

RESEARCH

Open Access



# Salicylic acid derivatives as potential $\alpha$ -glucosidase inhibitors: drug design, molecular docking and pharmacokinetic studies

Khalifa Sunusi Aminu<sup>1,2\*</sup>, Adamu Uzairu<sup>1</sup>, Abdullahi Bello Umar<sup>1,2\*</sup> and Muhammad Tukur Ibrahim<sup>1</sup>

## Abstract

**Background:** Diabetes mellitus (DM) is one of the most defying health risk in the twenty-first century promoting a high rate of morbidity and mortality that could possibly increase if no intervention is in place. However, drugs for curing DM are available but are associated with adverse side effect necessitating the pursuit for a safe antidiabetic drugs. The present study was conducted in order to develop a QSAR model that would be used to predict the activities of salicylic acid derivatives, as well as to determine the binding interactions of the compounds with  $\alpha$ -glucosidase using molecular docking studies.

**Results:** Model one was selected and reported as the best model based on its fitness with the following validation keys:  $R^2_{(trng\ set)} = 0.968$ ,  $R^2_{(adj)} = 0.957$ ,  $Q^2_{(cv)} = 0.932$ ,  $LOF = 0.085$  and  $R^2_{(test\ set)} = 0.864$ . Five potent analogues were designed using the ligand-based method with their predicted activities been calculated and found to be higher compared to the lead compound. Furthermore, binding interactions of the designed analogues within the active site of  $\alpha$ -glucosidase (pdb id:3L4V) illustrate a good binding affinities than kotalanol and acarbose. However, the ADMET and drug-likeness properties predicted the design analogues to be pharmacologically and orally safe by not having more than one violation of Lipinski's (Ro5) criteria.

**Conclusions:** The present findings therefore showed that the salicylic acid derivatives could serve as  $\alpha$ -glucosidase inhibitors. The compounds can be studied further for a hunts of promising drug candidates against diabetes mellitus.

**Keywords:** Diabetes mellitus, Quantitative structure activity relationship,  $\alpha$ -glucosidase, Salicylic acid derivatives, Molecular docking

## Background

Diabetes mellitus (DM) is a metabolic disease defined by hyperglycemia as a result of a shortage in insulin action and/or secretion promoting disturbances in metabolic pathways of the body's nutrients. Globally, the disease is considered to be among the most serious threats to humans affecting approximately 173 million people worldwide leading to higher morbidity and mortality rates that could increase severely by 2030 if no

intervention is in place (Wu et al. 2014). A high quantity of blood glucose level can render severe health complications such as neuropathy, retinopathy, microangiopathy and cardiovascular diseases (Chenafa et al. 2021).

The approach utilized for blood glucose control in DM relies on retarding the glucose absorption in the intestinal system by inhibiting  $\alpha$ -glucosidase enzymes which is a key enzyme in catabolism of complex carbohydrates (Chaidam et al. 2021). Considering the importance of the enzyme in DM, a lot of scientific endeavors were put in place to finding promising agents to inhibit the enzyme activity coupled with the fact that the already available drugs are faced with drawbacks (Barber et al. 2021). A study conducted showed natural compounds

\*Correspondence: aminukhalifa1@yahoo.com

<sup>1</sup> Department of Chemistry, Ahmadu Bello University, Zaria, Nigeria  
Full list of author information is available at the end of the article

from *Salacia reticulata* to possessed excellent efficacy against N-terminal human intestinal-maltase glucoamylase (ntMGAM), and therefore, inhibiting the enzyme by the compounds could help in alleviating the negative consequences of DM, since the investigated  $K_i$  values of the salacinol compounds against the enzyme were lower compared to acarbose suggesting their higher potencies than the standard drug (Sim et al. 2010).

Also, another study conducted showed the possibility of salicylic acid derivatives as potential  $\alpha$ -glucosidase inhibitors where a series of salicylic acid derivatives generated from phthalimide exhibited a promising anti- $\alpha$ -glucosidase activities (Chen et al. 2019). These increase our quest towards searching of structural scaffolds from the salicylic acid derivatives against ntMGAM in complex with kotalanol (pdb id: 3L4V) by exploring quantitative structure–activity relationship (QSAR).

QSAR enables the identification of the structural and physicochemical properties of compounds by modulating their activities as well as helping and guiding the design and adoption of effective and reliable drug candidates (Ferreira et al. 2021). To this effect, we herein, aimed to develop a QSAR model from salicylic acid derivatives and use the model to predict the activities of the derivatives. Based on the derivatives, new structural analogues were designed and their activities, alongside molecular docking and their in silico ADMET and drug-likeness properties were evaluated.

## Methods

### Generation of datasets, structure and geometry calculations

A set of thirty-one (31) salicylic acid derivatives (Chen et al. 2019) with their inhibitory concentrations ( $IC_{50}$ ) measured in mM against  $\alpha$ -glucosidase inhibitors were selected for the study. The  $IC_{50}$  of all the compounds were normalized to  $pIC_{50}$  using Eq. 1 (Ibrahim et al. 2020).

$$pIC_{50} = -\log\left(IC_{50} \times 10^{-3}\right) \quad (1)$$

Thereafter, Chemdraw version 12.0 was used to draw the 2-dimensional structures of all the compounds and the resultant structures were transformed automatically to 3-dimensional using the Spartan versions 14 software. Energy minimization and stability geometry of the molecules were carried out to reduce constraint in the structures as well as finding the most stable geometry of the studied molecules using density functional theory (DFT) utilizing Bee -3- Lee Yang Par method (B3LYP) and 6-311G\* level of theory (Abdullahi et al. 2021). The optimized compounds were saved in a Spatial Document file format (sdf).

### Descriptors calculation, data pretreatment and dataset splitting

To gain more information on the compounds under study, the thermodynamic, electronic, autocorrelation and geometric descriptors of each compound were determined by importing the 3D structures in (sdf) file format into the Pharmaceutical Data Exploration Laboratory (PaDEL) descriptor tool kit and the calculation was performed (Lawal et al. 2020). Data pretreatment software version 1.2 was used to pretreat the calculated descriptors of all the compounds under study followed by manual pretreated to eliminate unwanted and persistent descriptors. Following pretreatment, data splitting software was used to split the data into training (modeling) set and test (validation) set by the application of the Kennard–Stone algorithm (Idris et al. 2021). In this study, twenty-one (21) compounds were used as training sets and ten (10) compounds were used as test sets.

### Models building, validation and assessments of the chosen model

Models were developed using material studio version 8.0 software and also applying the genetic function approximation (GFA) which takes into consideration  $pIC_{50}$  as dependent variables and molecular descriptors as the independent variables. The models were scored based on Friedman's lack of fit (Abdullahi et al. 2021). Furthermore, the equation length including both (initial and maximum equation length) was set at 5, mutation probability of 0.1, population and maximum generation parameters were set at 1000 each, and the number of top equations returned was set at 4, respectively. The square correlation coefficient of the training set, adjusted ( $R^2_{adj}$ ), lack of fit (LOF) and cross-validation coefficient ( $Q^2_{cv}$ ) were determined from the generated models.

