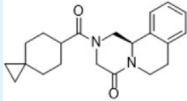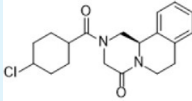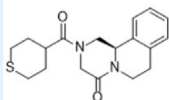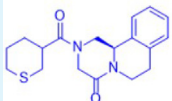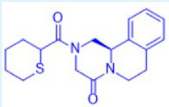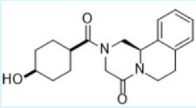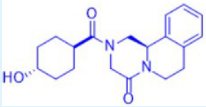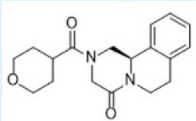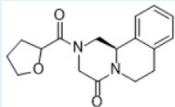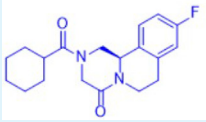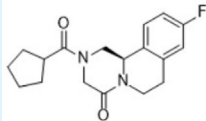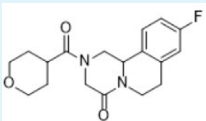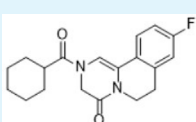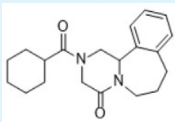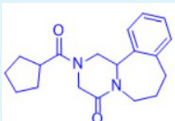
ID	Structure	pEC$_{50}$	Predicted pEC$_{50}$	Residual
1		6.301	6.377	−0.076
2		4.607	4.83	−0.223
3		6.523	6.377	0.145
4		5.602	5.683	−0.081
5		5.495	5.727	−0.232
6		5.495	5.629	−0.134
7		5.018	4.96	0.058
8		5.081	5.084	−0.004
9		4.77	4.678	0.091
10		4.706	4.821	−0.116
11		5.276	5.223	0.053
12		6.398	6.462	−0.064
13		5.602	6.087	−0.484

Table 1 (*continued*)

ID	Structure	pEC$_{50}$	Predicted pEC$_{50}$	Residual
14		5.796	6.11	−0.314
15		4.87	4.992	−0.122
16		4.745	4.759	−0.014
17		5.092	5.039	0.052
18		4.478	4.256	0.222
19		5.886	6.191	−0.305
20		5.398	5.509	−0.111
21		4.857	4.895	−0.038
22		4.863	4.929	−0.066
23		5.796	5.593	0.203
24		5.398	5.168	0.23
PZQ		—	5.545	—

Key: ▬ training set compounds, ▬ test set compounds.

among descriptors; the mean effect, which indicates the effects of descriptors on activity; and Y-randomization (cR^2p), which indicates the model's statistical significance.[52]

$$VIF = 1 - \frac{1}{1 - R^2} \qquad (6)$$

where R^2 is the correlation coefficient for each descriptor pair.

$$ME_x = \frac{A_x \sum\limits_{x=1}^{w=n} dwx}{\sum\limits_{x}^{m} A_x \sum\limits_{w}^{n} d_{wx}} \qquad (7)$$

where ME_x is the mean effect of descriptor x in the model, A_x is the coefficient of descriptor x, and dwx is the value of descriptor x in the data matrix for each compound in the training set. m is the sum of descriptors present in the model, and n is the number of compounds in the training set. Moreover, the Y-scrambling test involves reshuffling the actual activity values while keeping the descriptors constant, and assessing the validation parameter coefficient, cR^2p, for Y-randomization. We expected that the new QSAR model would exhibit low Q^2 and R^2 values, and cR^2p would be greater than 0.5.[53]

Applicability domain

The applicability domain (AD) technique was used to validate the predictions of the integrated GFA model. The process involved generating a Williams plot, wherein standardized residual values were plotted against leverage values (h) to assess the model's AD. Compounds with leverage scores exceeding the threshold value and standardized residuals beyond ± 3.0 standard deviation units (σ) were considered to be outside the defined AD of the QSAR model and were labeled as outliers.[54] The evaluation included analysis of the leverage approach and the warning leverage with equations (8) and (9), respectively:

$$h_i = j_i \left(J^T J \right) - K j_i^T \qquad (8)$$

$$h^* = \frac{3(j+1)}{q} \qquad (9)$$

where h_i is the leverage calculation method, J^T indicates the transpose matrix used in constructing the model, J represents the descriptor matrix with dimensions' $n \times k$ for the training datasets, h^* is the critical leverage value, j is the total number of descriptors of the chosen model, and q is the number of compounds in the training set.

Homology modeling

The SmGST target protein was obtained from the Universal Protein Resource Knowledge base (UniProtKB) webserver with accession code P09792. Protein validation was conducted in the SWISSMODEL online workspace (https://www.swissmodel.expasy.org). The Basic Local Alignment Search Tool (BLAST) and Hidden Markov model-based lightning-fast iterative sequence search (HHblits) were used to identify suitable templates for modeling. The most promising template was selected from the alignment results, and the ProMod3 modeling engine was used to create a novel protein model with minimized energy. The reliability of the three-dimensional structural model was assessed with the Qualitative model Energy Analysis (QMEAN) and Global Model Quality Estimation (GMQE) scores.[55-59]

Ligand based drug design

The criteria for selecting a lead compound for analog design relied exclusively on information acquired from the QSAR model. The chosen lead compound had the lowest residual value and reasonable pEC_{50}, fell within the preferred AD, and did not contravene Lipinski's rule of five (Ro5). Compound 17 from the PZQ derivatives was chosen as the lead compound and subsequently underwent modifications by substitution of various groups at specific positions indicated on the template compound. The selection of functional groups to generate new structural scaffolds was based on the significant mean effect value of the descriptors used in constructing the model.

Retrieval of target protein, preparations, and molecular docking studies

The homology model of SmGST was obtained in Protein Data Bank (PDB) format through the SWISS-MODEL interface. Discovery Studio was used to remove complex ligands, thus yielding the chosen modeled protein for this research. The protein was prepared for docking simulations with Molegro Virtual Docker (MVD) software. A surface was generated, and as many as five cavities were identified and fixed to identify potential binding sites. The optimized analogs were imported into MVD for the docking study. The optimal binding cavity had specific characteristics including a volume of 82.944 Å; surface of 308.41 Å; XYZ coordinates of 17.58, 54.81, and 36.01, respectively; and a radius of 15 Å. MolDock (Grid) scoring with a default grid resolution of 0.3 Å was applied. The docking simulation was run independently ten times, each with a maximum of 1500 iterations and a population size of 50.[60] After completion of the docking procedure, the MolDock score, Rerank score, and hydrogen bond energies were generated, on the basis of the binding effectiveness. The resulting docked complexes were saved in PDB format, and their interactions were visualized and interpreted in Discovery Studio software. For verification of the precision of the docking process, the crystallized ligand was repositioned within the active site of the target receptor. The resulting docked configuration was compared with the original crystallized ligand, and the root mean square deviation (RMSD) value was calculated. The docking procedures were validated by ensuring that

the RMSD value did not exceed a predetermined threshold of ≤ 2.0 Å.[38]

Molecular dynamics studies

MD simulations of SmGST protein in the unbound (apo) state and bound to potential anti-*Schistosoma* ligand were performed in the Desmond program developed by DE Shaw Research. The first step involved generating the necessary topology. Subsequently, the apo proteins and protein—ligand complexes were placed within the OPLS forcefield to examine the quantity and robustness of their interactions. Subsequently, these complexes were immersed in an SPC water model at a temperature of 300 K. Energy minimization was performed through 5000 steepest descent steps. Sodium and chloride ions were then introduced to mimic an in vivo environment. The calculation of long-range electrostatic interactions was accomplished with the particle-mesh Ewald method. A stable temperature and pressure were maintained through the application of a Nose—Hoover thermostat and the Martina—Tobias—Klein method. The integration of motion equations was performed with the multistep RESPA integrator, with a time step of 2.0 fs for interactions, both bonded and non-bonded, within the short-range cutoff. Periodic boundary conditions were also applied.

After the system equilibrated, the target proteins and their complexes with the most favorable potential anti-*Schistosoma* ligand were subjected to a production run for 100 ns in the N (total atoms in the system), P (system pressure), and T (system temperature) ensemble. The evaluation of RMSD, root mean square fluctuation (RMSF), full ligand contacts, and interaction fractions throughout the MD simulation provided valuable insights into the stability of the proteins and ligands in the bound state. Comparative RMSD and RMSF plots for the apo-form and protein—ligand complex were generated in the Simulation Interaction Diagram (SID) module from Schrödinger.

