

A comparative QSAR analysis, 3D-QSAR, molecular docking and molecular design of iminoguanidine-based inhibitors of HemO: A rational approach to antibacterial drug design

Emmanuel Israel Edache^{1,2*}, Adamu Uzairu², Paul Andrew Mamza² and Gideon Adamu Shallangwa²

¹Department of Pure and Applied Chemistry, University of Maiduguri, Brono State, Nigeria.

²Department of Chemistry, Ahmadu Bello University, P.M.B. 1044, Zaria, Nigeria.

*Corresponding author. Email: edacheson2004@gmail.com; Tel: +2348066776802.

Copyright © 2020 Edache et al. This article remains permanently open access under the terms of the [Creative Commons Attribution License 4.0](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Received 4th June, 2020; Accepted 20th July, 2020

ABSTRACT: QSAR models of 25 iminoguanidine derivatives with inhibitory HemO were developed. The QSAR model was built by using DFT-MLR and the best QSAR model has R^2 , Q^2 values of 0.6569 and 0.5493 for cross-validated and non-cross-validated. The predictive ability of QSAR model was further validated by a test set of 7 compounds, giving R^2_{pred} value of 0.7123. 3D-QSAR and docking studies were used to find the actual conformations of chemicals in active site of HemO, as well as the binding mode pattern to the binding site in HemO enzyme. Molecular dynamics and simulations study revealed that A-chain of HemO protein was stable at and above 100ps with respect to temperature (at and above 298 K), electrostatic (at and above 57500 kJ/mol), kinetic energy (at and above 12200 kJ/mol) and total energy (at and above 30500 kJ/mol). The information provided by QSAR, 3D-QSAR and molecular docking may provide a better understanding of the structural requirements of iminoguanidine-based inhibitors of HemO and help to design potential antibacterial molecules. Four (4) new compounds (4A, 5A, 6A, and 7A) with high predicted antibacterial activities have been theoretically designed and they are expected to be confirmed experimentally.

Keywords: CoMFA, docking, heme oxygenase (HemO), iminoguanidine derivatives, MD simulations, molecular design, *Pseudomonas aeruginosa*, QSAR.

INTRODUCTION

Pseudomonas aeruginosa infections are a main cause of mortality and morbidity of cystic fibrosis (CF) patients. *P. aeruginosa* bets on acquiring iron from its host, and various iron acquisition systems may be active during infection (Marvig et al., 2014). Heme (called protoporphyrin IX) which composed of iron and organic part, serves as a prosthetic group of many proteins. Heme enables a wide form of processes like electron transport, enzyme catalysis and reversible binding of gases. Heme also involved in regulation of protein synthesis and cells differentiation (Ponka, 1999; Podkalicka et al., 2018). Heme released from intracellular heme proteins causes cellular and tissue damage. Toxic and pro-oxidant properties of heme are the costs of iron-dependent reactive oxygen species generation, lipids peroxidation

and disruption of cellular membranes. As the result, an excessive level of circulating heme leads to several pathologies (Kumar and Bandyopadhyay, 2005; Ochsner et al., 2000). Heme is a critical source of iron for *P. aeruginosa* on infection of the host. The opportunistic infectious agent, *P. aeruginosa* converts two distinct heme uptake systems: (1) *Pseudomonas* heme uptake (phu) and (2) heme assimilation system (has) (Ochsner et al., 2000). Phu system is the primary iron acquisition system in *P. aeruginosa* chronic infection of the cystic fibrosis lung and up-regulated coincident with a decrease in the levels of the major secreted siderophore pyoverdine (Marvig et al., 2014; Nguyen et al., 2014). This increase is in large part caused by point mutations within the promoter region of the phu operon leading to increased expression of TonB-

dependent outer membrane receptor (PhuR) (Marvig et al., 2014).

The main agreed antimicrobial drug classes target essential bacterial cell functions including cell wall formation, protein synthesis, and DNA synthesis (Fàbrega et al., 2009; Fair et al., 2012). The outgrowth of anti-biotic-resistant infectious agents of mycobacterium species is a serious health problem globally nowadays (Ray, 2012). Because of outgrowth of multidrug resistance against *P. aeruginosa* drugs, there is an urgent need for the development of new drug candidate as well as gaining deeper knowledge of the molecular mechanism of action of current and forthcoming active compounds. Understanding the molecular mechanisms by which *P. aeruginosa* acquires and uses heme may provide a therapeutic strategy in the treatment of *P. aeruginosa* chronic infections. Therefore, the aim of the research is to detect and filtering successful compounds and the knowledge of polarization, new molecules with high predicted antibacterial activity were theoretically designed. In this circumstance, QSAR, 3D-QSAR, Molecular Docking, MD simulations and Molecular Design studies were performed to the antibacterial activities reported by PubChem (Antibacterial activity against *P. aeruginosa* JSRI-1 clinical isolate after 12 hours).

MATERIAL AND METHODS

Computational method

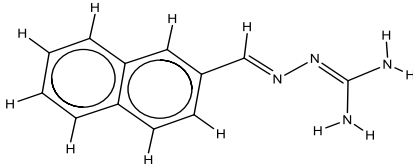
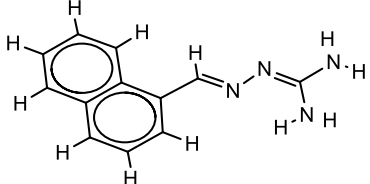
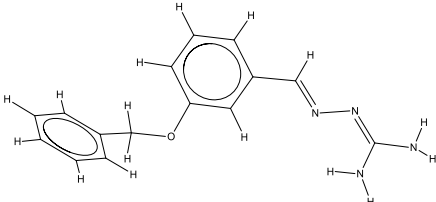
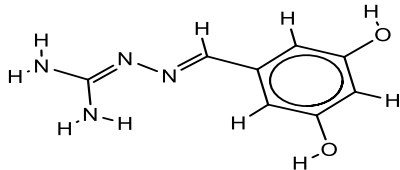
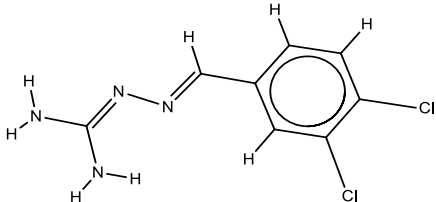
All the compounds and biological activity data used in this study were obtained from literature reported by PubChem (National Center for Biotechnology Information, 2019). The experimental MIC₅₀ values of all compounds in ug.mL⁻¹ where converted in pMIC₅₀ by taking $-\text{Log}(1/\text{MIC}_{50})$ and were used as the dependent variable. In this study, Kennard-Stone method (DTC-Lab v1.2) [<http://dtclab.webs.com/software-tools>] was performed for the classification of the compounds into training and test sets. The data set contains training set (17 compounds, 70%), and test set (8 compounds, 30%) and has uniform distribution. Structures and associated inhibitory activities were listed in Table 1 (the testing set is marked by *). The molecular structural scheme of compound of iminoguanidine based inhibitors of HemO and the substitutions are illustrated in Figure 1. All the 2D structures were generated by Marvin View software package v19.15 and transferred to Spartan'14 v14.1.4 software to create the 3D structure, pre-optimized using semi-empirical PM3 minimization. Then a more precise optimization was performed to make the conformations with least potential energy using Merck molecular force field (MMFF) with density function theory (DFT) with Becke's three-parameter hybrid functional (Becke, 1993) using LYP correlation functional (Lee et al., 1988). The standard Pople's 6-311+G** basis set was used. Descriptors were calculated using Spartan'14 v1.1.4

(Wavefunction, 2015), and PaDEL-Descriptor v2.20 (Yap, 2011) which include: quantum-chemical, molecular orbital, thermodynamics, charged partial surface area, electrostatic geometrical, and topological descriptors.

CoMFA and CoMSIA Modelling

3D-QSAR model was carried out through Molecular Interaction Fields (MIFs) analysis by dividing the dataset of 25 molecules into training set of 16 molecules and test set of 7 molecules in a random manner. At First, the molecules were superimposed using the atom-based and pharmacophore-based alignment by the Open3DALIGN tools v2.3 (Tosco et al., 2011). The energy minimized and optimized structures were aligned by the template-based method, and compound 18 was selected as the template to construct other compounds because of its high representative chemical structure and the alignment was completed by Open3DALIGN software. The alignment corresponding to the highest cumulative O3A score was selected for further analysis. Figure 2 shows the best alignment in which compound 18 (O3A_score = 2961.61) was selected as the template. CoMFA is a versatile method to describe 3D-QSAR quantitatively. Open3DQSAR software (version 2.3) is open-source software available for high-throughput chemometric analysis of molecular interaction fields (MIFs) (Tosco and Balle, 2011). The molecular structures were exported to the software Open3DQSAR (Tosco and Balle, 2011) for rigorous statistical manipulation of the molecular interaction fields written by Tosco and Balle (2011). Considerable attention has been set on computerization through the application of a scriptable interface, as well as on high computational performance achieved by algorithm parallelization. Tractability and the ability to exchange and use information with existing molecular modeling software make Open3DQSAR a powerful tool in pharmacophore assessment and ligand-based drug design (Ghasemi and Shiri, 2012). After importing MIFs descriptors from Open3Dalign medium into the Open3DQSAR, the following steps were performed: (1). Zeroing (sets to zero grid values which are close to zero); (2). Standard deviation cut-off (removes variables having a standard deviation among different objects lower than a user-defined threshold, in order to improve the signal to-noise ratio); (3). Removes variables assuming only a few different values across the different objects to prevent them from biasing the model (N-level variable elimination). Also, the following variable clustering and selection procedures (Baroni et al., 1993) have been implemented in Open3DQSAR; (4) D-optimal design variable selection (De Aguiar et al., 1995); (5). Smart region definition (SRD), as previously described by Pastor and co-workers (Pastor et al., 1997). Smart region definition (SRD) groups variables on the basis of their original localization in 3D space. This technique reduces dismissal arising from the

Table 1. Structures and structural formulae of compounds and their pMIC₅₀ values.