External validation was conducted on the developed QSAR models using the test sets compounds in which the  $R^2_{test\ value}$  was calculated using Eq. (2) and the best model was chosen. For an external regression equation, the value of  $R^2$  should be closed to one (Abdullahi et al. 2021).

$$R^2_{test} = 1 - \frac{\sum (Y_{exp} - Y_{pred})^2}{\sum (Y_{exp} - Y_{train})^2} \quad (2)$$

where  $Y_{exp}$ ,  $Y_{pred}$  and  $Y_{train}$  are experimental, predicted and average training set activities of the model compounds.

The selected model was assessed by the following numerical measures: mean effect and Y-scrambling test. The mean effect was used to estimate the contribution of each descriptor to the chosen model using Eq. (3).

However, the significance of these descriptors coupled with their mean effect values indicates their influence on the activities of the compounds (Umar et al. 2021).

$$ME_j = \frac{B_j \sum_{i=1}^n d_{ij}}{\sum_j^m B_j \sum_i^n d_{ij}} \quad (3)$$

where  $ME_j$  is the mean effect of descriptor  $j$  in a model,  $B_j$  is the coefficient of the descriptor  $j$  in that model and  $d_{ij}$  is the value of the descriptor in the data matrix for each molecule in the model building set,  $m$  is the number of descriptors that appear in the model and  $n$  is the number of molecules in the model building set (Ibrahim et al. 2020). Furthermore, the Y-scrambling test was used to justify the robustness of the chosen model. The test was performed by evaluating the coefficient of validation parameter for Y-randomization  $cR^2_p$  which is done by rearranging the actual activities and keeping the descriptors unchanged. It was expected that the new QSAR model will have low  $Q^2$  and  $R^2$  values and  $cR^2_p$  must be greater than 0.5 (Idris et al. 2021).

#### Applicability domain

The applicability domain was assessed using William's plot (i.e., plot of standardized residuals versus the leverage values) in order to determine whether the chosen model has a prevailing or peripheral molecules in the specific datasets (Umar et al. 2021). The assessment was carried out by evaluating the leverage approach as well as the warning leverage using Eqs. (4) and (5):

$$h_i = x_i(X^T X)^{-1} X_i^T \quad (4)$$

$$h^* = 3(x + 1)/q \quad (5)$$

where  $h_i$  is the leverage approach,  $X$  is the  $n \times k$  descriptor matrix of the training sets,  $X^T$  is the transpose matrix used in generating the model,  $h^*$  is the warning leverage,  $x$  is the number of descriptors of the selected model and  $q$  is the number of compounds in the training sets (Umar et al. 2021).

#### Ligand-based analogues design

The selection criteria for the choice of a compound to be considered as a lead compound for analogues design were focused solely on the information obtained from the model (i.e., compound with low residual value, reasonable/high  $pIC_{50}$ , found to be within the preferred domain (AD) and do not violate Lipinski's rule of five. However, compound 6 from the salicylic acid derivatives was selected (lead compound) and was modified by insertion and substitution of different groups on XYZ positions

(template compound) based on their notable mean effect value.

#### Molecular docking

A HP Z-Book computer system having the following specifications: An Intel® Core™ i7-4800MQ vPro Dual CPU @2.70 GHz 2.70 GHz, 12 GB of RAM and also utilizing Molegro Virtual Docker (MVD) 2013 version 6.0.1 was used to investigate the binding interactions between the active site of  $\alpha$ -glucosidase and the design analogues. Before docking analysis, analogues were prepared and optimized using Spartan versions 14 software and saved in sdf file format (Ibrahim et al. 2020).  $\alpha$ -Glucosidase with (pdb id: 3L4V) in complex with kotalanol as the co-crystallized ligand was obtained from protein data bank database (<https://www.rcsb.org/>) and MVD was used to prepare the enzyme before proceeding to dock study. During the preparation, water molecules and co-crystal ligands were removed from the crystal structure, the surface was generated and a maximum number of cavities was identified and fixed to 5 for detection of possible binding cavities; execution of docking study was carried out by importing the prepared analogues into the MVD software. MolDock score grid as the scoring function, binding site radius of 15Å, grid resolution of 0.30Å were selected while keeping other parameters as default (Abdullahi et al. 2021). Following a successful docking technique, the MolDock score and hydrogen bond energy were computed. The docked complexes were stored in a PDB file format, and discovery studio version 16.1.0 software was utilized to further visualize the interactions of the docked complexes.

#### ADMET and Drug-likeness of the design analogues

It is viable to assess the absorption, distribution, metabolism, excretion, toxicity and drug-likeness of analogues via silico assays. The drug-likeness and ADMET properties of the design analogues were carried out using SWISSADME (<http://www.swissadme.ch/index.php>) and pkCSM (<http://structure.bioc.cam.ac.uk/pkcsml>) which happens to be a certified free web tool for predicting and evaluating the ADMET and drug-likeness of small compounds (Kar and Chatterjee 2020). However, one of the main aspects in the preclinical phase of analogues design is to know whether if those analogues are orally bioavailable, in this regard the Lipinski's rule of five criteria was used to access the analogues under study which states that a chemical is weakly absorbed if it disobeys more than two of these conditions (molecular weight  $\leq 500$ , number of hydrogen bond donors  $\leq 5$ , number of hydrogen bond acceptors  $\leq 10$  and calculated Log  $p \leq 5$ ) (Chen et al. 2020) which were applied to the design analogues to predict their drug-likeness and ADMET properties.

## Results

### QSAR analysis

The experimental activities of the datasets were normalized to  $pIC_{50}$ . Also, the predicted  $pIC_{50}(s)$  of the compounds was determined (Table 1). The differences between the experimental and predicted activities of the compounds (residual values) were further calculated (Table 1).

Subsequently, models were generated using the genetic function approximation. Four models were generated and shown thus;

$$\text{Model 1 : } pIC_{50} = 1.45 * (ATSC2e) - 14.87 * (MATS1e) + 2.23 * (GATS2c) - 3.04 * (SpMin6\_Bhm) + 0.71 * (CrippenLogP) - 0.51$$

$$\text{Model 2 : } pIC_{50} = - 2.04 * (ATSC1e) + 0.03 * (ATSC2s) - 0.06 * (VR2\_Dzs) - 5.40 * (SpMin8\_Bhp) + 0.93 * (CrippenLogP) + 4.56$$

$$\text{Model 3 : } pIC_{50} = 1.50 * (ATSC2e) - 0.16 * (AATSC2m) - 3.79 * (AATSC6s) - 5.57 * (MATS1e) + 7.26 * (MATS8s) + 0.07$$

$$\text{Model 4 : } pIC_{50} = 0.02 * (AATS8v) - 1.21 * (ATSC1e) + 1.62 * (ATSC2e) - 2.71 * (AATSC6s) + 3.69 * (MATS8s) - 2.77$$

Following models building, validation was conducted (Table 2). The  $R^2_{\text{trng set}}$ ,  $R^2_{\text{adj}}$ ,  $Q^2_{\text{cv}}$ , as the internal validation parameters of the models were greater than 0.920 (closer to 1) while only model one passed the external validation with  $R^2_{\text{test set}}$  of 0.864 (Table 2). Furthermore, an XY (scatter) plot of residuals against experimental  $pIC_{50}$  from both the training and test sets compounds is shown in Fig. 1. The anomalous occurrence of these residuals on both side of zero on the plot indicates the absence of feasible mistakes in the model building (Fig. 1).