Drug-likeness and ADMET predictions

After successful docking of the designed analogs into the active site of the SmGST receptor, we evaluated the designed derivatives' ADMET properties and drug-likeness, to assess their potential as drug candidates. This analysis used the pkCSM (https://biosig.lab.uq.edu.au/pkcsm/) and Swiss-ADME (http://www.swissadme.ch/) web tools to assess ADMET profiles and drug-likeness characteristics.[61]

Density functional theory calculations

The structural and electronic properties of the newly designed compounds were computed with the DFT/B3LYP method with the 6-31G* basis set in Spartan 14. DFT calculations are a valuable computational tool for understanding the molecular properties, interactions, and reactivity of drug candidates, to further guide the rational design of more effective, safer drug candidates.[62] The generated parameters included the energies of the frontier molecular orbitals (HOMO, LUMO, and energy gap), as well as other reactivity parameters such as chemical hardness (h), softness (s), electronegativity (e), and chemical potential (p).[63]

List of abbreviations

ADMET: Absorption, distribution, metabolism, excretion and toxicity; B3LYP: Bee—3—Lee Yang Par; DFT: Density functional theory; GFA: Genetic function approximation; ME: Mean effect; MVD: Molegro Virtual Docker; PaDEL: Pharmaceutical Data Exploration Laboratory; PZQ: Praziquantel; Q^2cv: coefficient of determination for cross-validation; QSAR: Quantitative structure—activity relationship; R^2: internal validation coefficient; R^2_{adj}: adjusted coefficient of determination; SmGST: Schistosoma mansoni Glutathione S-transferase; SSE: sum of squares of errors in QSAR model; VIF: variance inflation factor.

Results

The experimental activity values of the derivatives were converted to pEC_{50} values for normalization. Five models were generated with the genetic function approximation combined with multi-linear regression (Table 2). Additionally, the predicted pEC_{50} values of the compounds were determined (Table 1). The residual values, i.e., the differences between the observed and predicted activity values of the compounds, were also computed (Table 1).

After model construction, an evaluation of internal validation parameters (R^2_{train}, R^2_{adj}, LOF, and Q^2cv) indicated that all five generated models demonstrated values above the recommended threshold (Table 3). However, during external validation (R^2_{test}), three of the five models passed, whereas the remaining two models failed, on the basis of negative R^2_{test} values.

Model 1, which exhibited the most favorable assessment parameters, was chosen for additional statistical validation to evaluate its predictive ability. The P-value, mean effect (ME), and VIF were computed (Table 4), with a significance threshold of $P < 0.05$ for all descriptors. VIF values ranging from 1.242 to 1.678 confirmed the absence of intercorrelation among the descriptors. In the ME analysis, positive values were observed for GATS4m, RDF90e, and L2s, whereas a negative value of -1.168 was observed for the RPCG molecular descriptor.

Table 2: Generated QSAR models.

Model 1	$pEC_{50} = -3.742 \times GATS4m + 62.635 \times RPCG - 0.137 \times RDF90e - 3.821 \times L2s + 12.259$
Model 2	$pEC_{50} = -3.607 \times GATS4m + 64.348 \times RPCG - 0.135 \times RDF90e - 3.879 \times L2e + 11.882$
Model 3	$pEC_{50} = 0.718 \times XlogP - 0.230 \times RDF90u + 0.208 \times RDF105m - 0.205 \times RDF85i + 10.404$
Model 4	$pEC_{50} = -0.351 \times minHBa + 0.552 \times XlogP - 0.506 \times RDF90p - 0.117 \times RDF40s + 10.769$
Model 5	$pEC_{50} = -4.047 \times GATS4m + 63.539 \times RPCG - 0.101 \times RDF90i - 3.827 \times L2s + 12.455$

Table 3: Validation parameters for all generated models, with their respective recommended threshold values.

Models	Validation parameters				
	R^2_{train}	R^2_{adj}	LOF	Q^2cv	R^2_{test}
1	0.957	0.941	0.101	0.906	0.783
2	0.955	0.938	0.106	0.899	0.763
3	0.954	0.938	0.107	0.911	−0.479
4	0.951	0.934	0.114	0.884	−0.111
5	0.951	0.933	0.114	0.890	0.787
Recommended threshold	*Close to 1*	*Close to 1*	*Low value*	*>0.5*	*>0.6*

Table 4: Pearson's correlation matrix, P-value, VIF, and ME of the chosen descriptors in QSAR model 1.

	GATS4m	RPCG	RDF90e	L2s	P-value	VIF	ME
GATS4m	1				2.32×10^{-4}	1.242	0.576
RPCG	−0.321	1			1.14×10^{-6}	1.351	−1.168
RDF90e	−0.055	0.028	1		7.78×10^{-8}	1.243	0.255
L2s	−0.372	0.452	−0.352	1	9.88×10^{-8}	1.678	1.337

Table 5: Y-scrambling assessment.

Model	R	R^2	Q^2
Original	0.919	0.844	0.734
Random 1	0.598	0.358	−0.125
Random 2	0.382	0.146	−0.553
Random 3	0.368	0.136	−0.220
Random 4	0.409	0.167	−0.505
Random 5	0.485	0.235	−0.563
Random 6	0.248	0.062	−0.937
Random 7	0.351	0.123	−0.504
Random 8	0.451	0.204	−0.830
Random 9	0.732	0.536	0.189
Random 10	0.250	0.062	−0.551
Random model parameters			
Average R:	0.427		
Average R^2:	0.203		
Average Q^2:	−0.460		
cRp^2:	0.747		
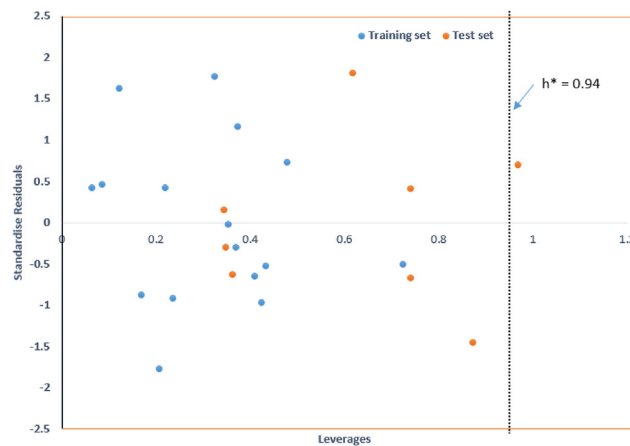
Figure 2: Williams plot of model 1.
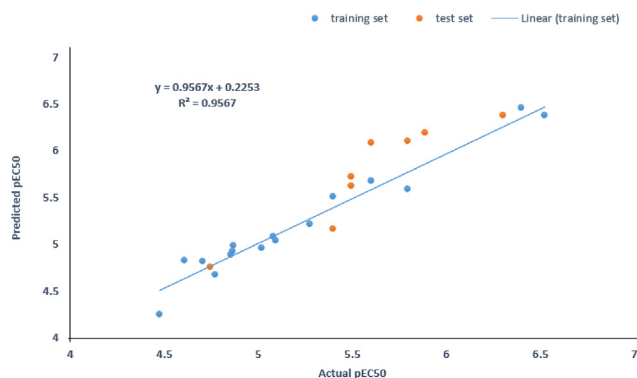
Figure 1: Activity plot of model 1.
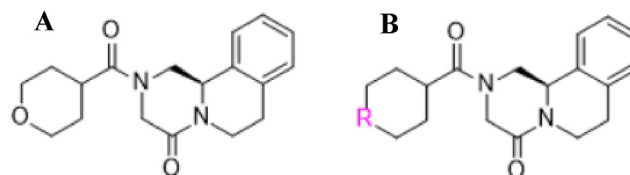
Figure 3: Structure of the lead compound (compound 17, (**A**)) and adopted design template (**B**).

An activity plot was used to further assess the concordance between the model's predicted activity and the experimental data (Figure 1). In the plot, narrower scatter indicates consistent accuracy, whereas wider scatter suggests greater prediction variability, as demonstrated by
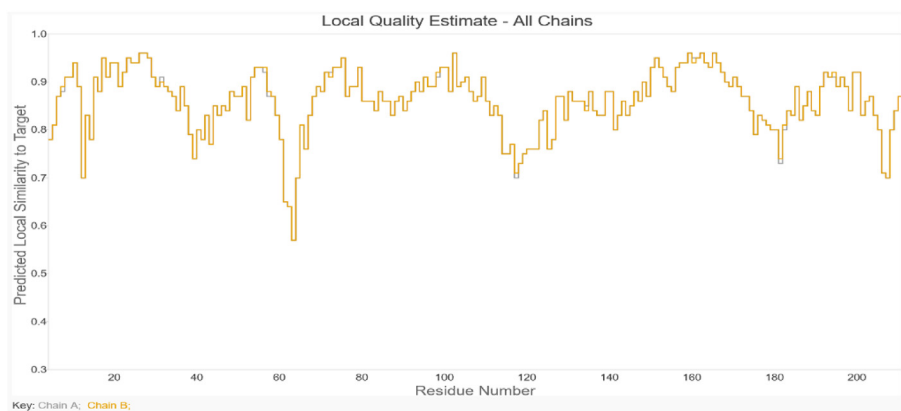
Figure 4: Plot of local quality estimate of the residues for structural validation of the SmGST homology model.