Compound	R ₁	R ₂	R ₃	ugmL ⁻¹	pMIC ₅₀
1*	-Cl	-H	-H	10.7	1.0294
2	-NC ₂ H ₆	-H	-H	32.5	1.5119
3	-H	-OH	-H	58.5	1.7672
4	-H	-OCH ₃	-H	38.4	1.5843
5	-NO ₂	-H	-H	61.4	1.7882
6	-H	-F	-H	28.6	1.4564
7				31.7	1.5011
8	-H	-H	-Cl	19.2	1.2833
9	-H	-Cl	-H	16.3	1.2122
10	-CH(CH ₃)CH ₃	-H	-H	20.6	1.3139
11	-OCH ₃	-H	-H	30.5	1.4843
12				30.4	1.4829
13*	-F	-H	-H	87.4	1.9415
14*	-Br	-H	-H	19.7	1.2945
15				25.5	1.4065
16*	-H	-Br	-H	62.9	1.7987
17*	-H	-OH	-Br	145	2.1614
18	-C ₆ H ₅	H	H	34.6	1.5391
19				338	2.5290
20	-OH	-H	-H	443	2.6464
21	-H	-H	-OH	391	2.5922
22	-H	-H	-OCH ₃	48.6	1.6866
23*	-H	-H	-F	70	1.8451
24*	-H	-H	-Br	23	1.3617
25*				24.1	1.3820

*Test set compound.

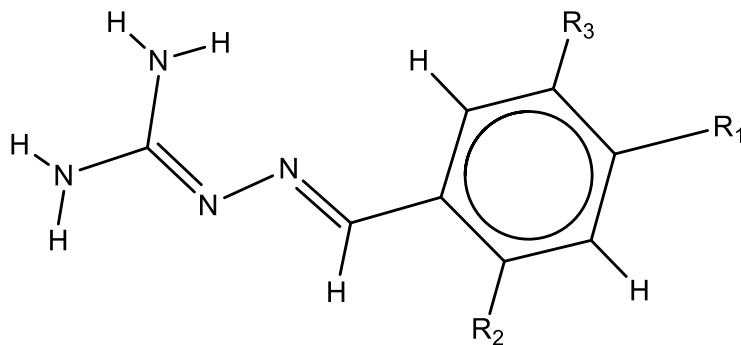


Figure 1. Molecular structural scheme of compound of iminoguanidine-based inhibitors of HemO.

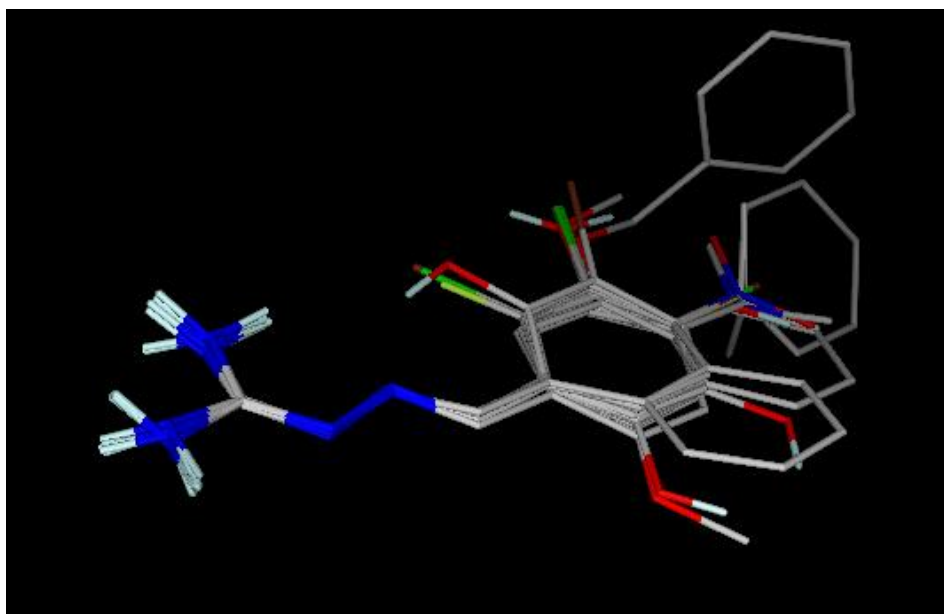


Figure 2. The best alignment in which compound 18 (O3A_ score = 2961.61) was selected as the template.

existence of multiple nearby descriptors which mainly encode the same kind of information; (6). Fractional factorial design (FFD) variable collection as initially described by Baroni et al. (1992) and implemented in GOLPE. Fractional factorial design (FFD) collection aims at choosing the variables which have the major effect on predictivity, and can operate on both single variables or on groups identified by a previous SRD run (Ghasemi and Shiri, 2012).

Docking and molecular dynamics simulation

Molecular docking studies were performed to further explain the results of the 3D-QSAR study and explore the possible binding modes of these compounds using

Discovery Studio visualizer v16.1.0.15350. The X-ray crystal structure of PhuS as a heme-degrading enzyme from *Pseudomonas aeruginosa* (PDB code: 1SK7) (Friedman et al., 2004) was obtained from the RCSB Protein Data Bank. All the Compounds were selected as ligand molecules to perform the docking study. After preparation, ligands were docked into the active site of PhuS as a heme-degrading enzyme from *P. aeruginosa* (PDB code: 1SK7), and ten (10) poses were written for each ligand. In order to further prove the reliability of the docking results and confirm the stability of the complex of all the ligands and PhuS as a heme-degrading enzyme from *P. aeruginosa*, a molecular dynamics (MD) simulations study was performed with CHARMM27 force field using NAMD (Phillips et al., 2005). Minimization was performed to optimize the initial structure of protein-ligand

Table 2. predicted activities of statistically significant models obtained by 2D models.

MLR Equation: $\text{PMIC}_{50} = 0.77105(+/-0.47069) - 0.00009(+/-0.00004) \text{ATS7v} + 3.19824(+/-0.87789) \text{ATSC0c} - 0.35816(+/-0.29462) \text{minsCl}$ ----- Model 1	
Internal Validation Parameters: SEE :0.2953; r^2 :0.6569; r^2 adjusted :0.5777; F :8.2955 (DF :3, 13) Leave-One-Out (LOO) Result: Q2 :0.5493; PRESS :1.4895; SDEP :0.296	
External Validation Parameters (Without Scaling): r^2 :0.7445; r_0^2 :0.7163; reverse r_0^2 :0.4242; $\text{rm}^2(\text{test})$:0.6195; reverse $\text{rm}^2(\text{test})$:0.3231; average $\text{rm}^2(\text{test})$:0.4713; $\Delta \text{rm}^2(\text{test})$:0.2964; rmsep :0.2008; rpred^2 :0.7123; Q2f1 :0.7123; Q2f2 :0.6941.	
Some External Validation Parameters (After Scaling): $\text{rm}^2(\text{test})$:0.6548; reverse $\text{rm}^2(\text{test})$:0.4257; average $\text{rm}^2(\text{test})$:0.5402; $\Delta \text{rm}^2(\text{test})$:0.2290.	
Golbraikh and Tropsha acceptable model criteria's:	
1. Q^2	0.54926 Passed (Threshold value $Q^2 > 0.5$)
2. r^2	0.74445 Passed (Threshold value $r^2 > 0.6$)
3. $ r_0^2 - r^2 $	0.29214 Passed (Threshold value $ r_0^2 - r^2 < 0.3$)
4. k	1.03432 $[(r^2 - r_0^2)/r^2]$ 0.03783 OR* k' 0.95341
	$[(r^2 - r_0^2)/r^2]$ 0.43025 Passed (Threshold value: $[0.85 < k < 1.15$ and $((r^2 - r_0^2)/r^2) < 0.1$] OR* $[0.85 < k' < 1.15$ and $((r^2 - r_0^2)/r^2) < 0.1$])

Table 3. Order of entered descriptors in MLR method.

Descriptor name	Description about selected variables	VIF	MF / (%)
ATS7v	Broto-Moreau autocorrelation of lag 7 (log function) weighted by van der Waals volume	1.0686	-0.4684 / (-23.26%)
ATSC0c	Centered Broto-Moreau autocorrelation - lag 0 / weighted by charges	1.1616	1.5069 / (74.83%)
minsCl	Minimum atom-type E-State: -Cl	1.2155	-0.0385 / (-1.91)

complex. After that, the temperature of system was gradually heated up from 0 K to 310 K in 100ps. Finally, the system was equilibrated at 310 K for 100ps with NVT ensemble.