Based on the initial validations, model one was selected as the best model due to the fact that it has the highest  $R^2_{\text{test set}}$  of 0.864 and was subjected for further assessment. From the *ME* analysis, positive *ME* values were observed in ATSC2e, MATS1e, GATS2c and Crippen-LogP with CrippenLogP descriptor having the highest *ME* value. Conversely, negative *ME* value of -1.01 with SpMin6-Bhm descriptors was noticeable (Table 3).

Moreover, assessment on the robustness and obtainability of the model whether by chance correlation or not was conducted using the Y-randomization test (Table 4). From the result obtained, the  $R^2$  and  $Q^2$  values were 0.19 and -0.57, respectively, while the  $cR^2p$  of the model was 0.79 (Table 4).

Williams plot was used in order to find compounds that undesirably stimulate the performance of the model within the domain of the model (Fig. 2). From the plot, one compound from the test set was found to be outside the favored domain, i.e., warning leverage ( $h^* > 0.86$ ) (Fig. 2). The compound was identified as compound 28.

After establishing the applicability domain, ligands were designed using the ligand-based approach coupled with information obtained from the model, compound 6 was tag as the lead compound (Fig. 3A), and template (Fig. 3B) was adopted. The design was performed by

addition of different groups on the X, Y and Z position on the template with model one utilized in predicting their activities.

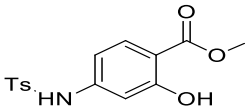
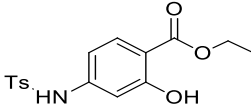
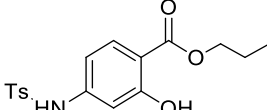
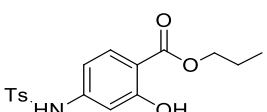
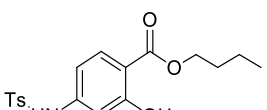
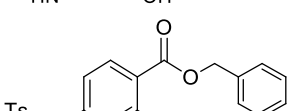
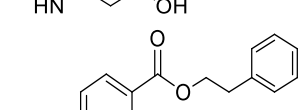
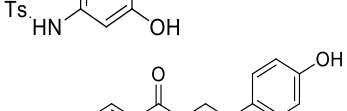
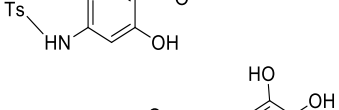
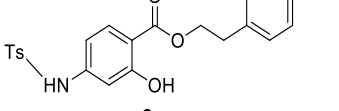
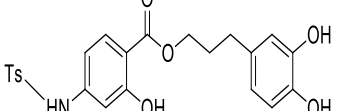
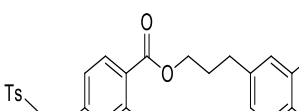
Among the designed analogues, five had predicted  $pIC_{50}$  relatively higher than the lead compound (Table 5). Interestingly, the analogues showed values between 4.32 and 5.11 with analogue A4 having the highest value of 5.11 (Table 5).

Molecular docking studies conducted between the designed analogues and  $\alpha$ -glucosidase (pdb id:3L4V) showed an interesting finding. The graphical structure of the protein was shown, and the active site of the enzyme was encircled (Fig. 4).

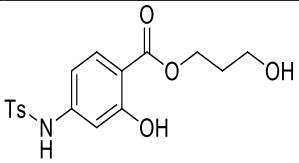
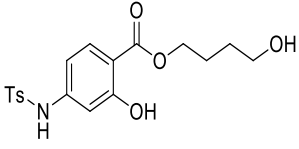
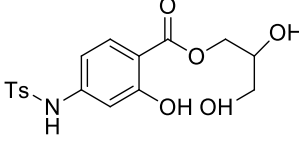
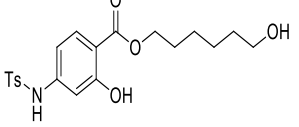
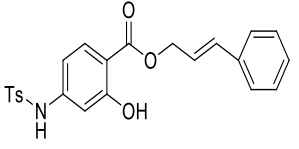
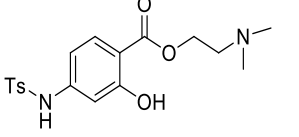
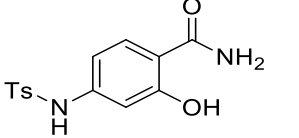
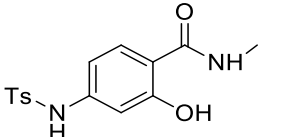
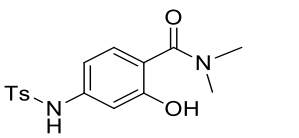
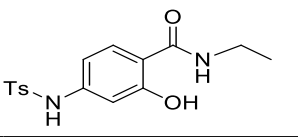
The MolDock scores of the analogues were in the order Analogue A2 < Analogue A3 < Analogue A1 < Analogue A5 < Analogue A4 (having a value of -196.1). However, it is important to note that the co-crystallized ligand (kotalanol) and acarbose have MolDock scores of -148.4 and -146.6, respectively (Table 6).

Visualization of the docking results was performed using discovery studio visualizer. Notably, Analogue A1 interacted with Ser288, Ile523, Lys776, Phe535 and Leu286 residues of the protein through conventional hydrogen bond interactions, respectively (Table 6, Fig. 5A and a). Also, Analogue A2 was bound to the protein via