Table 6: Ramachandran statistical parameters of the SmGST homology model.

Model SmGST	Residues	%
Residues in most favored regions [A, B, L]	338	92.9
Residues in additional allowed regions [a, b, l, p]	21	5.8
Residues in generously allowed regions [~a, ~b, ~l, ~p]	3	0.8
Residues in disallowed regions	2	0.5
No. of non-glycine and non-proline residues	364	100
No. of end-residues (excluding Gly and Pro)	6	
No. of glycine residues (shown as triangles)	28	
No. of proline residues	20	
Total no. of residues	418	

Table 7: Molecular structures, predicted pEC$_{50}$, docking scores, and RMSD values of the newly designed novel SmGST inhibitors.

ID	Molecular structure	Predicted pEC$_{50}$	Moldock score/kcal mol^{-1}	RMSD/Å
17		5.039	−86.974	0.003
17a		7.022	−93.403	0.041
17b		6.687	−100.226	0.121
17c		6.339	−107.363	0.025
17d		5.299	−114.753	0.124
17e		5.081	−111.974	0.032
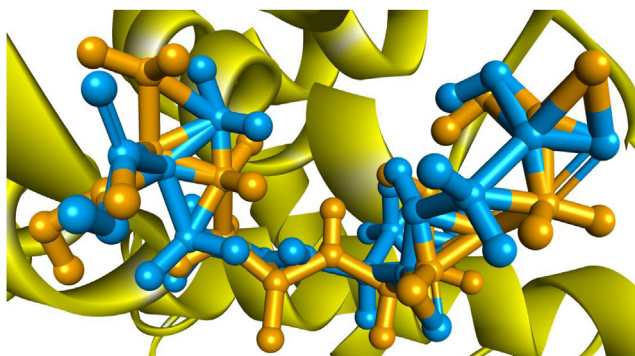
Figure 5: Superimposed co-crystallized ligand for docking validation, with RMSD of 1.387 Å.

Abdullahi and co-workers.[64] The spread of data points around the diagonal line indicated that the chosen model was robust and reliable.

Furthermore, to evaluate the model's robustness and determine whether the correlation was due to chance, we performed a Y-randomization test (Table 5). The R^2 and Q^2 values were 0.203 and -0.460, respectively, whereas the cR^2p of the model was 0.747.

A Williams plot was used to detect compounds outside the specified chemical space in which the model's predictions are considered reliable and accurate. Through analysis of leverage values and standardized residuals (Figure 2), we identified compound 16 as an influential outlier ($h^* > 0.94$).

After successfully passing all validation steps, and showing excellent predictive power, model 1 was used for ligand-based drug design to create novel and improved derivatives as potential anti-schistosomiasis agents. Using information from the AD plot and descriptors of the selected model, we chose compound 17 as the lead compound (Figure 3A) and applied a template for further design (Figure 3B). The exploration of new analogs provided valuable insights into the structure−activity

Table 8: Interactions of the designed compounds (17a−17e) with SmGST active site amino acid residues.

ID	Interacting amino acid residues	Types of interactions					
		Conventional H-bond	C−H bond	π−σ	π−π	Alkyl	π−alkyl
17	Glu106	−	✔ (2.310 Å)				
	Arg16					✔	
	Leu113					✔	✔
	His107						✔
	His110						✔
17a	Gly15	−	✔ (2.502 Å, 3.049 Å)				
	Phe211			✔	✔		
	Phe38				✔		
	Arg16					✔	
	His110						✔
	Pro210						✔
17b	Gly15	−	✔ (2.433 Å)				
	Tyr10		✔ (2.829 Å, 3.003 Å)				
	Phe211			✔	✔		
	Phe38				✔		
	Arg16					✔	
	Leu113					✔	
	His110			✔			✔
	Pro210						✔
17c	Gly15	−	✔ (2.395 Å)				
	Tyr10		✔ (2.935 Å, 3.016 Å)				
	Phe211			✔	✔		
	Phe38				✔		
	Arg16					✔	
	Leu113					✔	
	His110			✔			✔
	Pro210						✔
17d	Gly15	−	✔ (2.427 Å)				
	Tyr10		✔ (2.919 Å, 3.017 Å)				
	Phe211			✔	✔		
	Phe38				✔		
	Arg16					✔	
	Leu113					✔	
	His110			✔			✔
	Pro210						✔
17e	Lys120	✔	−	−			
	Phe38				✔		✔
	Pro210					✔	
	Leu113					✔	
	Phe211						✔

Figure 6: 2-Dimensional interactions of SmGST in complex with designed derivatives. **A**: compound 17a; **B**: compound 17b; **C**: compound 17c; **D**: compound 17d and **E**: compound 17e.

relationships of PZQ derivatives. Modifications at the three-ringed core were not well tolerated, and the introduction of polar functional groups, such as amines, hydroxy groups, and oxides, significantly decreased the activity. Novel compounds were designed through

modification at the R position, and their activity was predicted with the selected model.

The SWISS-MODEL template library was used to search for related structures matching the target sequence, by using BLAST and HHblits. The closest template found was the

