Insilico design of highly potent anti-salmonella typhi drug candidates from schiff bases

Ameji Philip John[1], Adamu Uzairu[1], Gideon Adamu Shallangwa[1]

[1]Department of Chemistry, Ahmadu Bello University Zaria, Nigeria.

Corresponding author: Ameji Philip John
Address: Sabon Gari, 810222, Zaria, Zaria, Kaduna Nigeria;
Telephone: 07036492681; E-mail: philo4real55@gmail.com

Abstract

Aim: To develop good and rational Quantitative Structure Activity Relationship (QSAR) mathematical models that can predict to a significant accuracy the anti-*Salmonella typhi* minimum inhibitory concentration (MIC) of Schiff bases and using the information provided by the model to design highly potent novel potential drug candidates against the pathogen.
Methods: A set of 35 Schiff bases with their antibacterial activities in terms of minimum inhibitory concentration (MIC) against *Salmonella typhi* were selected for 0D, 1D, 2D and 3D quantitative structure activity relationship (QSAR) analysis by means of Density Functional Theory (DFT) using the Becke's three-parameter hybrid functional (B3LYP) and 6-31G* basis set. The computed descriptors were correlated with their experimental MIC. Genetic function approximation (GFA) based Multi-linear regression analysis (MLR) was used to derive the most statistically significant QSAR model.
Results: Among the obtained QSAR models, the most statistically significant one was a tetra-parametric linear equation with the squared correlation coefficient (R^2) value of 0.8026, adjusted squared correlation coefficient (R^2adj) value of 0.7587 and Leave one out (LOO) cross validation coefficient (Q^2) value of 0.7453. An external set was used for confirming the predictive power of the model, its $R^2pred.$= 0.7000, LOF score = 0.0635, F-value = 18.2945, $\leq$0.05. Four new drug candidates against *Salmonella typhi* were designed.
Conclusion: The QSAR results reveal the predominance of electro-topological descriptor, SHsOH in influencing the anti-*Salmonella typhi* bioactivity of the studied Schiff bases. The QSAR model was found to be stable and reliable.

Keywords: descriptor, GFA, MIC, QSAR, validation.

Introduction

Enteric fever is a pervasive disease in the tropical and subtropical regions of the world. It is caused by the bacterium *Salmonella typhi* (*S. typhi*). It is a systemic disease and occurs by ingestion of infected food or water usually from a feco-oral source (1). It is a global problem and widely prevalent in the tropical and subtropical regions of the world with an estimated 12-33 million cases and at least 250,000 deaths occurring annually, constituting a serious source of morbidities and mortalities in these regions (2).

S. typhi bacterium has gained resistance to antibiotics like ampicillin, ceftriaxone and cotrimoxazole, besides developing resistance to previously efficacious drugs like ciprofloxacin. The current worldwide increase in antimicrobial resistance by *S. typhi* and lack in the development of new antibiotics have serious public health and economic implications. The reason for increasing resistance is multifactorial, but the main cause is the high level of inappropriate antibiotic usage (2). Thus, the need to urgently search for new antibiotics that will arrest this dangerous trend of multi-drug resistance by this organism (*S. typhi*) has become a sine qua non.

The enormous inhibitory activity of Schiff bases against the growth of *S. typhi* (3-5) has made them potential drug candidates in man's quest to curb the dangerous trend of multi-drug resistance posed by this pathogenic micro-organism.

The knowledge of the dominant structural features influencing the observed anti-bacteria activities of these molecules is an essential step towards the discovery and development of novel anti-*S. typhi* Schiff bases as this will aid the optimization of the existing structures to obtain candidates with higher bioactivity against the organism.

An essential step of drug discovery involves the modification of the hits in order to improve the biological properties of the compounds (6). The application of quantitative structure activity relationship (QSAR) methodologies has potential to decrease substantially the time and effort required to discover new medicines or to improve current ones in terms of their efficacy by avoiding the conventional trial and error approach employed in the discovery and development of novel medicines by avoiding leads unlikely to be successful. Thus, promoting green and greener chemistry by reducing waste and increasing efficiency. Moreover, such *in silico* methods in addition to theoretically helping to modify the compounds to exhibit the most potency, also help to obtain drug candidates with most selectivity, best pharmacokinetics and least toxicity.

QSAR establishes the mathematical relationship between physico-chemical properties or biological activities of interest and measurable or computable parameters called molecular descriptors (7). The fundamental principle underlying QSAR is that the difference in structural properties is responsible for the variations in biological activities of the compounds. It assumes that the potency of a certain biological activity exerted by a series of congeneric compounds is a function of various physico-chemical parameters of the compounds. Once statistical analysis shows that certain physico-chemical properties are favorable to the concerned activity, the concerned activity can be optimized by choosing such substituents which would enhance such physico-chemical properties (8).