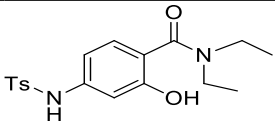
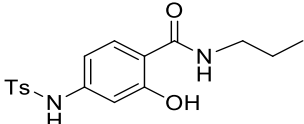
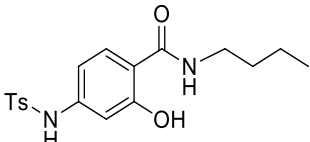
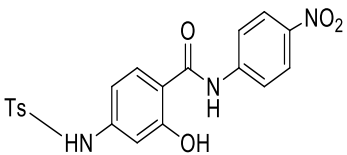
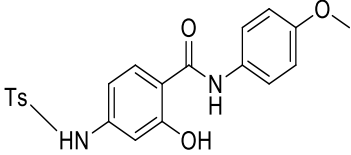
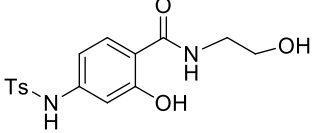
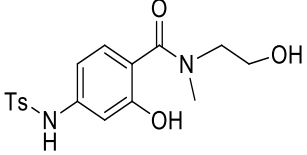
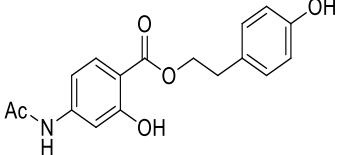
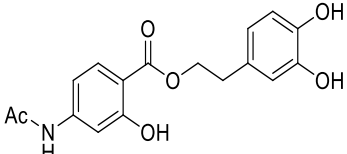
**Table 1** The structures, experimental  $pIC_{50}$ , predicted  $pIC_{50}$  and the residual values of salicylic acid derivatives against  $\alpha$ -glucosidase

Compounds	Structures	Experimental $pIC_{50}$	Predicted $pIC_{50}$	Residuals
1 <sup>X</sup>		2.76	2.53	0.23
2 <sup>Y</sup>		3.00	3.01	-0.01
3 <sup>Y</sup>		3.29	3.25	0.04
4 <sup>Y</sup>		3.05	3.21	-0.15
5 <sup>Y</sup>		3.46	3.41	0.05
6 <sup>Y</sup>		3.33	3.32	0.01
7 <sup>Y</sup>		3.36	3.54	-0.19
8 <sup>X</sup>		3.70	3.67	0.03
9 <sup>Y</sup>		3.82	3.87	-0.04
10 <sup>Y</sup>		4.07	4.02	0.05
11 <sup>Y</sup>		3.02	3.06	-0.04
12 <sup>X</sup>		2.40	2.34	0.05

**Table 1** (continued)

Compounds	Structures	Experimental $pIC_{50}$	Predicted $pIC_{50}$	Residuals
13 <sup>x</sup>		2.52	2.56	-0.04
14 <sup>x</sup>		2.78	2.68	0.10
15 <sup>y</sup>		1.72	1.99	-0.27
16 <sup>y</sup>		3.10	3.00	0.10
17 <sup>y</sup>		3.72	3.63	0.10
18 <sup>y</sup>		2.24	2.12	0.12
19 <sup>y</sup>		2.60	2.62	-0.02
20 <sup>x</sup>		2.65	2.30	0.35
21 <sup>x</sup>		2.22	1.94	0.28
22 <sup>y</sup>		2.76	2.60	0.15

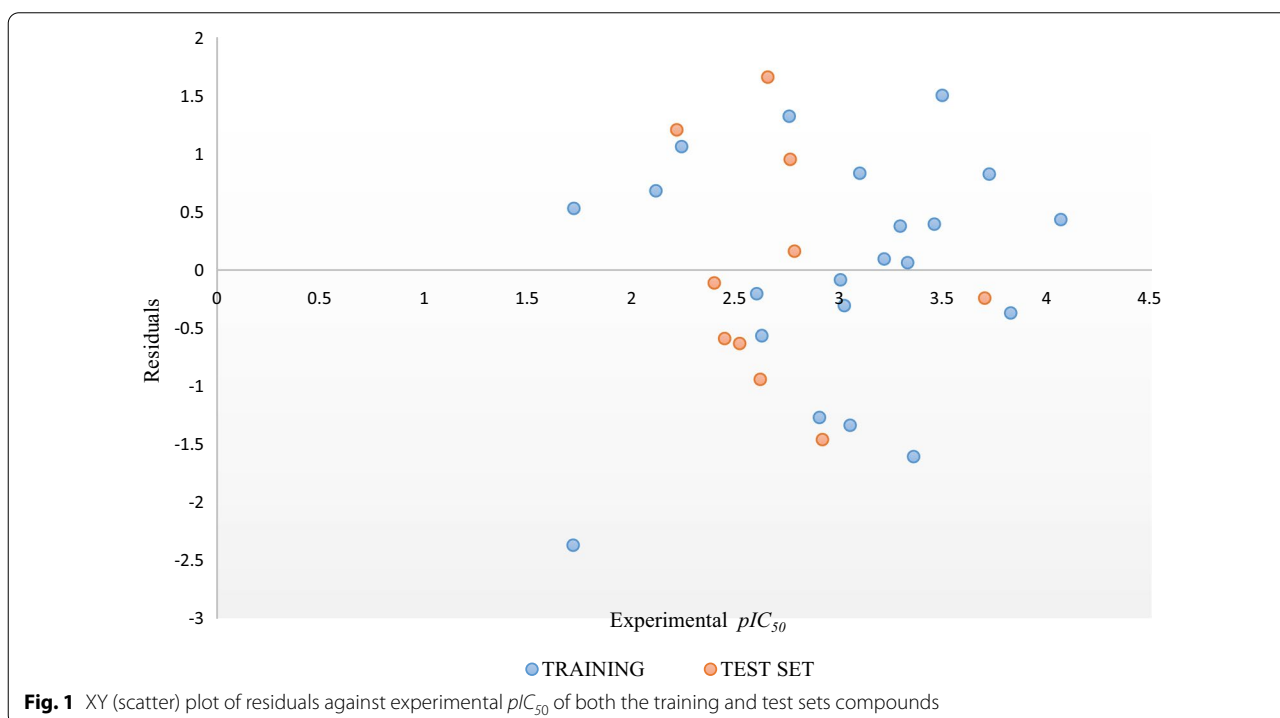
**Table 1** (continued)

Compounds	Structures	Experimental $pIC_{50}$	Predicted $pIC_{50}$	Residuals
23 <sup>x</sup>		2.45	2.48	-0.03
24 <sup>x</sup>		2.62	2.71	-0.09
25 <sup>y</sup>		2.63	2.70	-0.07
26 <sup>y</sup>		3.22	3.20	0.01
27 <sup>x</sup>		2.92	3.09	-0.18
28 <sup>y</sup>		2.12	2.04	0.08
29 <sup>y</sup>		1.72	1.66	0.06
30 <sup>y</sup>		2.90	3.05	-0.15
31 <sup>y</sup>		3.50	3.32	0.17
Reference drug (Acarbose)	6.35	-	-	

<sup>y</sup> Training set<sup>x</sup> Test set

**Table 2** Validation parameters of the models generated with the recommended values for assessing a valid QSAR model

Models	Parameters					
	$R^2_{\text{trng set}}$	$R^2_{\text{adj}}$	$Q^2_{\text{cv}}$	$R^2_{\text{trng set}} - Q^2_{\text{cv}}$	LOF	$R^2_{\text{test set}}$
Model 1	0.968	0.957	0.932	0.036	0.085	0.864
Model 2	0.966	0.955	0.931	0.035	0.087	0.116
Model 3	0.966	0.955	0.931	0.035	0.088	0.451
Model 4	0.965	0.954	0.936	0.029	0.091	0.275
Recommended values	Close to unity	Close to unity	>0.5	≤0.3	–	≥0.6