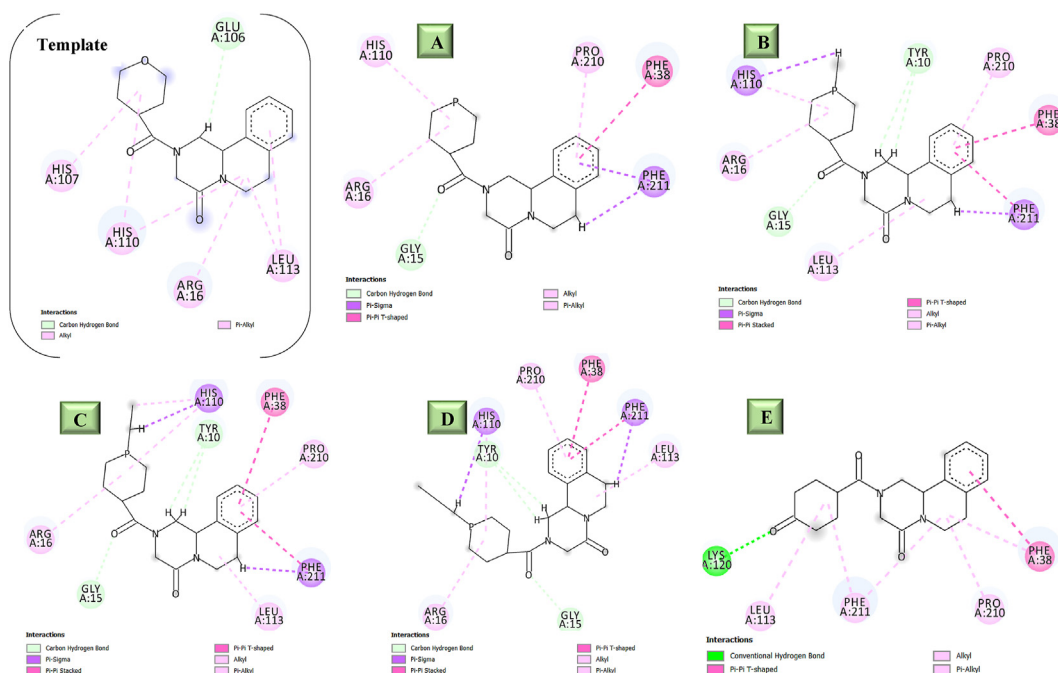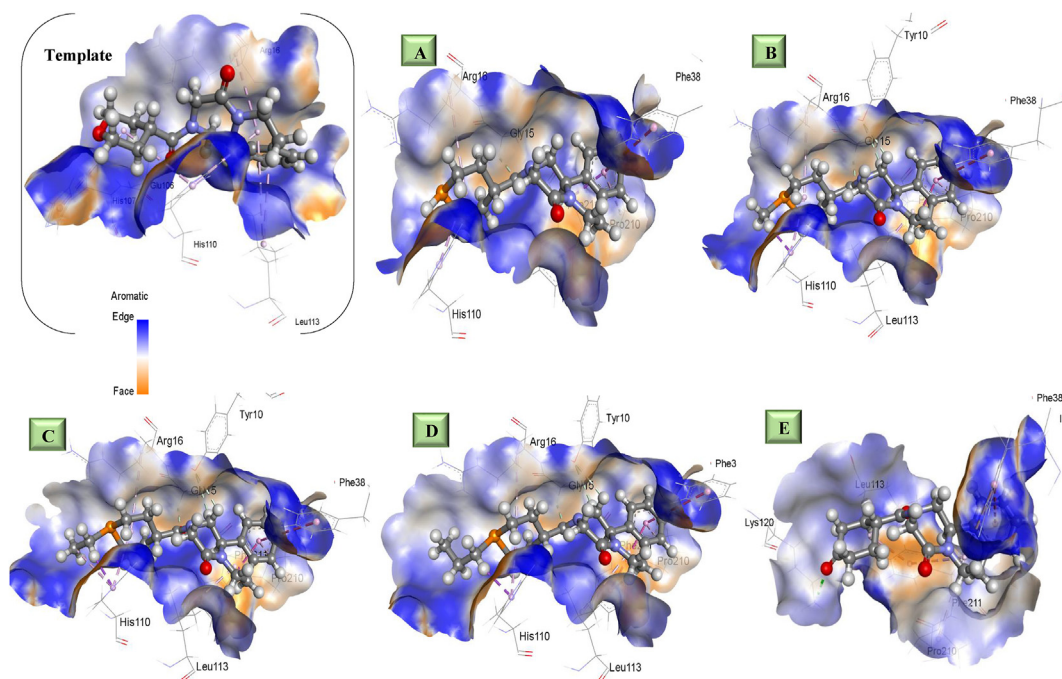**Figure 7:** Three-dimensional interactions of SmGST in complex with the designed derivatives. **A**: compound 17a; **B**: compound 17b; **C**: compound 17c; **D**: compound 17d, and **E**: compound 17e.
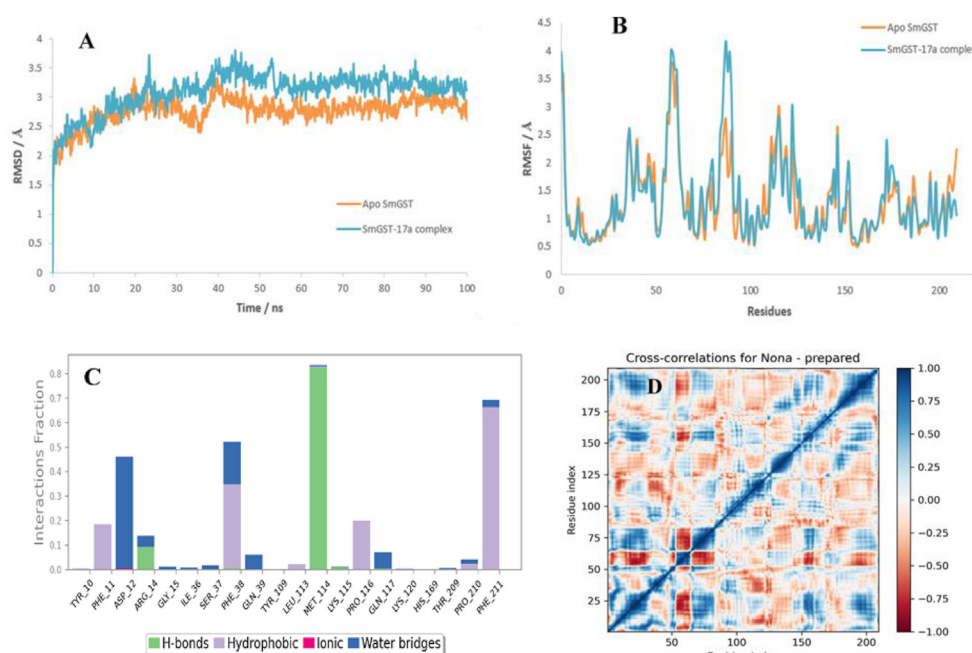
Figure 8: (**A**) RMSD plot of the apo SmGST and SmGST-17a complex, (**B**) RMSF plot of the apo SmGST and SmGST-17a complex, (**C**) SmGST-17a complex active site interactions, and (**D**) density correlation and covariance matrices of the SmGST-17a complex during MD simulations at 100 ns.

Table 9: Drug-likeness and ADMET parameters of the designed compounds.

ID	Drug-likeness					ADMET properties				
	Solubility class	Lipinski	B/S	Lead likeness	S/A	I/A (%)	CNS permeability (log PS)	Total clearance (log ml/min/kg)	Hepato-toxicity	Skin sensitization
17	+++	Yes	0.55	Yes	2.92	97.205	−2.890	0.376	Yes	No
17a	+++	Yes	0.55	Yes	3.24	100	−2.830	0.909	No	No
17b	+++	Yes	0.55	Yes	3.81	94.733	−2.262	0.302	Yes	No
17c	++	Yes	0.55	No	3.92	94.559	−2.194	0.333	No	No
17d	++	Yes	0.55	No	4.05	93.932	−2.123	0.352	Yes	No
17e	+++	Yes	0.55	Yes	2.89	96.230	−2.890	0.335	No	No

Key: +++ = soluble, ++ = moderately soluble, + = insoluble; B/S = bioavailability score; S/A = synthetic accessibility; I/A = human intestinal absorption.

SmGST class-mu 28 kDa isozyme (PDB: 1U3I.1.A), with a high level of sequence identity (99.53%) and query coverage (100%) (Figure 4, Table 6).

Multiple analogs were designed, five of which exhibited higher predicted activity than that of the lead compound. The predicted activity (pEC$_{50}$) of the designed analogs ranged from 5.081 to 7.022 (Table 7). Molecular docking studies of these designed analogs with the modeled SmGST target demonstrated stronger binding affinity in the range of −93.403 to −111.974 kcal/mol, surpassing that of the template (−86.974 kcal/mol) and the standard drug (−90.641 kcal/mol) (Table 7). Before docking of the designed compounds, we validated the docking process by repositioning the co-crystallized ligand. The resulting RMSD was 1.387 Å (Figure 5). Visualization of the docking studies in Discovery Studio revealed that most of the designed compounds formed carbon−hydrogen bond interactions with GLY15 and TYR10 within the

ligand−receptor complex. Furthermore, other hydrophobic interactions, including alkyl, π-alkyl, π−σ, and π−π interactions, were observed (Table 8, Figures 6 and 7).

An MD simulation lasting 100 ns was used to explore how the newly created derivatives behaved in the active sites of SmGST targets and to assess the stability of their complexes, according to the approach demonstrated by Hospital and colleagues.[65] Various parameters, including the RMSD, RMSF, interactions within the active site, and dynamic cross-correlation matrix plot, were computed to conduct an in-depth analysis (Figure 8).

Subsequent evaluation of drug-likeness and ADMET properties revealed the proposed compounds' promising potential as drug candidates with good ADMET properties (Table 9). Consequently, optimized derivatives were subjected to DFT calculations, thus yielding positive outcomes for the proposed compounds (Table 10). In addition, the investigation of HOMO and LUMO surface

Table 10: Frontier molecular orbital energies and quantum chemical descriptors of the designed compounds.