RESULTS AND DISCUSSIONS

QSAR study was carried for a series of 25 iminoguanidine derivatives, in order to determine a quantitative relationship between structure and potential antibacterial activities. The results obtained for QSAR using MLR method are represented in Table 2.

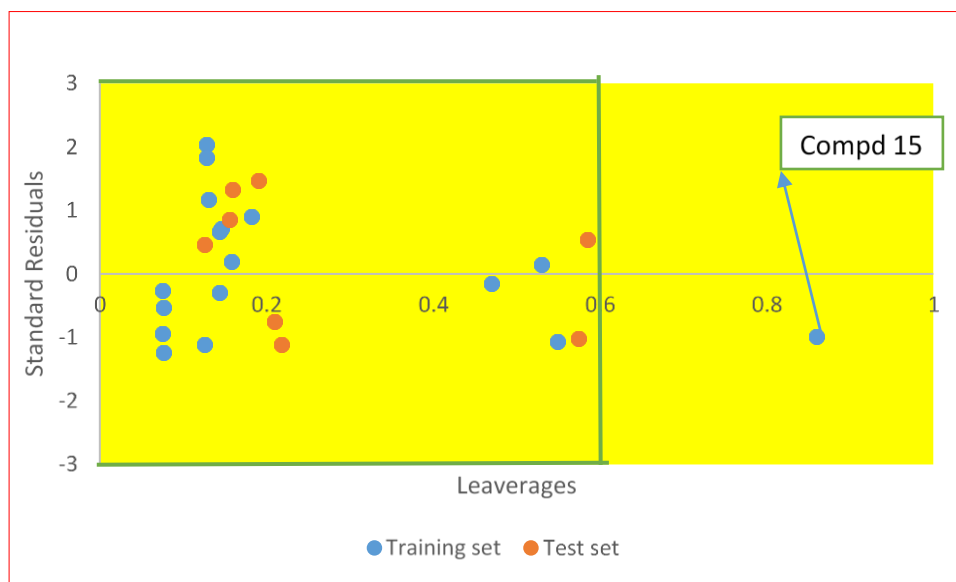
To observe the comparative value as well as the influence of each descriptor in the model, the value of the mean effect (MF) was calculated for each descriptor (Massart et al., 1997). The MF value indicates the relative importance of a descriptor, related with the other descriptors in the model. The mean effect values are shown in Table 3. As can be seen, the ATSC0c descriptor have great mean effect values than the other descriptors which means that these descriptors have a large effect on the pMIC_{50} of the studied compounds. The statistically best model for antibacterial activity against *Pseudomonas aeruginosa* JSRI-1 with a coefficient of determination (R^2) = 0.6569 was considered, as the model showed an internal

predictive power ($Q^2 = 0.5493$) of 55% and a predictivity for the external test set ($R^2_{\text{pred}} = 0.7123$) of about 71%. This model indicates the negative contribution of ATS7v (-23.26%) (This descriptor signifies Broto-Moreau autocorrelation of lag 7 (log function) weighted by van der Waals volume) with biological activity in the QSAR models which suggests that decreasing the ATS7v of the compounds will lead to increased activity and descriptors ATSC0c define the Centered Broto-Moreau autocorrelation - lag 0 / weighted by charges and make a positive contribution (74.83%) to activity which suggests that increasing the ATSC0c of the compounds will lead to increased activity.

The last descriptor minsCl in model represents minimum atom-type E-State: -Cl. The descriptor is negatively correlated (-1.91%) with biological activity in the QSAR models which suggests that decreasing the minsCl of the compounds will lead to increased activity. The correlation matrix between the descriptors and the activity for the models 1 is presented in Table 4. The multi-collinearity between the above three descriptors was detected by calculating their variation inflation factors (VIF). The corresponding VIF values of the three (3) descriptors (Table 3), all the variables have VIF values of less than four, indicating that the obtained model has statistical significance, and the descriptors were found to reasonably

Table 4. Correlation matrix for descriptors influencing the antibacterial activity.

	PMIC50	ATS7v	ATSC0c	minsCl
PMIC50	1			
ATS7v	-0.3246	1		
ATSC0c	0.7309	-0.0488	1	
minsCl	-0.3679	-0.2158	-0.3506	1

**Figure 3.** The Williams plot, the plot of the standardized residuals versus the leverage value.

orthogonal.

The leverage values (abscissa) can be calculated for every compound and plotted vs. standardized residuals (ordinate), and it allows a graphical detection of both the outliers and the structural outlier in a model. Figure 3 shows the Williams plot. It is noticeable that all compounds in the test set fall inside the domain of the MLR model (the warning leverage limit is 0.59). There are only one chemicals (compound 15 in the training set) which have the leverage higher than the warning h^* value ($h^* = 3(k+1)/n$, where k is the number of model parameters and n is the number of compounds (Netzeva et al., 2005; OECD, 2007), therefore, they can be regarded as structural outliers. Luckily in this case, the data predicted by the model are good for compound 15, so, they are “good leverage” compound. For all the compounds in the training and test sets, their standardized residuals are smaller than three standard deviation units (3δ). Hence, they are no outliers.

Validation of 3D-QSAR modeling and its Steric and electrostatic contribution plot

The comparative molecular field analysis (CoMFA) model

showed LOO cross-validation $Q^2 = -0.2289$ with other statistical value, non-cross-validation $R^2 = 0.9687$, F (F-test) value = 61.9012, SDEC = 0.0627 and SDEP = 0.3926; the steric and electrostatic contributions were 60.29 and 39.71%, respectively. The correlation between experimental and predicted pMIC₅₀ for training set is shown in Figure 4. Regularly, a high Q^2 value ($Q^2 > 0.5$) is considered as a proof of high predictive ability of the model (Bohm et al., 1999). The developed model was found not to be statistically significant due to its low cross-validation Q^2 . The CoMSIA model showed cross validation $Q^2 = 0.5483$ with other statistical value, non-cross-validation $R^2 = 0.9778$, F-test = 88.1020, SDEC = 0.0528 and SDEP = 0.238; the steric and electrostatic contributions were 67.53 and 32.47%, respectively. To test the stability and predictive ability of the 3D-QSAR model, seven compounds were selected as the test set for validation. The predicted pMIC₅₀ for training and test set is listed in Table 4. With the CoMFA model, the external cross-validated value (R^2_{pred}) was -0.5581 and SDEP = 0.7238. And with the comparative molecular similarity indices analysis (CoMSIA) model, the R^2_{pred} value was -0.7631 and SDEP = 0.2380. The correlation between experimental and predicted pMIC₅₀ for training set is shown in Figure 6. It was shown that the CoMFA model

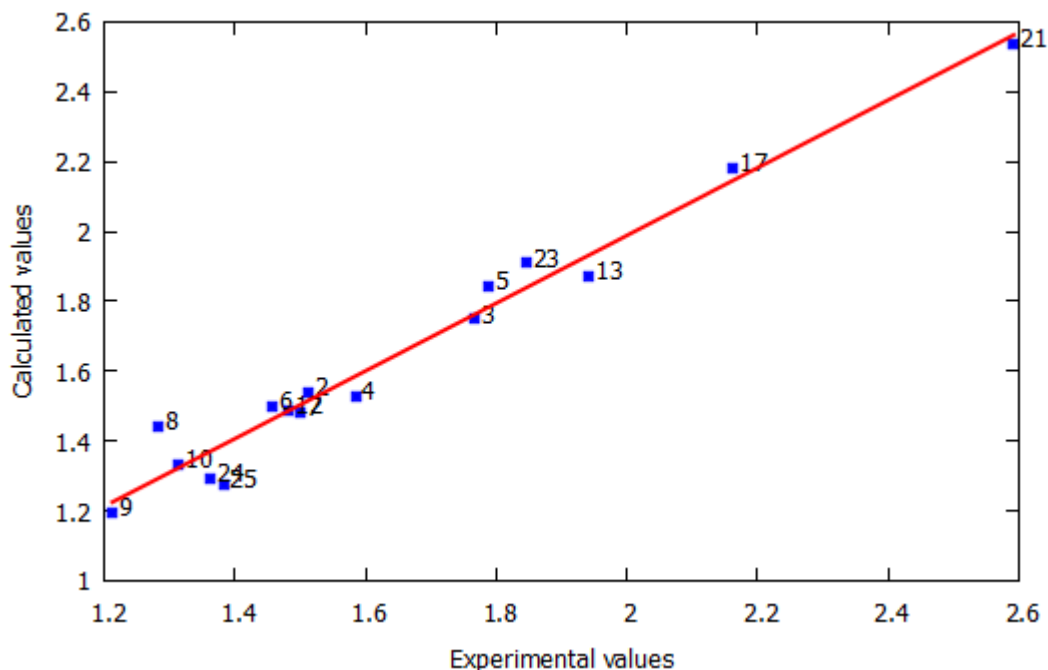


Figure 4. Plots of experimental and calculated pMIC₅₀ for CoMFA analysis.