**Table 3** Pearson's correlation and mean effect values of the descriptors for model one

Symbol	<i>ATSC2e</i>	<i>MATS1e</i>	<i>GATS2c</i>	<i>SpMin6_Bhm</i>	<i>CrippenLogP</i>	ME
<i>ATSC2e</i>	1					0.71
<i>MATS1e</i>	0.85	1				0.11
<i>GATS2c</i>	−0.05	0.39	1			0.47
<i>SpMin6_Bhm</i>	0.19	0.34	0.56	1		−1.01
<i>CrippenLogP</i>	0.67	0.62	0.21	0.59	1	0.72

four conventional H-bond with Ser288, Ile523, Leu286 and Arg520, respectively (Table 6, Fig. 5B and b). Two conventional H-bond interaction was observed in Analogue A3 with Arg283 and Leu286, respectively (Table 6, Fig. 5C and c). Moreover, compound A4 was bound to

the protein via Ser288, Lys776, Met567, Gly533, Leu286 and Phe535 conventional H-bond (Table 6, Fig. 5D and d). Also, Analogue A5 interacted with the protein via five conventional H-bond with Pro287, Ser288, Lys776, Arg283 and Pro284, respectively (Table 6, Fig. 5E and



**Table 4** Y-randomization test result

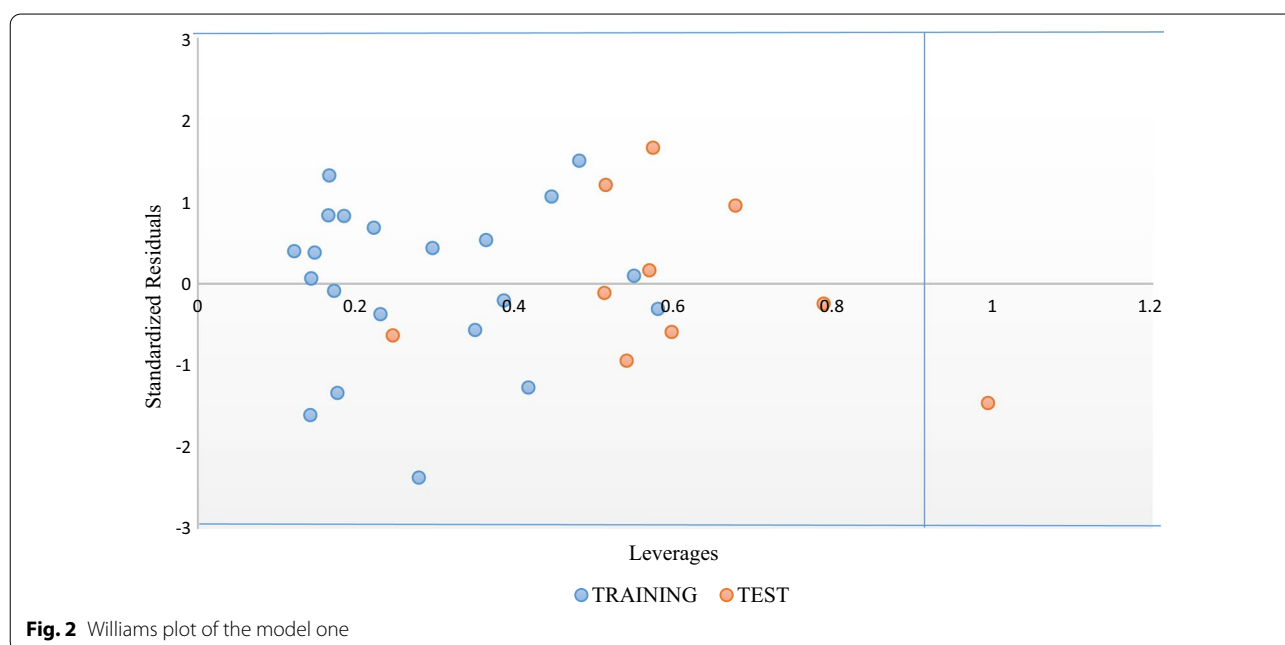
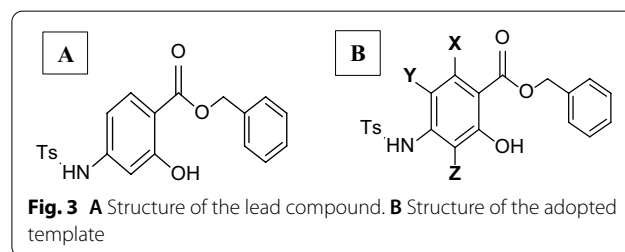
Iterations	R	R <sup>2</sup>	Q <sup>2</sup>
Original	0.93	0.87	0.74
Random 1	0.75	0.56	0.02
Random 2	0.55	0.30	-0.25
Random 3	0.09	0.01	-1.04
Random 4	0.35	0.12	-0.66
Random 5	0.51	0.26	-0.44
Random 6	0.29	0.08	-0.73
Random 7	0.19	0.04	-0.88
Random 8	0.52	0.27	-0.40
Random 9	0.47	0.22	-0.37
Random 10	0.22	0.05	-0.90
<i>Random models parameters</i>			
Average R:	0.39		
Average R <sup>2</sup> :	0.19		
Average Q <sup>2</sup> :	-0.57		
cRp <sup>2</sup> :	0.79		

e). However, the kotalanol and acarbose was bound to the protein through six and eight conventional H-bond interactions, respectively (Table 6, Fig. 5F and f, G and g). Noticeably, other interactions such as carbon H-bond, alkyl, pi-alkyl, pi-sulfur, pi-cation, pi-sigma and pi-donor H-bond were observed (Table 6).

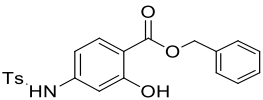
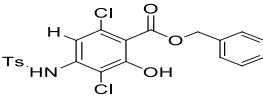
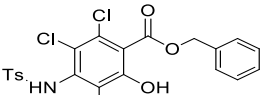
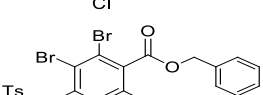
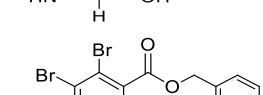
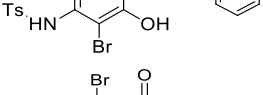
The ADMET properties of the designed analogues (Table 7) showed an excellent intestinal absorption ranging between 91.087 and 92.961%, The BBB permeability and central nervous system permeability were also

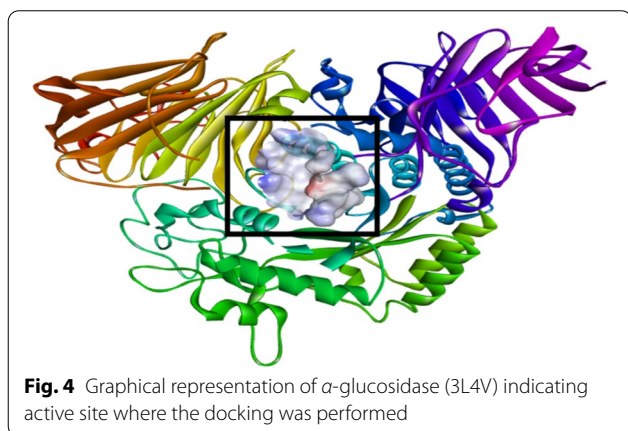
found to be within the recommended value which is  $\log BB < -1$  and  $\log PS < -3$ , respectively. The cytochrome P450 (CYP) superfamily which are responsible for so many drug interactions were examined with the designed analogues serving as substrate of 3A4 and inhibitors of 2C19, 2C9 and 3A4, respectively. Moreover, the designed analogues are found to be nontoxic with good total clearance potential within the range of  $-0.367$  to  $0.007$  (Table 7).