ID	E-HOMO/ Ev	E-LUMO/ eV	Band gap (ΔE)	Chemical hardness (η)	Chemical softness (σ)	Electronegativity (χ)	Chemical potential (μ)
17	-6.40	-0.38	6.02	3.01	0.33	3.39	-0.020
17a	-6.38	-0.38	6.00	3.00	0.33	3.38	-0.019
17b	-6.04	-0.36	5.68	2.84	0.35	3.20	-0.017
17c	-6.01	-0.35	5.66	2.83	0.35	3.18	-0.017
17d	-5.99	-0.35	5.64	2.82	0.35	3.17	-0.016
17e	-6.49	-0.46	6.03	3.02	0.33	3.48	-0.020
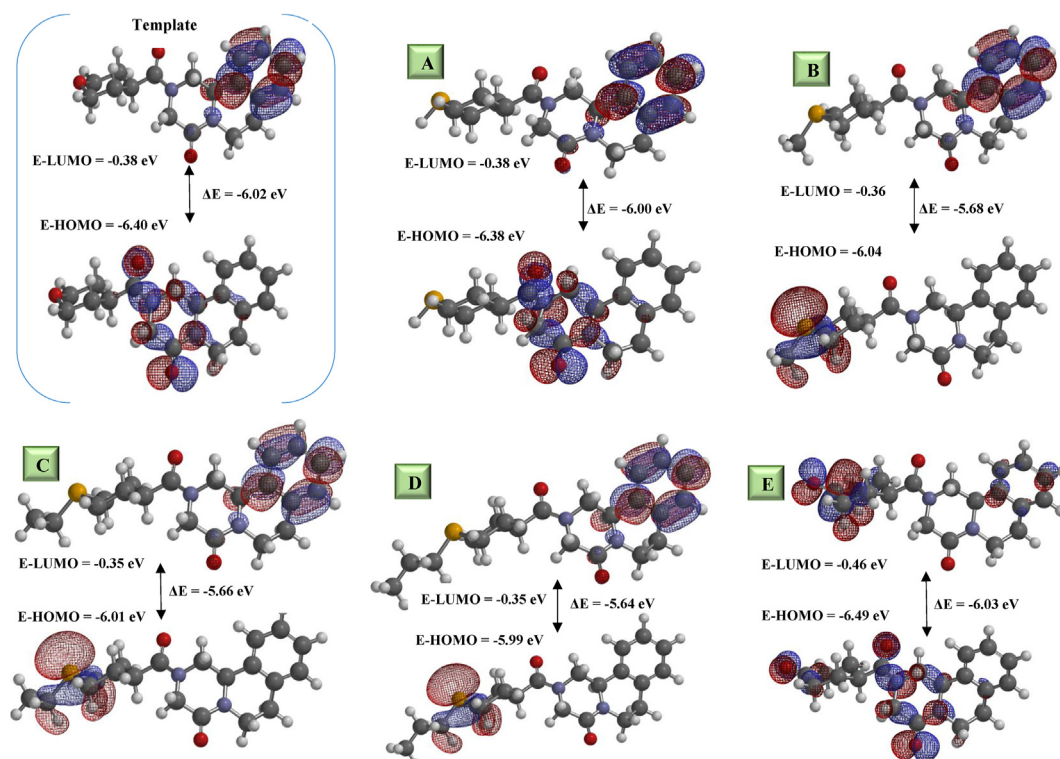
Figure 9: Frontier molecular orbital surface diagrams of the designed derivatives. **A**: compound 17a; **B**: compound 17b; **C**: compound 17c; **D**: compound 17d, and **E**: compound 17e.

diagrams provided valuable understanding of the molecular properties, reactivity, and binding interactions of the ligands (Figure 9).

Discussion

More drugs are needed to combat schistosomiasis, owing to factors including drug resistance, limited medicinal options, challenges in pediatric treatment, adverse effects, and the need to target various stages of the parasite's life.[6,12,14] Developing effective new drugs is crucial in the global effort to control and eliminate this widespread and debilitating disease.[14] To avoid the time-consuming and labor-intensive trial-and-error approach, QSAR modeling was used to screen, design, and predict the biological activity of potent PZQ derivatives, according to their chemical structures.[66] Five QSAR models were created and validated by using internal and external assessment parameters.[67]

Model 1 was chosen as the most suitable for activity prediction, because of its excellent validation parameters (R^2 of 0.957, R^2_{adj} of 0.941, LOF of 0.101, Q^2_{cv} of 0.906, and R^2_{test} of 0.783).

On the basis of the statistical outcomes (Table 3), the selected model surpassed the predefined threshold values for all parameters. This finding aligned with discoveries reported by Abdullahi and others.[60,68–70] The R^2 of 0.957 indicated that 95.6% of the overall variation in biological activity could be explained by the chosen model. This value indicates how well the model accommodates the compounds in the training dataset. The R^2_{adj} of 0.941 demonstrated the model's dependability and an absence of overfitting. Additionally, the Q^2_{cv} at 0.906 indicated the model's proficiency in predicting the activity of the compounds within the training dataset. The R^2_{test} value of 0.783 notably suggested the model's effectiveness in predicting the activity of the test set compounds. Recent research has demonstrated

that compound activity can be predicted when R^2 values exceed the widely accepted threshold of 0.6.[54]

The activity plot (Figure 1), illustrating the predicted pEC_{50} compared with the actual biological activity, indicated an R^2 value of 0.957, which aligned with the R^2 obtained from the selected QSAR model (R^2 of 0.957). A shared R^2 value between the activity plots and the QSAR model indicates model stability and highlights the chosen descriptor's strong predictive capacity for compound activity, thus underscoring that the model constructed around this descriptor is resilient and consistently effective in making predictions.[71]

Additional statistical evaluations of model 1 included P-values, VIF, mean effect, and Y-scrambling (Table 4). The calculated P-values ranged from 9.88×10^{-8} to 2.32×10^{-4}, thus indicating the model's robustness. A P-value below a chosen significance level, typically 0.05, indicates that the QSAR model is statistically significant and has no multicollinearity issues.[72,73] Multicollinearity among independent variables, specifically descriptors, was further assessed through VIF analysis. The VIF values for the selected model were within the range of 1.242−1.678, all of which were below the threshold value of 5. This finding indicated no significant multicollinearity among the descriptors and other independent variables.[73] A high VIF value, often within the range of 5−10, would suggest strong collinearity with other variables, thus potentially rendering the estimated coefficients unreliable.[74]

The mean effect values of the molecular descriptors present the physicochemical properties in numerical form and provide structural information for each descriptor.[71] Each descriptor has a unique ability to establish relationships with biological activity, thereby enabling the prioritization of compounds.[75] In model 1, positive molecular descriptors (GATS4m, RDF90e, and L2s) favorably affected anti-*Schistosoma* activity. Higher values of these descriptors correlate with greater activity of the derivatives, and vice versa. In contrast, the negative coefficient molecular descriptor (RPCG) had adverse effects on the activity of PZQ derivatives. Higher values of RPCG indicate lower activity, and vice versa.[76] These findings highlighted the importance of the three-ringed core and the fixed geometry with closely positioned atoms within the cyclic ring, thus resulting in shorter distances between neighboring atoms, and inducing the activity of PZQ derivatives.[21] The cRp^2 value of 0.747 suggested that the model's performance on the original dataset was approximately 74.7% of its performance on the scrambled dataset.[68] This finding indicated a moderate dependency on the actual structure for making accurate predictions.

Furthermore, the model's AD was evaluated with a Williams plot (Figure 2), contrasting standardized residuals with leverage values across the entire dataset. The plot revealed that compound 16 was an influential compound with leverage values exceeding the established threshold (h*) of 0.940.[77] This compound was found to lie outside the preferred domain, possibly because of changes in its stereochemistry. Notably, within the specified AD, compound 17 showed the lowest residual value and a favorable pEC50 value, and consequently was chosen as the lead compound for the ligand-based design of novel schistosomiasis inhibitors. Modifications were made at R

positions, as indicated on the template. The choice of the substituent was guided by the RCPG and L2s descriptors, which are known to have the highest ME values.[78] On the basis of these descriptors, we designed five new analogs with higher activity than that of the lead compound. The incorporation of phosphate-containing groups and a carbonyl functional group improved the activity of the compounds, in a manner potentially influenced by changes in the distance between atoms within the compound and/or affecting the ring profile.

A homology model was constructed with a high GMQE score of 0.97 and QMEAN score of 0.99, thus indicating confidence and robustness. Local assessment parameters indicated reasonable quality estimates for the model, and most residue scores were close to 1 (Figure 4). The Ramachandran statistical parameters demonstrated that 92.9% of the residues were in the most favored region (Table 6), thereby validating the suitability of the SmGST homology model for anti-schistosomiasis drug discovery and development.

The docking method's credibility was confirmed by re-docking of the co-crystallized ligand, which resulted in an RMSD value of 1.387 (Figure 5), within the accepted threshold of RMSD ≤2.0.[38] The MVD docking procedure accurately positioned the co-crystallized ligand within the SmGST binding site, thus demonstrating the method's efficiency. The results of molecular docking studies conducted on the five newly designed analogs against the SmGST model protein yielded favorable results. All five derivatives displayed MolDock scores within the range of −93.403 to −114.753 kcal/mol, surpassing that of the template, at −86.974 kcal/mol. The docking studies indicated that the analogs effectively bound the receptor's active site, as evidenced by their favorable MolDock scores in comparison to that of the template. This finding indicated the promising potential of the designed compounds, in terms of their binding affinity. Among the compounds, 17a was notably the most active designed PZQ derivative, because of its highest predicted activity score of 7.022, and impressive stability, as reflected by the MolDock score and re-rank score. Compound 17a forms several interactions with amino acid residues within the binding site. Notably, it engages in two carbon−hydrogen bonding interactions between the phosphinane ring moiety and the Gly10 residue, at distances of 2.502 Å and 3.049 Å. Additionally, the hydrogen from the substituted piperazine scaffold establishes a hydrophobic π−σ interaction with Phe211 at 2.748 Å. Furthermore, compound 17a participates in various hydrophobic interactions, including a π−π stacking interaction with the Phe211 residue at 5.189 Å and a π−π T-shape interaction with Phe38 at 4.975 Å. Other hydrophobic interactions include alkyl and π-alkyl interactions with the Arg16, Phe211, Phe38, and Pro210 residues. The molecular interactions of the generated PZQ derivatives are depicted in Figures 6 and 7, and summarized in Table 8. All proposed PZQ derivatives exhibited superior binding interactions, improved hydrogen bond energies, and re-rank scores to those of both the lead compound 17 and PZQ.