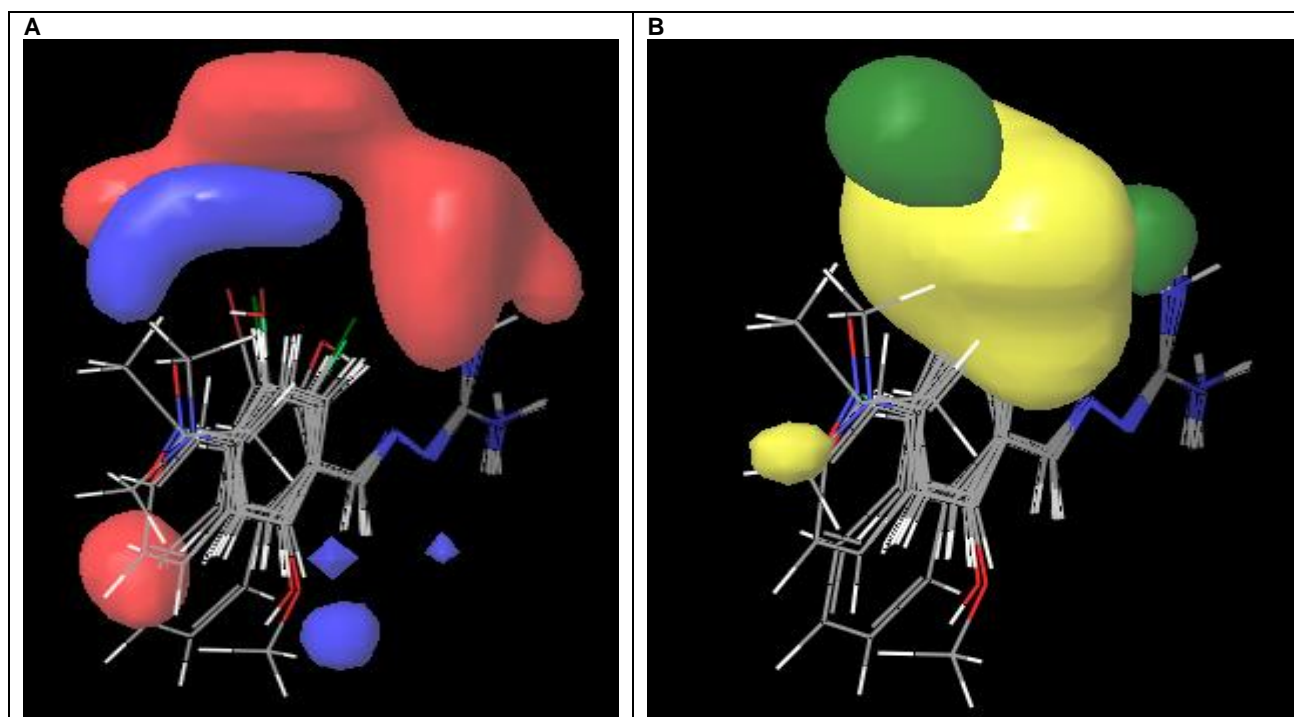


Figure 5. 3D comparative molecular field analysis (CoMFA) contour maps. A) Steric field: blue: sterically favorable; red: sterically unfavorable and B) Electrostatic field: red: negative potential favorable; blue: negative potential unfavorable.

were unstable and does not provide robust predictive ability, while CoMSIA model provide robust predictive ability in the training set, unstable in the external/test set. Figure 5 shows the CoMFA contour plots, to visualize the

information content of the derived 3D-QSAR models. These contour plots showed regions in 3D space around the molecules, where variation in specific molecular properties increase or decrease the activity. CoMFA steric

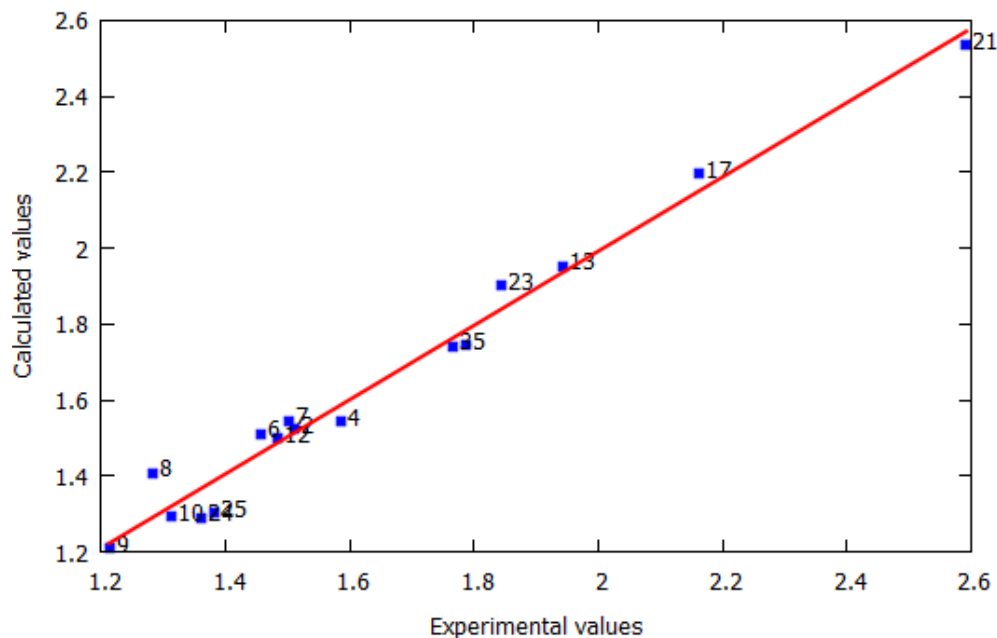


Figure 6. Plots of experimental and calculated pMIC₅₀ for CoMSIA analysis.

(75.87%, Figure 5(A)) and electrostatic (24.13%, Figure 5(B)) contour plots using all the training set compounds as a reference molecular. In Figure 5(A), the blue contours indicate regions, where bulky groups will increase the activity, while the red contours indicate regions of steric hindrance to activity. In Figure 5(B), green contours indicate regions, where electron-donating groups will increase activity, and yellow contours indicate regions, where electron-withdrawing groups will increase the activity. As shown in Figure 5, the R₁ and R₃ site of all compounds was oriented toward red and yellow contours indicating that a less bulky substituent would be favored. This may explain why compounds 20 and 21, which possess less bulky substituents at R₁ and R₃ position have significantly better activity than other compounds, which have major groups at the R₁ and R₃ position. The blue and green contours at the R₂ and R₃ site of all the compounds indicated that bulky groups at this position might increase the activity.

Figure 7 provides the steric and electrostatic contour for 16 compounds of the CoMSIA model. The steric interaction for CoMSIA model was shown in Figure 7A. The blue contours display the hydrogen bond donor groups are favorable while blue contours state these groups would decrease activity. One medium size OCH₃ or (phenol) contour around R₁ and R₃ suggests that hydrogen donor group in this region would increase activity. In hydrogen bond acceptor field, the blue and red plots indicate favorable and unfavorable situations. In Figure 7B, one H-bond acceptor favorable to yellow contour located near R₁, revealed that both hydrogen bond donor and acceptor groups can increase activity. Great

number of the derivatives involved in this study possessed hydrogen bond acceptor group (OCH₃) at this site which demonstrates the importance of the hydrogen bond acceptor substituent.

Docking results

The structure, based virtual screening was done using 25 structural analogs of the existing ligand of Allosteric Inhibitors of the Iron-Regulated Heme Oxygenase (HemO) of *Pseudomonas aeruginosa* through AutoDockVina (Dallakyan and Olson, 2015) in PyRx software (Trott and Olson, 2010) and the docking simulation energies were determined for each ligand with 8 exhaustiveness. Lamarckian geometric algorithm (LGA) were carried out to calculate Root Mean Square Deviation (RMSD). Active site grid dimensions were set at X = 11.0731Å, Y = 1.3633Å and Z = 10.7386Å for the center and total size dimensions were set as X = 46.2499Å, Y = 46.4106Å and Z = 39.6166Å. All the best leads superimposed core cavity interactions were visualized through Ligplot v2.1 and Discovery Studio visualizer v16.1.0.15350 software (Figure 8 and 9) respectively. All iminoguanidine derivatives were docked into the binding site of HemO and the binding energy scores for each molecule are shown in Table 5, where precise correlations could be found between docking scores and pMIC₅₀ values. Docking studies showed that compound 18 is suitably situated at the binding site and there are various interactions between it and the binding region of the enzyme. The iminoguanidine binds to the HemO through three key