Subsequently, the drug-likeness properties depicted in (Table 8) of the design analogues showed that only analogue A1 was found to respect Lipinski's (Ro5) without violating any of the criteria set by the rule while analogues A2, A3, A4 and A5 were found to have one violation (i.e., their molecular weight exceeds 500). The bioavailability score was also observed to be 0.55 except for analogue A4 which is 0.17. The synthetic accessibility having a recommended range from 1 (very easy to synthesize) to 10 (very hard to synthesize) was also analyzed with the designed analogues found to be within the range of 3.14 to 3.43 (Table 8).



**Table 5** Structures, predicted  $pIC_{50}$  of the lead compound and the designed analogues

S/No	Structures	Predicted $pIC_{50}$
Lead compound		3.32
A1		4.32
A2		4.80
A3		4.48
A4		5.11
A5		4.50



## Discussions

Globally, the increased occurrence of DM is disturbing (Wu et al. 2014). The adverse side effects of the synthetic antidiabetic agents hassle the search for new therapeutic candidates (Banerjee et al. 2020). In this study, the fascinating in vitro activities of salicylic acid derivatives against  $\alpha$ -glucosidase (Chen et al. 2019) led to the structural activity relationship analysis of the derivatives. Most importantly, designing

potent analogues from the derivatives could serve as  $\alpha$ -glucosidase inhibitors.

The QSAR models were generated using genetic function approximation. This is because the GFA delivers more than one model which gives a degree of freedom to select the best (Abdullahi et al. 2021). Subsequently, the generated models were validated in order to ascertain their fitness and predictive power to improve the compound's biological activities (Umar et al. 2021). Model one was selected as the best based on the  $R^2_{\text{test set}}$  value of 0.864 which shows that the model was not over-fitted and sensible predictive activities for designed analogues is promising compared to other models.

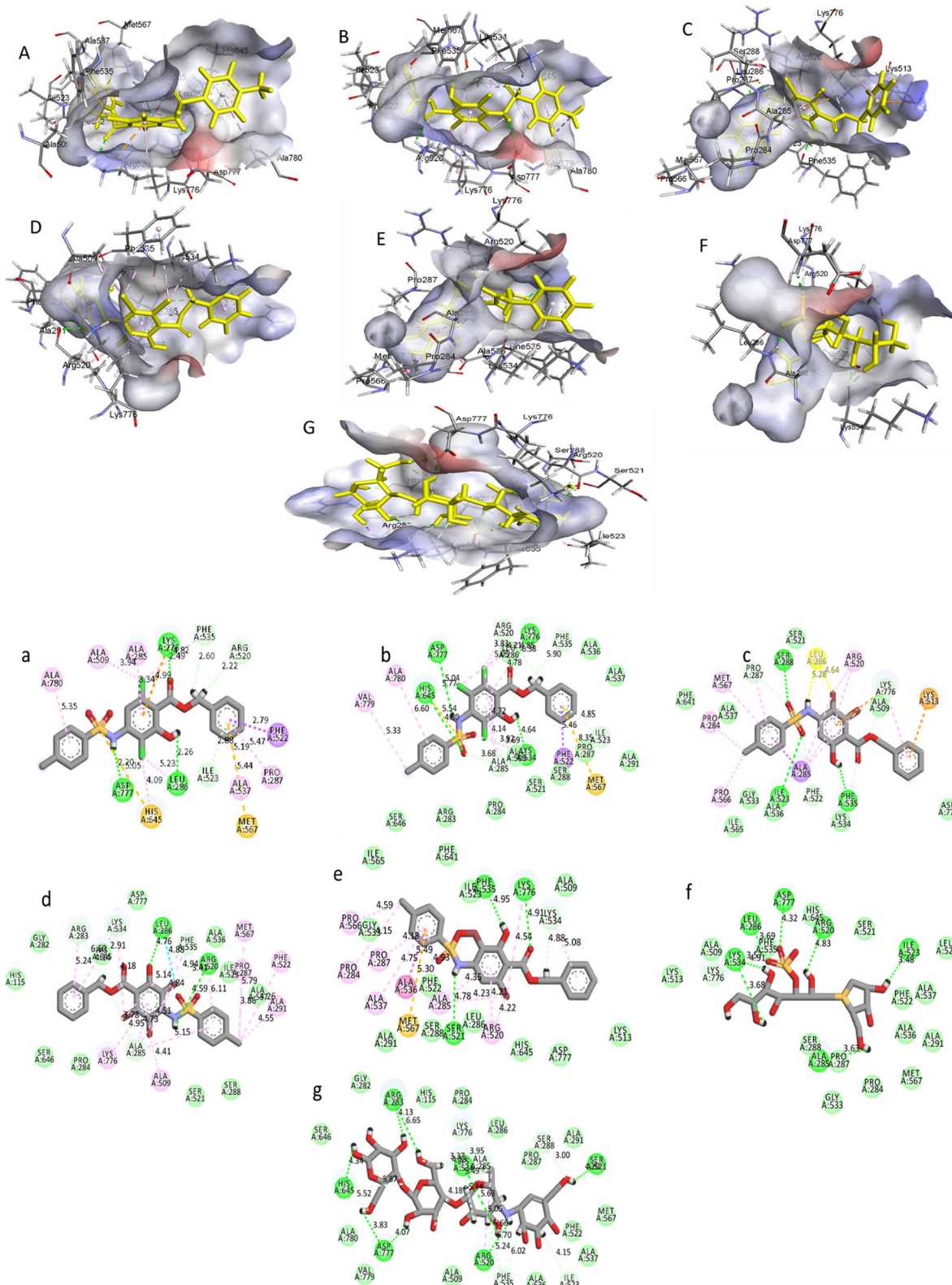
The Pearson's correlation shows the correlation of the descriptors in the model, and the mean effect values  $ME$  of the molecular descriptors represent the physicochemical properties that provides a structural information of each descriptor but in a numerical form, i.e., a unique information of each descriptor that can be employed to improve the activities of the compounds (Umar et al. 2021). However, positive coefficient descriptors  $ATSC2e$ ,  $MATS1e$ ,  $GATS2c$  and  $CrippenLogP$  as seen from the result indicate a positive influence of these descriptors to the activities of salicylic acid derivatives, i.e., the greater the value of these types of descriptors the greater the antidiabetic activity and vice versa. Consequently, the descriptor with a negative coefficient  $SpMin6-Bhm$  indicates a negative influence to the activities of salicylic acid derivatives, i.e., the lower the value of this type of descriptor, the greater the antidiabetic activities. Thus, this shows the role of steric and electrostatic interactions in inducing the activities of salicylic acid derivatives against  $\alpha$ -glucosidase (Umar et al. 2021). Also, a  $cR^2p$  of 0.79 obtained indicates the robustness of model one and obtainability was not based on probability. This is because the recommended value for the parameter is greater than 0.5 (Idris et al. 2021).