We performed MD simulation to gain insights into the dynamic behaviors of atoms within protein targets and their complexes with ligands, as previously demonstrated.[65,79,80] This method provides valuable information regarding the

structural stability of both the target proteins and ligands, both before and after their interaction, as highlighted by De Vivo et al.[81] Notably, the RMSD of SmGST backbone atoms for both the apo form and the SmGST-17a complex was assessed to understand the dynamics of ligand binding (Figure 8A). The apo SmGST exhibited an RMSD range of 2.5–3.0 Å after equilibration, and the RMSD increased from 0 to 35 ns as the simulation stabilized. Subsequently, a consistently low RMSD value of less than 3.0 Å was maintained throughout the 100 ns simulation, thereby suggesting continuous conformational sampling of the apo-protein during the simulation. In contrast, the RMSD of the complex stabilized at approximately 20 ns and exhibited minimal deviations during the rest of the simulation (Figure 8A). The apo and 17a bound SmGST depicted similar RMSD patterns, thus suggesting that the ligand 17a bound complex is stable. The RMSF was also monitored to gauge the degree of local fluctuations of residues with respect to their average positions during the simulation. The observed RMSF values were largely similar for the SmGST-17a complex and the apo SmGST (Figure 8B). Although the protein maintained its secondary structure, we observed substantial fluctuation in the region Lys56-lys66 and Glu86-Ser94, corresponding to the secondary structure. These changes were due to the accommodation of ligand 17a, which made multiple contacts with the protein, thus minimizing fluctuations. Ile9-Gly15, Ile36-Gln39, Tyr109, Leu113-Gln117, Lys120, and His169 were among the points of contact with ligand 17a, and were stabilized by its binding. Our overall analysis indicated that 17a confers stability within the target binding site. Furthermore, the interactions between SmGST and 17a throughout the 100 ns simulation were monitored and categorized into four types: hydrogen bonds, hydrophobic, ionic, and water bridges. Intriguingly, nearly all binding amino acids identified in the molecular docking simulation of 17a were consistently observed during the 100 ns simulation (Figure 8C). Notably, Gly15 maintained 3% water bridge contacts with the ligand, Phe38 exhibited approximately 35% hydrophobic and 17% water bridge contacts, and Phe211 had a 0.7 value indicating that it maintained a hydrophobic interaction for 70% of the simulation duration (Figure 7C). To gain further insights into the dynamic behavior of the system, we performed dynamic cross-correlation matrix analysis, which generates a two-dimensional matrix illustrating correlations in residue motions throughout the MD simulation timeline. Regions with dense coloration represent positive and negative correlations between protein residues, whereas uncolored regions indicate no correlation in residue movement. The prominently correlated matrix in Figure 8D highlighted the strong interactions between the ligand molecule (17a) and the SmGST binding site, which result in coordinated motions across the entire protein structure. This high correlation underscores the good binding of the ligand to SmGST. The MD simulation analysis suggested that ligand 17a has high binding affinity toward SmGST, and the complex remains stable after binding.

Additionally, the drug-likeness and ADMET assessment indicated that the compounds met all parameters with reasonable synthetic accessibility scores.[82,83] The drug-likeness of the formulated entities was assessed: three

exhibited high water solubility, whereas two displayed moderate solubility (Table 9). All designed compounds adhered to Lipinski's Ro5, satisfying all criteria without any violations. The bioavailability score for the entire set of designed analogs was determined to be 0.55, thus indicating their promising potential for absorption and distribution within tissues. Three of the five compounds emerged as potential lead candidates for further exploration in drug discovery. Furthermore, the synthetic accessibility of the compounds was evaluated on a scale from 1 to 10, where 1 indicates easy synthesis, and 10 indicates difficult synthesis. The analyzed compounds were within the range of 2.89–4.05, thereby indicating favorable synthetic accessibility (Table 9). The intestinal absorption ranged from 93.932% to 100%. The central nervous system permeability was found to comply with the recommended value of log PS < -3. Additionally, the total clearance values ranged from 0.302 to 0.909, thereby suggesting favorable rates of drug elimination. Notably, all analogs exhibited zero skin sensitization potential and consequently their safety in terms of allergic reactions. However, caution is warranted, because three of the five compounds were found to display hepatotoxicity; therefore, further investigation and consideration during drug development are warranted.

Subsequently, we subjected the optimized analogs to DFT calculations, which yielded more positive outcomes for the proposed analogs than the template compound (Table 10). In addition, the HOMO and LUMO surface diagrams provided valuable understanding of the molecular properties, reactivity, and binding interactions of the ligands (Figure 9). The frontier molecular orbitals (HOMO and LUMO) of the designed ligands play crucial roles in charge–transfer interactions with the target's active site.[63] Ligands with high HOMO energy are good electron donors, whereas those with lower energy are weak electron acceptors. The small energy gap in the compounds facilitates intermolecular charge transfer and enhances their bioactivity by promoting electron movement. The ligands exhibit high η and low σ, thereby indicating resistance to electronic alteration during chemical reactions. The χ values of 3.17–3.47 eV suggest that the compounds act as electron donors. The negative μ values indicate good stability and the formation of stable complexes with the receptor. Overall, these molecular properties positively influence the ligands' binding affinity and potential as inhibitors for SmGST.[78]

Conclusions

Our results demonstrated that the genetic function approximation-derived model 1 exhibited the highest effectiveness, on the basis of its strong alignment with both internal and external validation parameters: R^2 of 0.957, R^2_{adj} of 0.941, LOF of 0.101, Q^2cv of 0.906, and R^2_{test} of 0.783. Furthermore, the five potent analogs developed through ligand-based techniques displayed greater activity than that of the lead compound. These compounds exhibited reasonable interactions within the active site of SmGST, and stable ligand-protein binding, as indicated by the 100 ns MD simulation studies. The drug-likeness and ADMET parameter assessment indicated compliance with Lipinski's Ro5 with reasonable

ADMET parameters. However, caution is warranted, because three of the five compounds displayed hepatotoxicity. Consequently, further investigation and consideration during drug development are necessary. The quantum chemical parameters computed through DFT calculations indicated the structural stability of the selected molecules, and the potential for formation of stable complexes with the receptor. Comprehensive computational investigations strongly suggested that the chosen compounds have potential as viable candidates against the SmGST target and may serve as promising agents in combating schistosomiasis.

Source of funding

Conflict of interest

The authors have no conflict of interest to declare.

Ethical approval

Not applicable.

Consent

Not applicable.

Authors contributions

SCJ designed and performed the study, interpreted the results, and wrote the manuscript. AU provided supervision and edited the manuscript. AC and IQ performed MD simulations and edited the manuscript. MSS interpreted the results and edited the manuscript. GIN interpreted the results and edited the manuscript. MTI interpreted the results and edited the manuscript. All authors have critically reviewed and approved the final draft and are responsible for the content and similarity index of the manuscript.

Acknowledgment

Availability of data and material

All the data produced or examined throughout this research are incorporated within the article.

References

1. Wall KM, Kilembe W, Vwalika B, Dinh C, Livingston P, Lee YM, et al. Schistosomiasis is associated with incident HIV transmission and death in Zambia. **PLoS Negl Trop Dis 2018**; 12(12):e0006902.

2. LoVerde PT, Schistosomiasis. **Advances in experimental medicine and biology**, 1154; 2019. pp. 45−70.

3. Schistosomiasis [https://www.nhs.uk/conditions/schistosomiasis/].

4. Schistosomiasis [https://www.who.int/news-room/fact-sheets/detail/schistosomiasis].

5. Rinaldo D, Perez-Saez J, Vounatsou P, Utzinger J, Arcand J-L. The economic impact of schistosomiasis. **Infect Dis Poverty 2021**; 10(1): 1−12.

6. Mawa PA, Kincaid-Smith J, Tukahebwa EM, Webster JP, Wilson S. Schistosomiasis morbidity hotspots: roles of the human host, the parasite and their interface in the development of severe morbidity. **Front Immunol 2021**; 12.

7. Aula OP, McManus DP, Jones MK, Gordon CA. Schistosomiasis with a focus on Africa. **Trop Med Infect Dis 2021**; 6(3).

8. Patel P, Rose CE, Kjetland EF, Downs JA, Mbabazi PS, Sabin K, et al. Association of schistosomiasis and HIV infections: a systematic review and meta-analysis. **Int J Infect Dis 2021**; 102: 544−553.