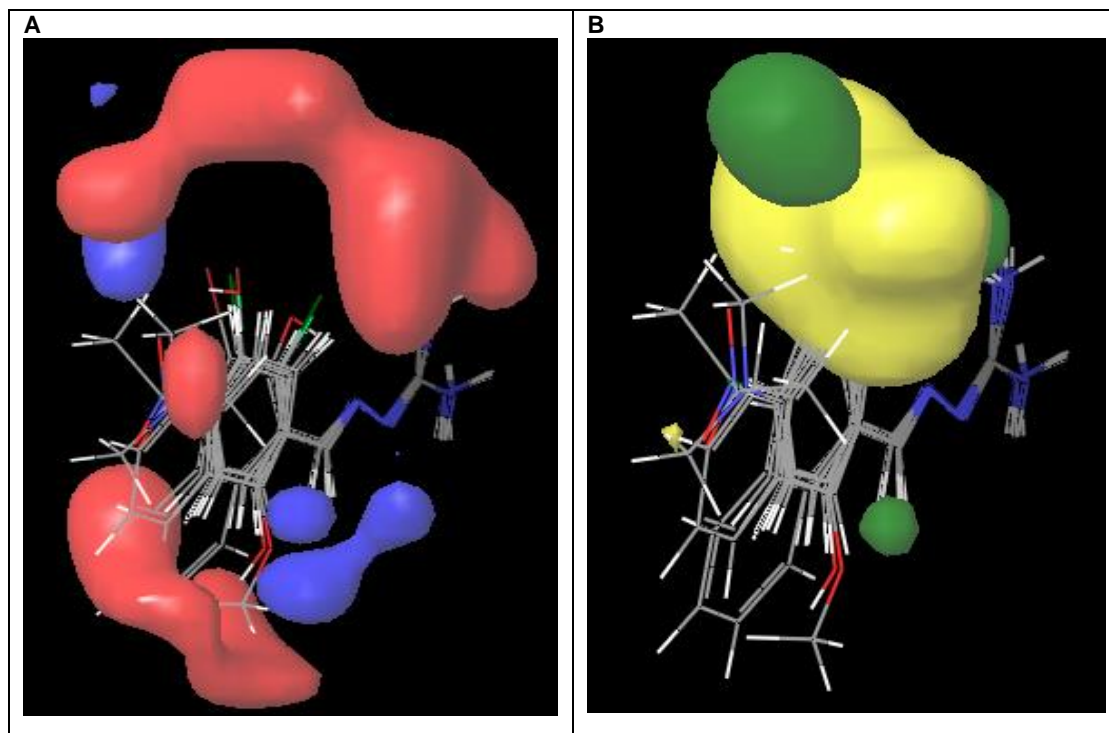


Figure 7. COMSIA contour maps based on 16 training set compounds (A), Steric (B) Electrostatic.

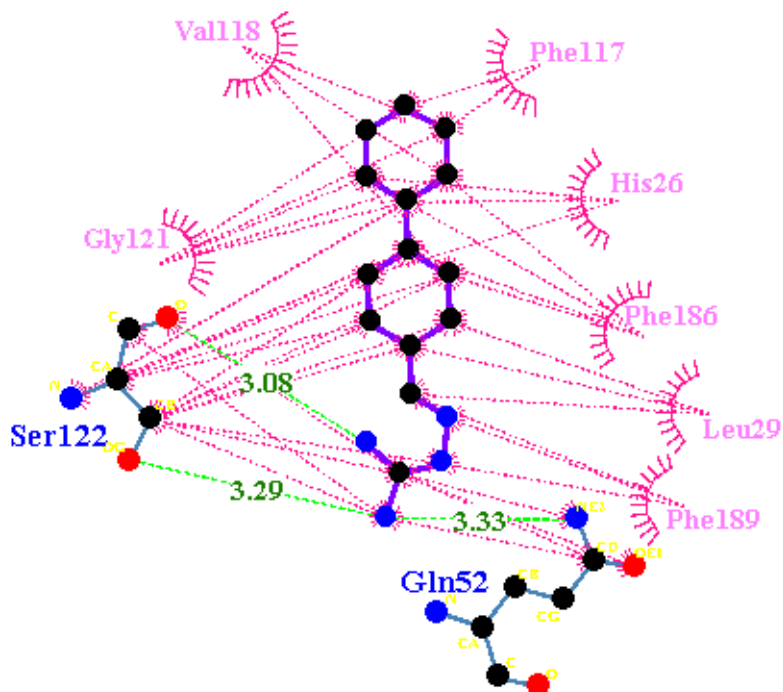


Figure 8. Connection of docked highest potent iminoguanidine compound 18 at the active site of 1SK7 produced using the PyRx program/LigPlus.

hydrogen bond interactions: (1) between the NH of Gln52 and the NH of the iminoguanidine substituent (2) between

the NH of Ser122 and the NH of the iminoguanidine substituent (3) between the O of Ser122 and the N of the

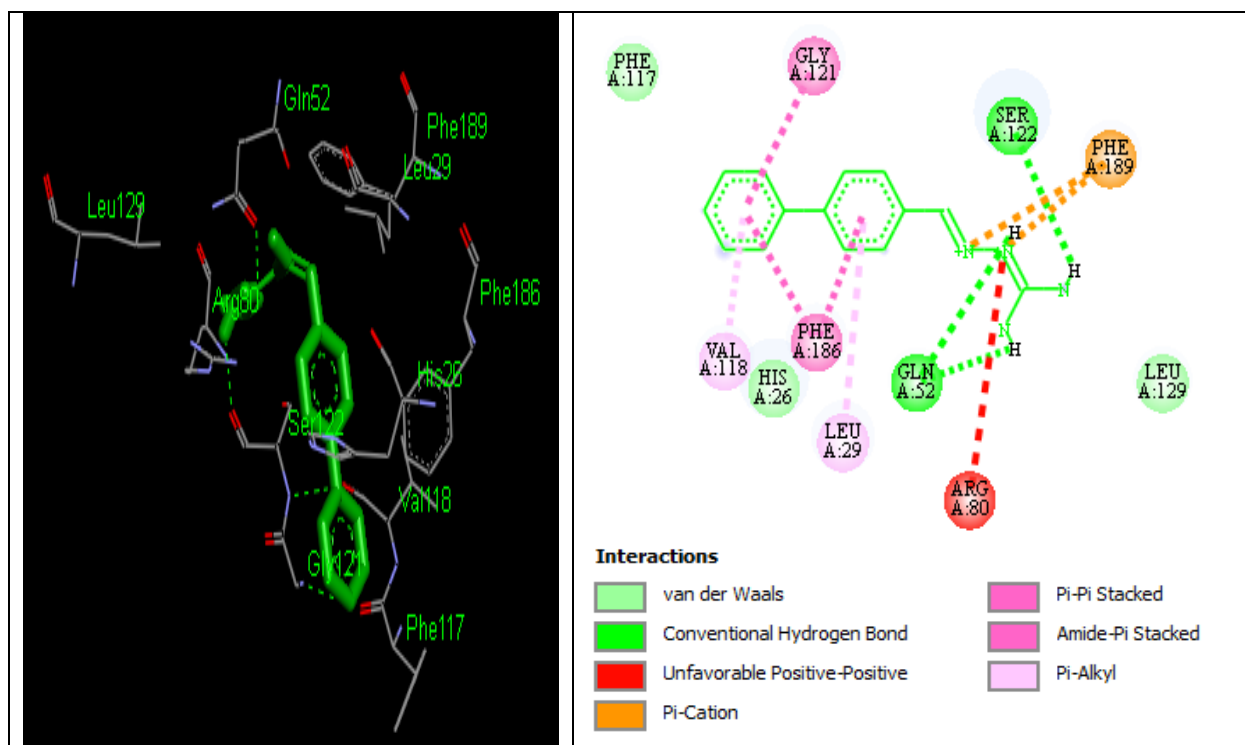


Figure 9. Connection of docked highest potent iminoguanidine compound 18 at the active site of 1SK7 produced using the PyRx program/Discovery Studio.

Table 5. The evaluation of PLS statistics results of 3D QSAR Models of CoMFA, CoMSIA and with PyRx docking results.

Comp No.	pMIC50	Docking Scores	CoMFA	CoMFA Residual	CoMSIA	CoMSIA Residual
1*	1.0294	-6.6	1.5200	-0.4906	1.428	-0.3986
2	1.5119	-6.6	1.5404	-0.0285	1.5232	-0.0113
3	1.7672	-6.4	1.7493	0.0179	1.7408	0.0264
4	1.5843	-6.4	1.5301	0.0542	1.5430	0.0413
5	1.7882	-7.0	1.8440	-0.0558	1.7444	0.0438
6	1.4564	-6.5	1.4995	-0.0431	1.5104	-0.054
7	1.5011	-8.0	1.4839	0.0172	1.5432	-0.0421
8	1.2833	-6.5	1.4418	-0.1585	1.4079	-0.1246
9	1.2122	-6.5	1.1978	0.0144	1.2082	0.004
10	1.3139	-6.9	1.3356	-0.0217	1.2923	0.0216
11*	1.4843	-6.5	1.3229	0.1614	1.3153	0.169
12	1.4829	-7.8	1.4882	-0.0053	1.4975	-0.0146
13	1.9415	-6.5	1.8738	0.0677	1.9538	-0.0123
14*	1.2945	-6.5	1.8482	-0.5537	1.9755	-0.681
15 ^R	1.4065	-8.2	-	-	-	-
16*	1.7987	-6.6	1.0654	0.7333	1.0871	0.7116
17	2.1614	-6.7	2.1837	-0.0223	2.1948	-0.0334
18 ^R	1.5391	-8.4	-	-	-	-
19*	2.5290	-6.5	1.6886	0.8404	1.6867	0.8423
20*	2.6464	-6.6	1.4187	1.2277	1.3255	1.3209
21	2.5922	-6.3	2.5353	0.0569	2.5345	0.0577
22*	1.6866	-6.5	2.2720	-0.5854	2.3256	-0.639
23	1.8451	-6.6	1.9091	-0.064	1.8997	-0.0546
24	1.3617	-6.6	1.2947	0.067	1.2898	0.0719
25	1.3820	-6.7	1.2782	0.1038	1.3017	0.0803

*Test set compounds; ^RRemove from the model.