Moreover, the salicylic acid derivatives were exposed to Williams plot in order to inspect for influential compounds that might undesirably affect the model performance (Ibrahim et al. 2020). Among the derivatives, compound 28 was considered as an outlier. The change in substituent and position of attachment could be the cause for the compound to be found outside the preferred domain.

Among the derivatives within the defined domain, compound 6 with low residual value, good  $pIC_{50}$  value was tag as lead compound and positions X, Y and Z were added as point of attachments as indicated in the adopted template. The choice of the substituent to be added was based on  $ATSC2e$  and  $CrippenLogP$  descriptors as it has earlier reported to have highest  $ME$  values. Based on the descriptors, five designed analogues were found to

**Table 6** MolDock score, H-bond energies and amino acid residues of the design analogues, kotalanol and acarbose

Ligands	MolDock score (kJ/mol)	H-bond energies (kJ/mol)	Conventional H-bond	Carbon-H bond	Pi-sulfur	Alkyl	Pi-alkyl	Pi-cation	Pi-sigma	Attractive charge	Pi-donor H-bond
A1	-187.2	-5.05	Ser288 Ile523 Lys776 Phe535 Leu286	Phe522, Lys534	Met567	Pro284, Pro566	Ala285, Arg520, Lys534, Pro287 Ala537	-	-	-	-
A2	-164.4	-6.20	Ser288 Ile523 Leu286 Arg520	Pro28 Phe522 Lys776 Lys534	Arg283 Pro284, Pro566	His645 Ala285 Arg520 Lys776, Ala285, Pro287 Met567	Lys776	-	-	-	-
A3	-166.9	-9.87	Arg283 Leu286	Phe522, Lys776, Lys534	Pro284, Pro566	Ala285 Arg520 Lys776 Pro287	-	Ala285	-	-	-
A4	-196.1	-10.67	Ser288 Lys776 Met567 Gly533 Leu286 Phe535	Ser288, Ile565, Pro566	Met567	Leu286, Arg520	His645 Ala285 Arg520 Pro287 Ala537 Pro566 Lys776	-	-	-	-
A5	-192.4	-7.55	Pro287 Ser288 Lys776 Arg283 Pro284	Pro287, Phe522	Pro284, Pro566	Ala285 Arg520 Ile523, Lys534 Pro287 Met567	-	-	-	-	-
Kotalanol	-148.4	-12.24	Ala285 Arg520 Ile523 Met567 Thr775 Lys776	Arg520 Ala536 Phe535 His645 Asp777 Lys776 Leu286	-	-	-	-	Lys776	-	-
Acarbose	-146.6	-13.42	Arg283 Arg520 Ile523 Lys534 Ser521 Asp777 Lys534 His645	Arg283, Ile534, Lys534 Phe535, Ser288	-	Ala285 Arg520	-	-	-	-	His645



**Fig. 5** 3D and 2D Interactions of Analogue A1 (A and a), Analogue A2 (B and b), Analogue A3 (C and c), Analogue A4 (D and d), Analogue A5 (E and e), kotalanol (F and f) and acarbose (G and g) with  $\alpha$ -glucosidase (3L4V)

**Table 7** ADMET properties of the designed analogues

	Absorption	Distribution	CNS	Metabolism						Excretion	Toxicity	
	Intestinal absorption (Human)	BBB perm. (Log BB)	Perm. (Log PS)	Substrate and inhibitors CYP						Total clearance	AMES toxicity	
				Substrate of CYP		Inhibitors of CYP						
				2D6	3A4	1A2	2C19	2C9	2D6			3A4
A1	92.552	-0.756	-2.072	No	Yes	No	Yes	Yes	No	Yes	0.007	No
A2	93.752	-0.896	-1.949	No	Yes	No	Yes	Yes	No	Yes	0.110	No
A3	92.024	-0.756	-2.026	No	Yes	No	Yes	Yes	No	Yes	-0.229	No
A4	92.961	-0.921	-1.881	No	Yes	No	Yes	Yes	No	Yes	-0.367	No
A5	92.024	-0.773	-2.026	No	Yes	No	Yes	Yes	No	Yes	-0.347	No

**Table 8** Drug-likeness properties of the designed analogues

	Molecular weight	H-bond acceptor	H-bond donor	MLog P	Lipinski's violations	Bioavailability score	Synthetic accessibility
A1	466.33	5	2	3.66	0	0.55	3.33
A2	500.78	5	2	4.13	1	0.55	3.38
A3	555.24	5	2	3.87	1	0.55	3.35
A4	634.13	5	2	4.45	1	0.17	3.43
A5	555.24	5	2	3.87	1	0.55	3.37

have a higher activities relative to the lead compound. Electron-withdrawing group particularly bromine group have shown to significantly enhance the activities of the designed analogues. However, similar substituents were shown to be used in recent studies conducted which were found to be promising in enhancing the activities of compounds (Abdullahi et al. 2021).

The docking studies of the five designed analogues against  $\alpha$ -glucosidase (pdb id:3L4V) were conducted. MolDock score is the scoring function which gives information on the binding energy of interaction. It was adopted for evaluating the interactions of the designed analogues against the active site of the  $\alpha$ -glucosidase. The result of the docking studies show that they were bind at the active site of the receptor with a good MolDock score compared to kotalanol and acarbose. Observably, the higher number of H-bond energies in acarbose and kotalanol was as a result of higher number of hydroxyl functional group which portrays a greater number of hydrogen bond interactions at the active site of the receptor after docking. Most of the interactions occur with the amino acid residues found within the active sites (Maurya et al. 2020).

The ADMET results showed that the analogues had passed the adsorption, distribution, metabolism, excretion and toxicity parameters with good bioavailability

score and orally safe. Most at times determines the success of a drug candidate (Zafar et al. 2020).