9. Kayange NM, Smart LR, Downs JA, Maskini M, Fitzgerald DW, Peck RN. The influence of HIV and schistosomiasis on renal function: a cross-sectional study among children at a hospital in Tanzania. **PLoS Negl Trop Dis 2015**; 9(1): e0003472.

10. Abdel-Naser MB, Altenburg A, Zouboulis CC, Wollina U. Schistosomiasis (bilharziasis) and male infertility. **Andrologia 2019**; 51(1):e13165.

11. Bishop H. Menace of schistosomiasis: its true neglected nature in Nigeria. **MOJ Public Health 2017**; 6(5): 421−426.

12. Molehin AJ, McManus DP, You H. Vaccines for human schistosomiasis: recent progress, new developments and future prospects. **Int J Mol Sci 2022**; 23(4).

13. Hoekstra PT, Casacuberta-Partal M, van Lieshout L, Corstjens PLAM, Tsonaka R, Assaré RK, et al. Limited efficacy of repeated praziquantel treatment in Schistosoma mansoni infections as revealed by highly accurate diagnostics, PCR and UCP-LF CAA (RePST trial). **PLoS Negl Trop Dis 2022**; 16(12):e0011008.

14. Cioli D, Pica-Mattoccia L, Basso A, Guidi A. Schistosomiasis control: praziquantel forever? **Mol Biochem Parasitol 2014**; 195(1): 23−29.

15. Kim Y, Cha SJ, Choi HJ, Kim K. Omega class glutathione S-transferase: antioxidant enzyme in pathogenesis of neurodegenerative diseases. **Oxid Med Cell Longev 2017**; 2017:5049532.

16. Padi N, Akumadu BO, Faerch O, Aloke C, Meyer V, Achilonu I. Engineering a pseudo-26-kDa schistosoma glutathione transferase from bovis/haematobium for structure, kinetics, and ligandin studies. **Biomolecules 2021**; 11(12).

17. Keen JH, Jakoby WB. Glutathione transferases. Catalysis of nucleophilic reactions of glutathione. **J Biol Chem 1978**; 253(16): 5654−5657.

18. Akumadu BO, Pandian R, Olfsen J, Worth R, Thulo M, Mentor T, et al. Molecular basis of inhibition of Schistosoma japonicum glutathione transferase by ellagic acid: insights into biophysical and structural studies. **Mol Biochem Parasitol 2020**; 240:111319.

19. Balloul JM, Sondermeyer P, Dreyer D, Capron M, Grzych JM, Pierce RJ, et al. Molecular cloning of a protective antigen of schistosomes. **Nature 1987**; 326(6109): 149−153.

20. McTigue MA, Williams DR, Tainer JA. Crystal structures of a schistosomal drug and vaccine target: glutathione S-transferase from Schistosoma japonica and its complex with the leading antischistosomal drug praziquantel. **J Mol Biol 1995**; 246(1): 21−27.

21. Park SK, Friedrich L, Yahya NA, Rohr CM, Chulkov EG, Maillard D, et al. Mechanism of praziquantel action at a

parasitic flatworm ion channel. **Sci Transl Med 2021**; 13(625): eabj5832.

22. Abdullahi SH, Uzairu A, Shallangwa GA, Uba S, Umar AB. Ligand-based drug design of quinazolin-4(3H)-ones as breast cancer inhibitors using QSAR modeling, molecular docking, and pharmacological profiling. **J Egypt Natl Cancer Inst 2023**; 35(1): 24.

23. Hassan Baig M, Ahmad K, Roy S, Mohammad Ashraf J, Adil M, Haris Siddiqui M, et al. Computer aided drug design: success and limitations. **Curr Pharm Des 2016**; 22(5): 572−581.

24. Abdullahi SH, Uzairu A, Danazumi AU, Finbarrs-Bello E, Umar AB, Shallangwa GA, et al. Computational design of quinoxaline molecules as VEGFR-2 inhibitors: QSAR modelling, pharmacokinetics, molecular docking, and dynamics simulation studies. **Biocatal Agric Biotechnol 2023**; 51:102787.

25. Abdullahi SH, Moin AT, Uzairu A, Umar AB, Ibrahim MT, Usman MT, et al. Molecular docking studies of some benzoxazole and benzothiazole derivatives as VEGFR-2 target inhibitors: in silico design, MD simulation, pharmacokinetics and DFT studies. **Intell Pharm 2023**.

26. Bassani D, Moro S. Past, present, and future perspectives on computer-aided drug design methodologies. **Molecules 2023**; 28(9): 3906.

27. Singh DB, Pathak RK. Computational approaches in drug designing and their applications. **Exp Protoc Biotechnol 2020**: 95−117.

28. Chtita S, Belaidi S, Qais FA, Ouassaf M, AlMogren MM, Al-Zahrani AA, et al. Unsymmetrical aromatic disulfides as SARS-CoV-2 Mpro inhibitors: molecular docking, molecular dynamics, and ADME scoring investigations. **J King Saud Univ Sci 2022**; 34(7):102226.

29. Nour H, Daoui O, Abchir O, ElKhattabi S, Belaidi S, Chtita S. Combined computational approaches for developing new anti-Alzheimer drug candidates: 3D-QSAR, molecular docking and molecular dynamics studies of liquiritigenin derivatives. **Heliyon 2022**; 8(12):e11991.

30. Chtita S, Fouedjou RT, Belaidi S, Djoumbissie LA, Ouassaf M, Qais FA, et al. In silico investigation of phytoconstituents from Cameroonian medicinal plants towards COVID-19 treatment. **Struct Chem 2022**; 33(5): 1799−1813.

31. Kwon S, Bae H, Jo J, Yoon S. Comprehensive ensemble in QSAR prediction for drug discovery. **BMC Bioinform 2019**; 20(1): 521.

32. Nour H, Abchir O, Belaidi S, Qais FA, Chtita S, Belaaouad S. 2D-QSAR and molecular docking studies of carbamate derivatives to discover novel potent anti-butyrylcholinesterase agents for Alzheimer's disease treatment. **Bull Kor Chem Soc 2022**; 43(2): 277−292.

33. Abdullahi SH, Uzairu A, Shallangwa GA, Uba S, Umar AB. Computational modeling, ligand-based drug design, drug-likeness and ADMET properties studies of series of chromen-2-ones analogues as anti-cancer agents. **Bull Natl Res Cent 2022**; 46(1): 177.

34. Aminu KS, Uzairu A, Umar AB, Ibrahim MT. Salicylic acid derivatives as potential α-glucosidase inhibitors: drug design, molecular docking and pharmacokinetic studies. **Bull Natl Res Cent 2022**; 46(1): 162.

35. Cousins KR. Computer review of ChemDraw ultra 12.0. **J Am Chem Soc 2011**; 133(21). 8388−8388.

36. Li Z, Wan H, Shi Y, Ouyang P. Personal experience with four kinds of chemical structure drawing software: review on ChemDraw, ChemWindow, ISIS/draw, and ChemSketch. **J Chem Inf Comput Sci 2004**; 44(5): 1886−1890.

37. Ibrahim MT, Uzairu A, Shallangwa GA, Uba S. In-silico activity prediction and docking studies of some 2, 9-disubstituted 8-phenylthio/phenylsulfinyl-9h-purine derivatives as Antiproliferative agents. **Heliyon 2020**; 6(1).

38. Abdullahi SH, Uzairu A, Shallangwa GA, Uba S, Umar AB. Molecular docking, ADMET and pharmacokinetic properties predictions of some di-aryl pyridinamine derivatives as estrogen receptor (Er+) kinase inhibitors. **Egypt J Basic Appl Sci 2022**; 9(1): 180−204.

39. Yap CW. PaDEL-descriptor: an open source software to calculate molecular descriptors and fingerprints. **J Comput Chem 2011**; 32(7): 1466−1474.

40. Yamari I, Abchir O, Mali SN, Errougui A, Talbi M, Kouali ME, et al. The anti-SARS-CoV-2 activity of novel 9, 10-dihydrophenanthrene derivatives: an insight into molecular docking, ADMET analysis, and molecular dynamics simulation. **Sci Afr 2023**; 21:e01754.

41. Umar AB, Uzairu A, Shallangwa GA, Uba S. Design of potential anti-melanoma agents against SK-MEL-5 cell line using QSAR modeling and molecular docking methods. **SN Appl Sci 2020**; 2(5): 815.

42. Rajer-Kanduč K, Zupan J, Hrastelj N. Separation of data on the training and test set for modelling: a case study for modelling of five colour properties of a white pigment. **Chemometr Intell Lab Syst 2003**; 65: 221−229.