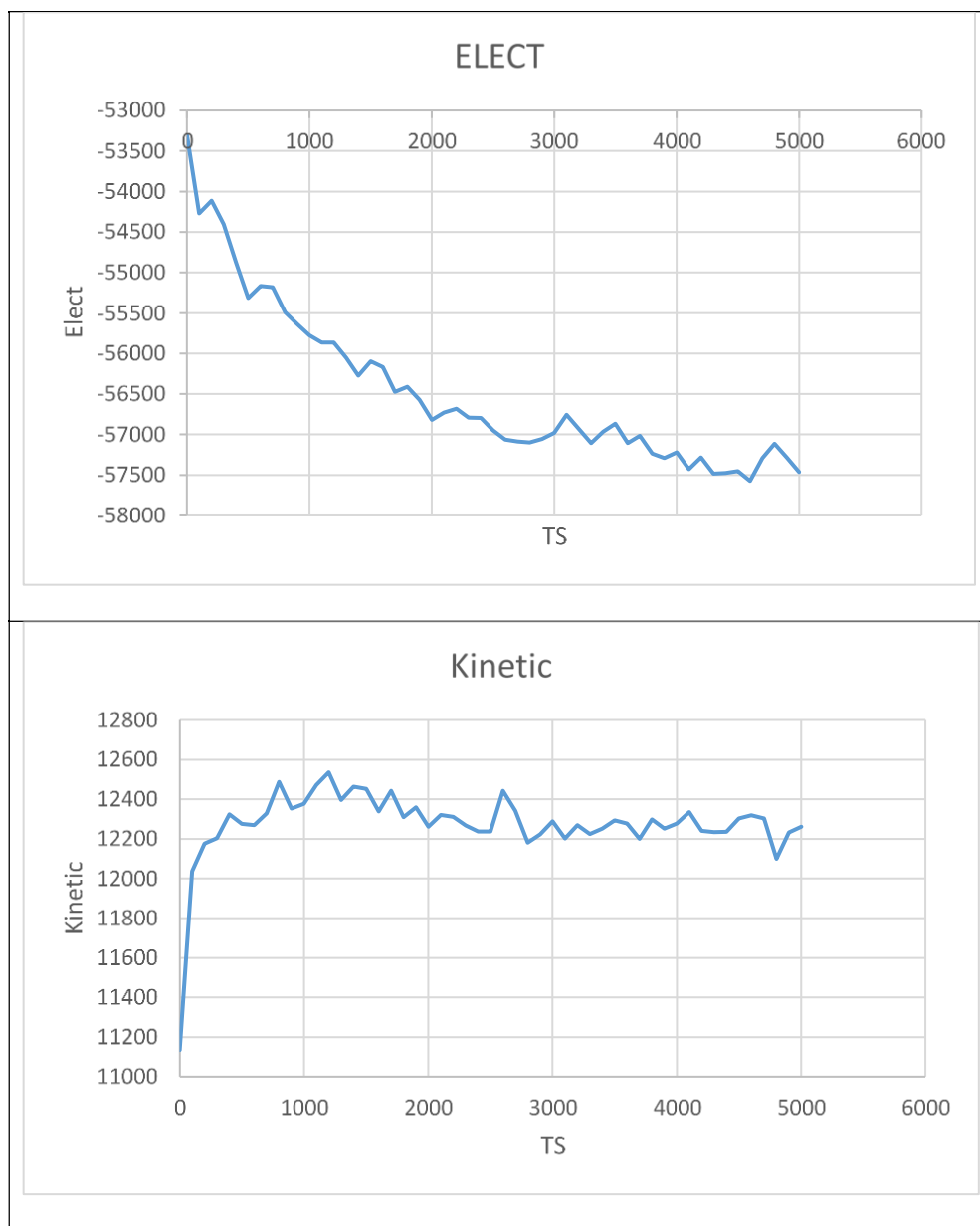


Figure 10. Molecular dynamics simulations study of A-chain of protein (PDB ID: 1SK7) with respect to (A) Time vs Electrostatic; (B) Time vs Kinetic energy.

iminoguanidine substituent. The hydrogen bonding distances observed were 3.33Å (NH₂-H-NH- Gln52), 3.08Å (O₂-H-NH- Ser122), 3.29Å (O₂-H-NH-Ser122) as shown in Figure 8. The docking results reported in Table 5, disclose that hydrogen bonding may be accountable for activity, which may be further increased on adding high electronegative substitutions.

The result of MD simulations proved the stability of protein-ligand complex and provided additional information about the binding mode of compound 18. The electrostatic, kinetic energy, total energy and temperature plot of complex during the MD simulations are shown in Figure 10. The simulation result reveals that the temperature

tends to be stable and fluctuated at about 296K, electrostatic at about 57500, kinetic energy at about 12200 kcal/mol and total energy is fluctuated at about 30500 kcal/mol after running 100ps, which implies that the structure of the complex is basically in a stable state. The conformation of the complex was compared after simulation with the initial. As shown in Figure 11, the compound 18 adopt a similar pose and were placed in the same region of protein, which means that the compound can be stably located in the docking site. And then, the residues which interact with compound 18 after the simulation were compared. As shown in Figure 11, compound 18 went slightly different into the HemO binding

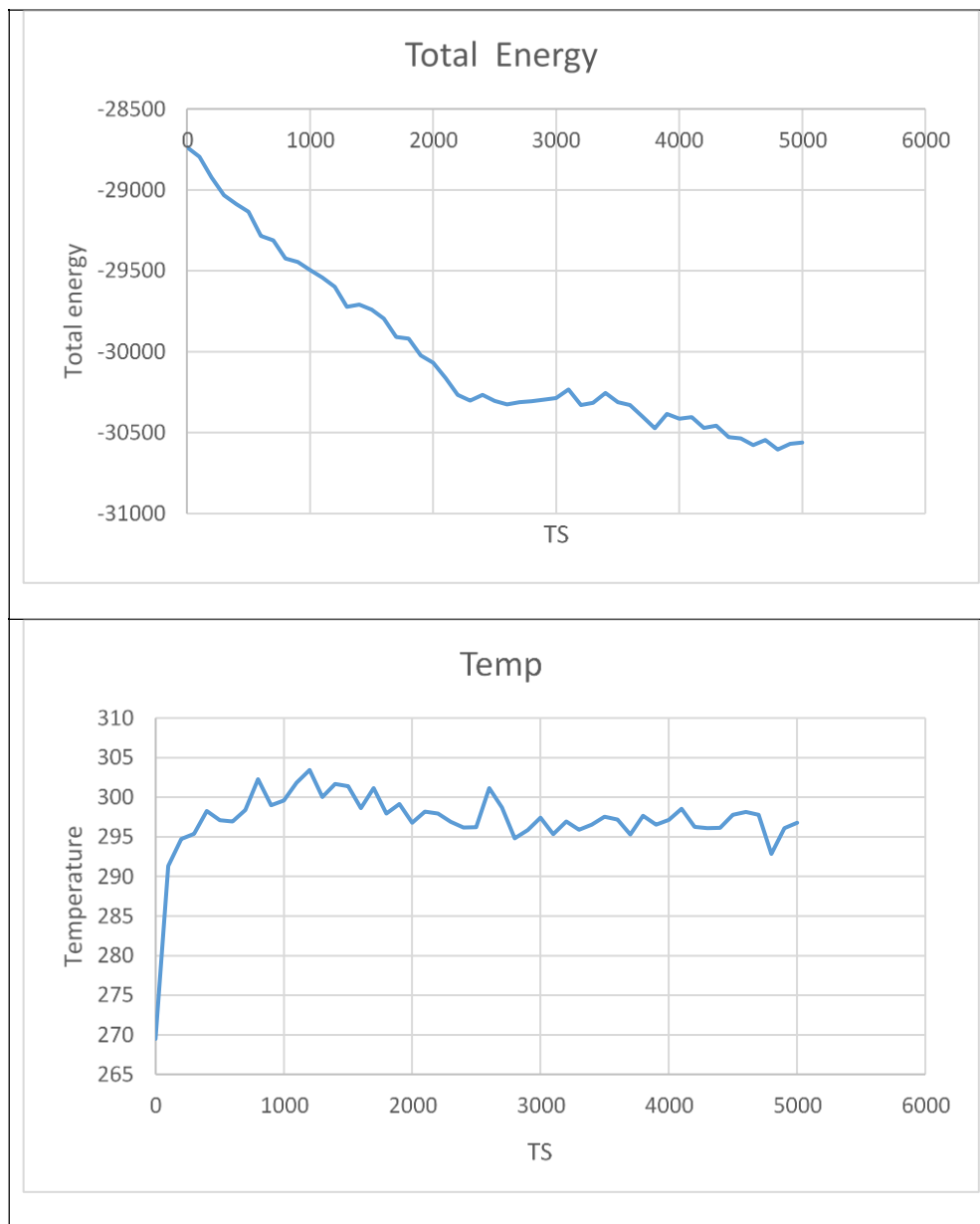


Figure 10 Contd. Molecular dynamics simulations study of A-chain of protein (PDB ID: 1SK7) with respect to (C) Time vs Total energy; (D) Time vs Temperature.

pocket. The iminoguanidine-based inhibitors of HemO still had hydrophobic interactions with surrounding residues. This result further explained the importance of phenyl, naphthalene OCH₃ and OC₆H₅ and on the inhibitory activity. Though this molecular modeling study (CoMFA, CoMSIA, molecular docking and MD simulations), there was a clear understanding of the structure–activity relationship and binding modes of iminoguanidine-based inhibitors of HemO. The results of 3D-QSAR revealed the effect of each substituent on the activity, and the docking results confirmed this effect. Meanwhile, the results of molecular docking also revealed some key interactions between the protein and the ligands. The later MD

simulations further explained the significance of phenyl, naphthalene, OCH₃, and OC₆H₅. All of this molecular modeling studies would provide useful information and a powerful tool for the next design and optimization of new HemO inhibitors.