## Conclusions

The present findings therefore showed that model one generated from genetic function approximation was the best due to its fitness with the following internal and external validation parameters of  $R^2_{(\text{trng set})} = 0.968$ ,  $R^2_{(\text{adj})} = 0.957$ ,  $Q^2_{\text{cv}} = 0.932$ ,  $\text{LOF} = 0.085$  and  $R^2_{(\text{test set})} = 0.864$ . However, the five potent analogues designed using the ligand-based were found to have higher activities than the lead compound. The compounds interacted reasonably well with the active site of the  $\alpha$ -glucosidase. Based on the present findings, the compounds could be used subsequently for the search of new antidiabetic agent.

## Abbreviations

QSAR: Quantitative structure–activity relationship; B3LYP: Bee–3–Lee Yang Par; GFA: Genetic function approximation; DFT: Density functional theory; ME: Mean effect; MVD: Molegro Virtual Docker; PaDEL: Pharmaceutical Data Exploration Laboratory; ADMET: Absorption, distribution, metabolism, excretion and toxicity.

## Acknowledgements

The author sincerely acknowledged all the contributors for their supervision and motivation during the research work and also acknowledged Ahmadu Bello University for providing the softwares used for the research.

**Author contributions**

AU and KSA conceptualized the research idea; KSA performed all the computational analysis and wrote the manuscript; AU, KSA, ABU, MTI analyze the result to ensure minimal errors, and AU supervised and proof read the manuscript. All authors read and approved the manuscript before final submission.

**Funding**

No funding was received by the authors for this research.

**Availability of data and materials**

All data generated or analyzed during this study are included in this published article.

**Declarations****Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Author details**

<sup>1</sup>Department of Chemistry, Ahmadu Bello University, Zaria, Nigeria. <sup>2</sup>Department of Pure and Industrial Chemistry, Bayero University, Kano, Nigeria.

Received: 12 April 2022 Accepted: 26 May 2022

Published online: 04 June 2022

**References**

- Abdullahi SH, Uzairu A, Ibrahim MT, Umar AB (2021) 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.* <https://doi.org/10.1186/s42269-021-00631-w>
- Banerjee M, Khursheed R, Yadav AK, Singh SK, Gulati M, Pandey DK, Prabhakar PK, Kumar R, Porwal O, Awasthi A, Kumari Y, Kaur G, Ayinkamiye C, Prashar R, Mankotia D, Pandey NK (2020) A systematic review on synthetic drugs and phytopharmaceuticals used to manage diabetes. *Curr Diabetes Rev* 16(4):340–356. <https://doi.org/10.2174/1573399815666190822165141>
- Barber E, Houghton MJ, Williamson G (2021) Flavonoids as human intestinal  $\alpha$ -glucosidase inhibitors. *Foods.* <https://doi.org/10.3390/foods10081939>
- Chaidam S, Saehlim N, Athipornchai A, Sirion U, Saeeng R (2021) Synthesis and biological evaluation of 1,6-bis-triazole-2,3,4-tri-O-benzyl- $\alpha$ -D-glucopyranosides as a novel  $\alpha$ -glucosidase inhibitor in the treatment of Type 2 diabetes. *Bioorganic Med Chem Lett* 50:128331. <https://doi.org/10.1016/j.bmcl.2021.128331>
- Chen J, Lu W, Chen H, Bian X, Yang G (2019) A new series of salicylic acid derivatives as non-saccharide  $\alpha$ -glucosidase inhibitors and antioxidants. *Biol Pharm Bull* 42(2):231–246. <https://doi.org/10.1248/bpb/b18-00661>
- Chen X, Li H, Tian L, Li Q, Luo J, Zhang Y (2020) Analysis of the physicochemical properties of acaricides based on Lipinski's rule of five. *Comput Biol* 27(9):1397–1406. <https://doi.org/10.1089/cmb.2019.0323>
- Chenafa H, Mesli F, Daoud I, Achiri R, Ghalem S, Neghra A (2021) In silico design of enzyme  $\alpha$ -amylase and  $\alpha$ -glucosidase inhibitors using molecular docking, molecular dynamic, conceptual DFT investigation and pharmacophore modelling. *J Biomol Struct Dyn.* <https://doi.org/10.1080/07391102.2021.18>
- Ferreira LT, Borba JVB, Moreira-Filho JT, Rimoldi A, Andrade CH, Costa FTM (2021) Qsar-based virtual screening of natural products database for identification of potent antimalarial hits. *Biomolecules* 11(3):1–12. <https://doi.org/10.3390/biom11030459>
- Ibrahim MT, Uzairu A, Shallangwa GA, Uba S (2020) In-silico activity prediction and docking studies of some 2, 9-disubstituted 8-phenylthio/phenylsulfanyl-9h-purine derivatives as anti-proliferative agents. *Heliyon* 6(1):e03158
- Idris MO, Abechi SE, Shallangwa GA (2021) Computational evaluation of some compounds as potential anti-breast cancer agents. *Future J Pharm Sci.* <https://doi.org/10.1186/s43094-021-00315-2>
- Kar S, Chatterjee S (2020) In silico meets in vitro techniques in ADMET profiling of drug discovery (Part I). *Curr Drug Metab* 21(10):745–745. <https://doi.org/10.2174/138920022110201126153901>
- Lawal HA, Uzairu A, Uba S (2020) QSAR, molecular docking, design, and pharmacokinetic analysis of 2-(4-fluorophenyl) imidazol-5-ones as anti-breast cancer drug compounds against MCF-7 cell line. *J Bioenerg Biomembr* 52(6):475–494. <https://doi.org/10.1007/s10863-020-09858>
- Maurya AK, Mulpuru V, Mishra N (2020) Discovery of novel coumarin analogs against the  $\alpha$ -glucosidase protein target of diabetes mellitus: pharmacophore-based QSAR, docking, and molecular dynamics simulation studies. *ACS Omega* 5(50):32234–32249. <https://doi.org/10.1021/acscomega.0c03871>
- Sim L, Jayakanthan K, Mohan S, Nasi R, Johnston BD, Mario Pinto B, Rose DR (2010) New glucosidase inhibitors from an ayurvedic herbal treatment for type 2 diabetes: structures and inhibition of human intestinal maltase-glucoamylase with compounds from *Salacia reticulata*. *Biochem* 49(3):443–451. <https://doi.org/10.1021/bi9016457>
- Umar AB, Uzairu A, Shallangwa GA, Uba S (2021) 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* 22(1):1–15. <https://doi.org/10.1186/s43042-020-00126>
- Wu Y, Ding Y, Tanaka Y, Zhang W (2014) Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention. *Int J Med Sci* 11(11):1185–1200. <https://doi.org/10.7150/ijms.10001>
- Zafar F, Gupta A, Thangavel K, Khatana K, Sani AA, Ghosal A, Tandon P, Nishat N (2020) Physicochemical and pharmacokinetic analysis of anacardic acid derivatives. *ACS Omega* 5(11):6021–6030. <https://doi.org/10.1021/acscomega.9b04398>

**Publisher's Note**

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

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)