43. Kennard RW, Stone LA. Computer aided design of experiments. **Technometrics 1969**; 11(1): 137−148.

44. Ibrahim MT, Uzairu A, Shallangwa GA, Uba S. Lead identification of some anti-cancer agents with prominent activity against non-small cell lung cancer (NSCLC) and structure-based design. **Chem Afr 2020**; 3(4): 1023−1044.

45. Abdullahi M, Adeniji SE, Arthur DE, Musa S. Quantitative structure-activity relationship (QSAR) modelling study of some novel carboxamide series as new anti-tubercular agents. **Bull Natl Res Cent 2020**; 44(1): 136.

46. Ibrahim MT, Uzairu A, Uba S, Shallangwa GA. Design of more potent quinazoline derivatives as EGFRWT inhibitors for the treatment of NSCLC: a computational approach. **Fut J Pharm Sci 2021**; 7(1): 140.

47. Ibrahim MT, Uzairu A, Shallangwa GA, Uba S. Computer-aided design of some quinazoline analogues as epidermal growth factor receptor inhibitors. **Egypt J Med Hum Genet 2021**; 22(1): 62.

48. Umar AB, Uzairu A, Shallangwa GA, Uba S. In silico evaluation of some 4-(quinolin-2-yl)pyrimidin-2-amine derivatives as potent V600E-BRAF inhibitors with pharmacokinetics ADMET and drug-likeness predictions. **Fut J Pharm Sci 2020**; 6(1): 61.

49. Veerasamy R, Rajak H, Jain A, Sivadasan S, Christopher PV, Agrawal R. Validation of QSAR models − strategies and importance. **Int J Drug Des Discov 2011**; 2: 511−519.

50. Kasuya E. On the use of r and r squared in correlation and regression. **Ecol Res 2018**; 34.

51. Pratim Roy P, Paul S, Mitra I, Roy K. On two novel parameters for validation of predictive QSAR models. **Molecules 2009**; 14(5): 1660−1701.

52. Ibrahim MT, Uzairu A, Shallangwa GA, Uba S. Structure-based design and activity modeling of novel epidermal growth factor receptor kinase inhibitors; an in silico approach. **Sci Afr 2020**; 9:e00503.

53. Mahmud AW, Shallangwa GA, Uzairu A. QSAR and molecular docking studies of 1,3-dioxoisoindoline-4-aminoquinolines as potent antiplasmodium hybrid compounds. **Heliyon 2020**; 6(3):e03449.

54. Umar AB, Uzairu A, Shallangwa GA, Uba S. Ligand-based drug design and molecular docking simulation studies of some novel anticancer compounds on MALME-3M melanoma cell line. **Egypt J Med Hum Genet 2021**; 22(1): 1−15.

55. McKie JH. Homology modelling of the dihydrofolate reductase-thymidylate synthase bifunctional enzyme of Leishmania major, a potential target for rational drug design in leishmaniasis. **Drug Des Discov 1994**; 11(4): 269−288.

56. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. **BMC Bioinform 2009**; 10(1): 421.

57. Remmert M, Biegert A, Hauser A, Söding J. HHblits: lightning-fast iterative protein sequence searching by HMM-HMM alignment. **Nat Methods 2012**; 9(2): 173−175.

58. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. **Nucleic Acids Res 2018**; 46(W1): W296−W303.

59. Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models. **Bioinformatics 2010**; 27(3): 343−350.

60. Abdullahi SH, Uzairu A, Ibrahim MT, Umar AB. Chemo-informatics activity prediction, ligand based drug design, molecular docking and pharmacokinetics studies of some series of 4, 6-diaryl-2-pyrimidinamine derivatives as anti-cancer agents. **Bull Natl Res Cent 2021**; 45(1): 167.

61. Petit J, Meurice N, Kaiser C, Maggiora G. Softening the rule of five − where to draw the line? **Bioorg Med Chem 2012**; 20(18): 5343−5351.

62. Tandon H, Chakraborty T, Suhag V. A brief review on importance of DFT in drug design. **Res Med Eng Stud 2019**; 39: 46.

63. Umar AB, Uzairu A. Virtual screening, pharmacokinetic, and DFT studies of anticancer compounds as potential V600E-BRAF kinase inhibitors. **J Taibah Univ Med Sci 2023**; 18(5): 933−946.

64. Abdullahi SH, Uzairu A, Shallangwa GA, Uba S, Umar AB. 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 Cent 2022**; 46(1): 2.

65. Hospital A, Goñi JR, Orozco M, Gelpí JL. Molecular dynamics simulations: advances and applications. **Comput Biol Chem Adv Appl 2015**; 8: 37−47.

66. Abdullahi M, Shallangwa GA, Uzairu A. In silico QSAR and molecular docking simulation of some novel aryl sulfonamide derivatives as inhibitors of H5N1 influenza A virus subtype. **Beni-Suef Univ J Basic Appl Sci 2020**; 9(1): 2.

67. Tropsha A. Best practices for QSAR model development, validation, and exploitation. **Mol inf 2010**; 29(6−7): 476−488.

68. Ibrahim MT, Uzairu A, Uba S, Shallangwa GA. Quantitative structure-activity relationship, molecular docking, drug-likeness, and pharmacokinetic studies of some non-small cell lung cancer therapeutic agents. **Beni-Suef Univ J Basic Appl Sci 2020**; 9(1): 49.

69. Umar AB, Uzairu A, Shallangwa GA, Uba S. QSAR modelling and molecular docking studies for anti-cancer compounds against melanoma cell line SK-MEL-2. **Heliyon 2020**; 6(3).

70. Aminu KS, Uzairu A, Abechi SE, Shallangwa GA, Umar AB. Ligand-based drug design, molecular docking and pharmacokinetic studies of some series of 1,4 − dihydropyridines derivatives as human intestinal maltase-glucoamylase inhibitor. **Chem Data Collect 2022**; 41:100911.

71. Adeniji SE, Uba S, Uzairu A. QSAR modeling and molecular docking analysis of some active compounds against *Mycobacterium tuberculosis* receptor (Mtb CYP121). **J Pathog 2018**; 2018:1018694.

72. Golbraikh A. Value of p-Value. **Mol Inf 2019**; 38(8−9): e1800152.

73. Yang L, Wang Y, Chang J, Pan Y, Wei R, Li J, et al. QSAR modeling the toxicity of pesticides against Americamysis bahia. **Chemosphere 2020**; 258:127217.

74. Ibrahim MT, Uzairu A. 2D-QSAR, molecular docking, drug-likeness, and ADMET/pharmacokinetic predictions of some non-small cell lung cancer therapeutic agents. **J Taibah Univ Med Sci 2023**; 18(2): 295−309.

75. Nantasenamat C, Isarankura-Na-Ayudhya C, Prachayasittikul V. Advances in computational methods to predict the biological activity of compounds. **Expet Opin Drug Discov 2010**; 5(7): 633−654.

76. Grisoni F, Consonni V, Todeschini R. Impact of molecular descriptors on computational models. In: Brown JB, editor. *Computational chemogenomics*. New York, NY: Springer New York; 2018. pp. 171−209.

77. Garg R, Smith C. Predicting the bioconcentration factor of highly hydrophobic organic chemicals. **Food Chem Toxicol 2014**; 69.

78. Amalia S, Lucia P. Molecular descriptors and properties of organic molecules. In: Takashiro A, editor. *Symmetry (group theory) and Mathematical treatment in chemistry*. Rijeka: IntechOpen; 2018. Ch. 9.

79. Abchir O, Daoui O, Nour H, Yamari I, Elkhattabi S, Errougui A, et al. Exploration of Cannabis constituents as potential candidates against diabetes mellitus disease using molecular docking, dynamics simulations and ADMET investigations. **Sci Afr 2023**; 21:e01745.

80. Abchir O, Yamari I, Nour H, Daoui O, Elkhattabi S, Errougui A, et al. Structure-based virtual screening, ADMET analysis, and molecular dynamics simulation of Moroccan natural compounds as candidates α-amylase inhibitors. **ChemistrySelect 2023**; 8(26):e202301092.

81. De Vivo M, Masetti M, Bottegoni G, Cavalli A. Role of molecular dynamics and related methods in drug discovery. **J Med Chem 2016**; 59(9): 4035−4061.

82. Abdullahi SH, Uzairu A, Shallangwa GA, Uba S, Umar AB. Pharmacokinetic profiling of quinazoline-4(3H)-one analogs as EGFR inhibitors: 3D-QSAR modeling, molecular docking studies and the design of therapeutic agents. **J Taibah Univ Med Sci 2023**; 18(5): 1018−1029.

83. Abdullahi S, Uzairu A, Shallangwa G, Uba S, Umar A. 2D and 3D-QSAR modeling of 1H-pyrazole derivatives as EGFR inhibitors: molecular docking, and pharmacokinetic profiling. **Chem Afr 2023**.