Design of new compounds with high predicted antibacterial activity

The chemical and biological information obtained using 2D-QSAR, Molecular Docking, 3D-QSAR pharmacophore modeling (CoMFA and CoMSIA) studies has been widely

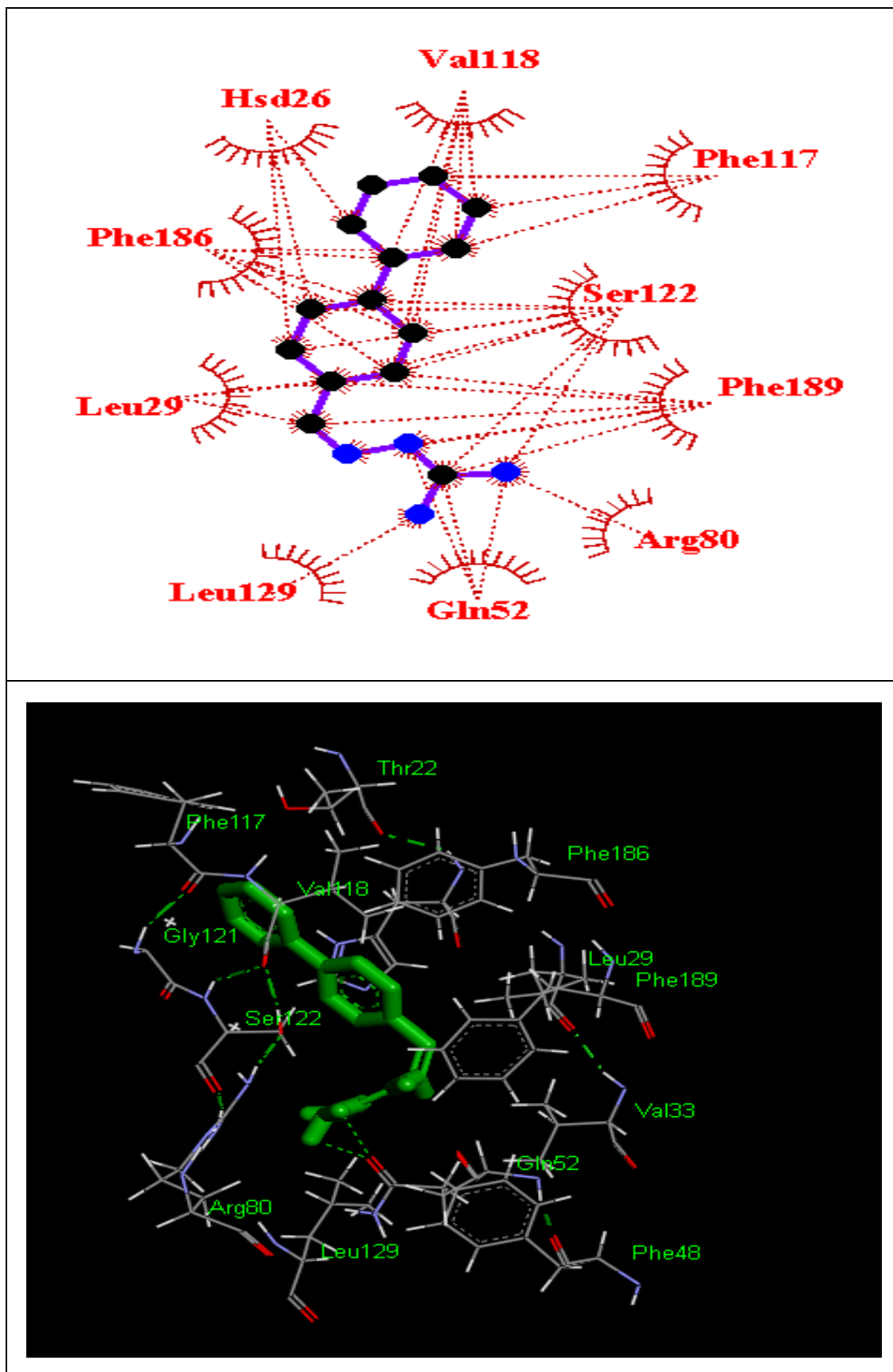


Figure 11. Binding interactions of compound 18 with simulated protein structure PDB: 1SK7.

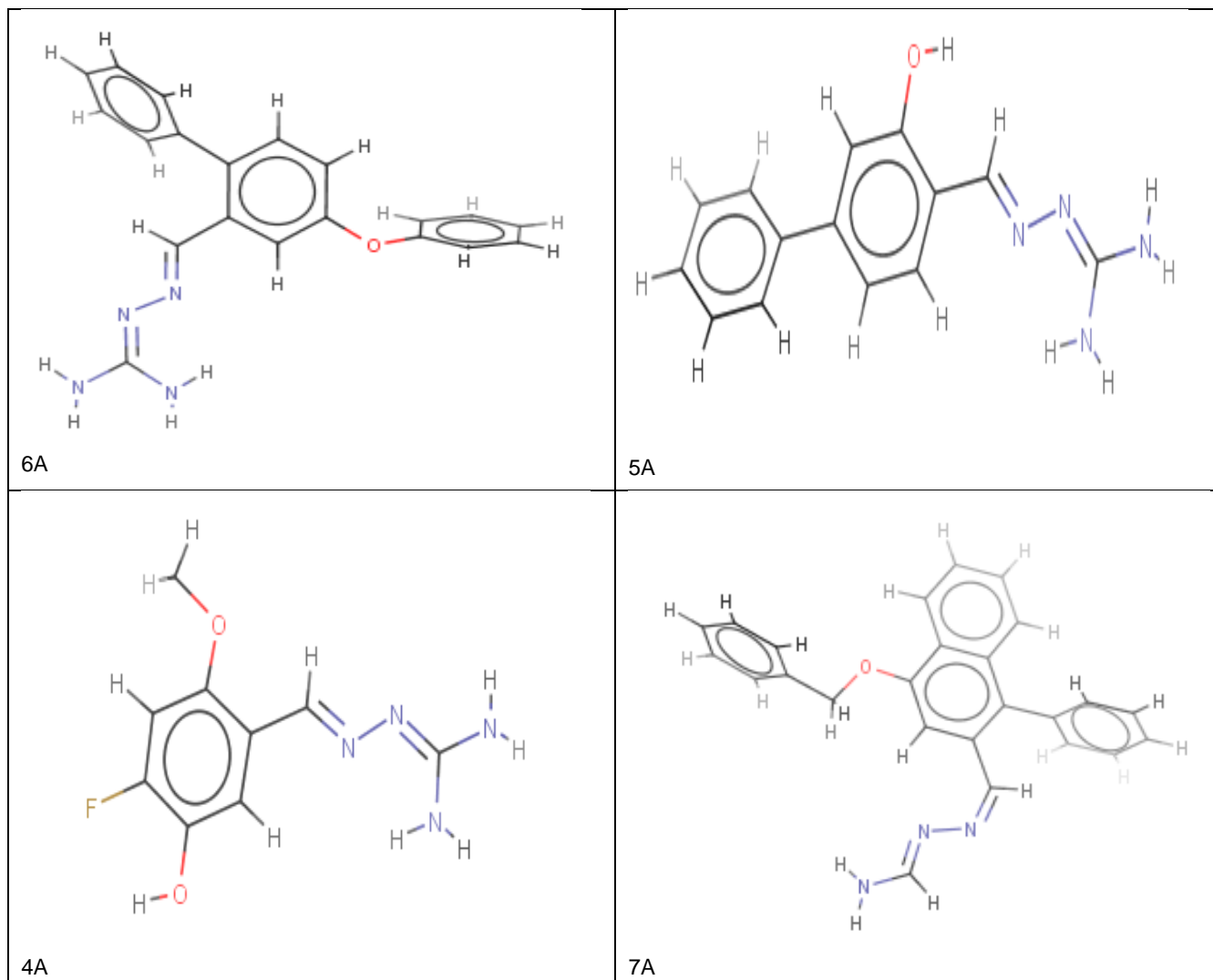


Figure 12. The structures of 4 new designed compounds of iminoguanidine-based inhibitors of HemO.

used to design new ligands with improved activity profile. In the present work, the above-mentioned techniques were used to identify structural elements of iminoguanidine-based inhibitors of HemO. The structural information therefore obtained was used to design a virtual library of molecules based on the aforementioned frameworks and the following were found: the three descriptors very important and display greater correlation to $pMIC_{50}$; likewise, they show close correlation to substituent R_1 to R_3 . Hence, through modifying R structures to improve the values of the three descriptors, 4 new compounds with higher antibacterial activities have been theoretically designed. The considerations of designing new compounds are presented as follows: since C—F is stronger than those of C—O, so F atom with the largest electronegativity and smaller volume can be adopted instead of O atom, leading the increase in the positive charges of Carbon and the decrease in the value of

minsCI. In addition, the R's structure is modified to reduce its ATS7v value to near zero. The structures of 4 new designed compounds are shown in Figure 12. The values of three parameters of 4 new compounds calculated using the same methods are given in Table 6. Docking results and the $pMIC_{50}$ values predicted according to the Model are also presented in Table 6.

Table 6 shows that the predicted activities of three compounds (5A and 4A) are much greater ($pMIC_{50} = 1.7848$ and 2.5494, respectively) than some of the 25 compounds mentions in Tables 1 and 5; only the activity of compound 4A ($pMIC_{50} = 2.5494$) is almost equal to that of compound 19, 20 and 21 with the highest activity ($pMIC_{50} = 2.529$, 2.6464 and 2.5922 respectively) in this series of studied compounds. Such results further indicate that the model established through the QSAR, 3D-QSAR and molecular docking studies is significant and predictive, and that the consideration on the molecular design is also reasonable.

Table 6. Calculated results for 4 designed compounds.

Comp	ATS7v	ATSCoc	minsCI	pMIC ₅₀ (pred)	ugmL ⁻¹ (pred)	Docking Scores
7A	24519.86	0.375843	0	-0.2337	0.583846	-8.7
6A	15652.74	0.491591	0	0.934531	8.600637	-9.4
5A	7751.453	0.535102	0	1.784804	60.92624	-8.5
4A	5015.476	0.697172	0	2.549381	354.3082	-6.8

Conclusion

The present work shows how iminoguanidine-based inhibitors of HemO may be treated statistically to uncover the molecular characteristics which are essential for high activity. The generated models were examined and confirmed for their statistical significance and external prediction power. The awareness and understanding of the descriptors, 3D-QSAR, docking and MD simulations involved in antibacterial activity of these compounds provide a great opportunity for the ligand structures design with appropriate features, and for the explanation of the ways in which these features affect the biological data upon binding to the respective receptor target. Four new compounds with higher predicted antibacterial activity have been hypothetically designed, and they are expected to be confirmed experimentally.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the Department of Chemistry, Ahmadu Bello University, Zaria (Samaru, Zaria-Nigeria); for computational studies and as part of the PhD thesis.

REFERENCES

- Baroni, M., Clementi, S., Cruciani, G., Costantino, G., Riganelli, D., & Oberrauch, E. (1992). Predictive ability of regression models. Part II: Selection of the best predictive PLS model. *Journal of chemometrics*, 6(6), 347-356.
- Baroni, M., Costantino, G., Cruciani, G., Riganelli, D., Valigi, R., & Clementi, S. (1993). Generating optimal linear PLS estimations (GOLPE): an advanced chemometric tool for handling 3D-QSAR problems. *Quantitative Structure-Activity Relationships*, 12(1), 9-20.
- Becke, A. D. (1993). Density-functional thermochemistry 3. The role of exact exchange. *Journal Chemical Physics*, 98, 5648-5652
- Böhm, M., Stürzebecher, J., & Klebe, G. (1999). Three-dimensional quantitative structure - activity relationship analyses using comparative molecular field analysis and comparative molecular similarity indices analysis to elucidate selectivity differences of inhibitors binding to trypsin, thrombin, and factor Xa. *Journal of Medicinal Chemistry*, 42(3), 458-477.
- Dallakyan, S., & Olson, A. J. (2015). Small-molecule library screening by docking with PyRx. In *Chemical Biology* (pp. 243-250). Humana Press, New York, NY.
- de Aguiar, P. F., Bourguignon, B., Khots, M. S., Massart, D. L., & Phan-Thau-Luu, R. (1995). D-optimal designs. *Chemometrics and Intelligent Laboratory Systems*, 30(2), 199-210.
- Fàbrega, A., Madurga, S., Giralt, E., & Vila, J. (2009). Mechanism of action of and resistance to quinolones. *Microbial biotechnology*, 2(1), 40-61.
- Fair, R. J., Hensler, M. E., Thienphrapa, W., Dam, Q. N., Nizet, V., & Tor, Y. (2012). Selectively guanidinylated aminoglycosides as antibiotics. *ChemMedChem*, 7(7), 1237-1244.
- Friedman, J., Lad, L., Li, H., Wilks, A., & Poulos, T. L. (2004). Structural basis for novel δ -regioselective heme oxygenation in the opportunistic pathogen *Pseudomonas aeruginosa*. *Biochemistry*, 43(18), 5239-5245.
- Ghasemi, J. B., & Shiri, F. (2012). Molecular docking and 3D-QSAR studies of falcipain inhibitors using CoMFA, CoMSIA, and Open3DQSAR. *Medicinal Chemistry Research*, 21(10), 2788-2806.
- Kumar, S., & Bandyopadhyay, U. (2005). Free heme toxicity and its detoxification systems in human. *Toxicology Letters*, 157(3), 175-188.
- Lee, C., Yang, W., & Parr, R. G. (1988). Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. *Physical Review B*, 37(2), 785-789.
- Marvig, R. L., Damkjaer, S., Khademi, S. H., Markussen, T. M., Molin, S., & Jelsbak, L. (2014). Within-host evolution of *Pseudomonas aeruginosa* reveals adaptation toward iron acquisition from hemoglobin. *MBio*, 5(3), e00966-14.
- Massart, D. L., Vandeginste B. G. M., Buydens, L. M. C., Jong, S. D. E., Leui, P. J., & Smeyers-Verbeke, J. (1997). *Hand book of chemometrics and qualimetrics: Part A*. Elsevier, The Netherlands
- National Center for Biotechnology Information. PubChem Database. Source=ChEMBL, AID=1315713. Retrieved July 7, 2019 from <https://pubchem.ncbi.nlm.nih.gov/bioassay/1315713>.
- Netzeva, T. I., Worth, A. P., Aldenberg, T., Benigni, R., Cronin, M. T., Gramatica, P., Jaworska, J. S., Kahn, S., Klopman, G., Marchant, C. A., Myatt, G., Nikolova-Jeliazkova, N., Patlewicz, G. Y., Perkins, R., Roberts, D. W., Schultz, T. W., Stanton, D. T., Van De Sandt, J. J. M., Tong, W., Veith, G., & Yang C (2005). Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships: The report and recommendations of ecvam workshop 52. *Alternatives to Laboratory Animals*, 33(2), 155-173.

- Nguyen, A. T., O'Neill, M. J., Watts, A. M., Robson, C. L., Lamont, I. L., Wilks, A., & Oglesby-Sherrouse, A. G. (2014). Adaptation of iron homeostasis pathways by a *Pseudomonas aeruginosa* pyoverdine mutant in the cystic fibrosis lung. *Journal of Bacteriology*, 196(12), 2265-2276.
- Ochsner, U. A., Johnson, Z., & Vasil, M. L. (2000). Genetics and regulation of two distinct haem-uptake systems, phu and has, in *Pseudomonas aeruginosa*. The GenBank accession numbers for the sequences reported in this paper are AF055999, AF127222, and AF127223. *Microbiology*, 146(1), 185-198.
- OECD (2007). Guidance Document on the Validation of (Quantitative) Structure–Activity Relationships [(Q)SAR] Models, Organisation for Economic Co-Operation and Development, Paris, France.
- Pastor, M., Cruciani, G., & Clementi, S. (1997). Smart region definition: A new way to improve the predictive ability and interpretability of three-dimensional quantitative structure–activity relationships. *Journal of Medicinal Chemistry*, 40(10), 1455-1464.
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L., & Schulten, K. (2005). Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*, 26(16), 1781-1802.
- Podkalicka, P., Mucha, O., Józkwicz, A., Dulak, J., & Łoboda, A. (2018). Heme oxygenase inhibition in cancers: possible tools and targets. *Contemporary Oncology*, 22(1A), 23-32.
- Ponka, P. (1999). Cell biology of heme. *The American Journal of the Medical Sciences*, 318(4), 241-256.
- Ray, S. (2012). QSAR study of Salicylanilides derivatives against various Mycobacterial strain at AM1 semi empirical level. *International Research Journal of Pharmacy*, 3(4), 251-254.
- Tosco, P., & Balle, T. (2011). Open3DQSAR: a new open-source software aimed at high-throughput chemometric analysis of molecular interaction fields. *Journal of Molecular Modeling*, 17, 201-208.
- Tosco, P., Balle, T., & Shiri, F. (2011). Open3DALIGN: an open-source software aimed at unsupervised ligand alignment. *Journal of Computer-Aided Molecular Design*, 25(8), 777-783.
- Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of computational chemistry*, 31(2), 455-461.
- Wavefunction, Inc. (2015). Chemistry at your fingertips. Retrieved February. 15, 2020 from <http://www.wavefun.com>.
- Yap, C. W. (2011). PaDEL-descriptor: An open source software to calculate molecular descriptors and fingerprints. *Journal of Computational Chemistry*, 32(7), 1466-1474.