Though, similar QSAR work has been carried out on the data set under investigation by Adawara et al (9) but with semi-empirical method of molecular optimization which has been proven to provide less accurate result compared to DFT deployed in this study. Likewise, the model generated by the aforementioned authors met the minimum criteria for generally acceptable QSAR models except that the model was not externally validated, making it difficult to estimate the true predictive power of the model.

The aim of the present study is to build robust, rational, and predictive DFT based QSAR models for *Salmonella typhi* inhibitory activity of Schiff

bases using Genetic Function Approximation and using the information obtained from the most statistically significant model to design highly potent novel anti-*Salmonella typhi* Schiff bases.

Methods

Data sources

A novel series of 35 Schiff bases with potent anti-*S. typhi* activities in vitro were collected from literature (3-5,10,11) to perform this study. The general molecular structures of the studied compounds, toxicity values and their CAS registry number are shown in Table 1. The anti-*S. typhi* activities expressed as μM were converted into the corresponding pMIC (pMIC = logMIC) values and used as dependent variables in the QSAR study.

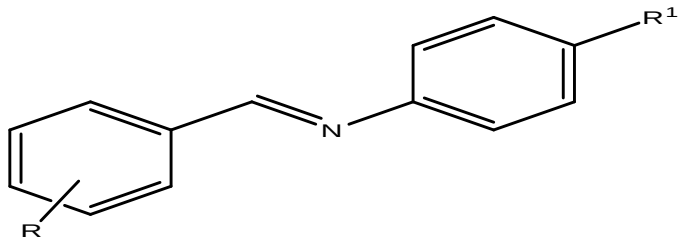
Molecular descriptors

The "Cascade method" of molecular optimization was used for this study (12). Here, the molecules were first pre-optimized with the molecular mechanics procedure included in Spartan'14 V1.1.0 software and the resulting geometries were further refined by means of Semi-empirical (pm3) and DFT (B3LYP) using 6-31G* basis set. The molecular descriptors used in the QSAR modelling were calculated using Padel descriptor tool kit, Spartan 14 software and ChmBio3D Pro 12.0.1V software.

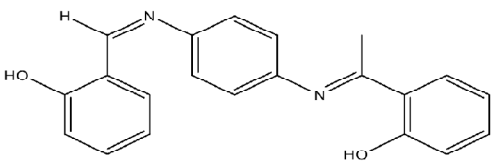
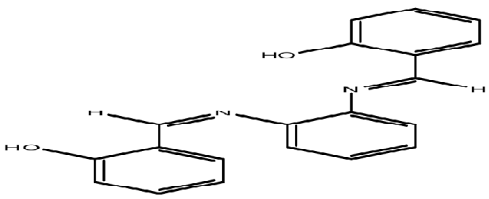
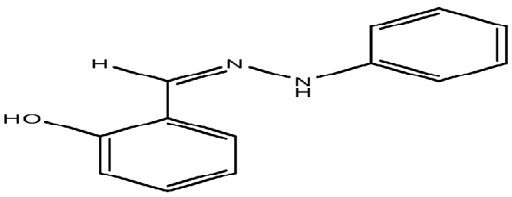
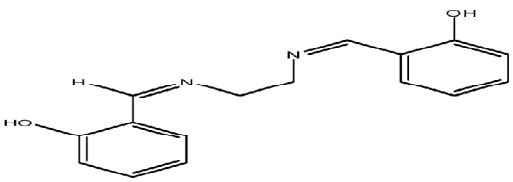
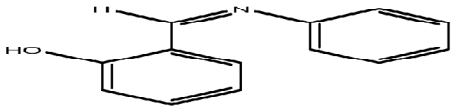
Statistical analysis

In this stage, the correlations between pMIC of the compounds and the calculated descriptors were obtained via correlation analysis using the Microsoft excel package in Microsoft office 2013. Pearson's correlation matrix was used as a qualitative model, in order to select the suitable descriptors for each regression analysis. The selected descriptors were subjected to regression analysis with the pMIC as the dependent variable and the selected descriptors as the independent variables using Genetic function approximation (GFA) method in Material studio software. It is a distinctive characteristic of GFA that it could create a population of models rather than a single model. GFA algorithm, selecting the basic functions genetically, developed better models than those made using stepwise regression methods. And then, the models were estimated using the "lack of fit" (LOF), which was measured using a slight variation of the original Friedman formula, so that best model received the best fitness score (13-16).

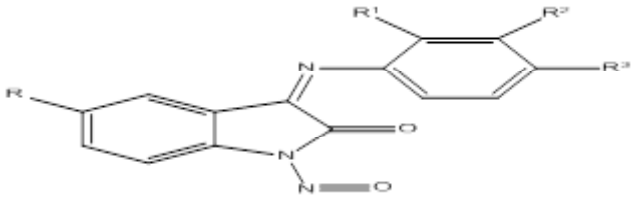
Table 1. Anti-*Salmonella typhi* activity of the compounds (MIC μg/ml and pMIC)



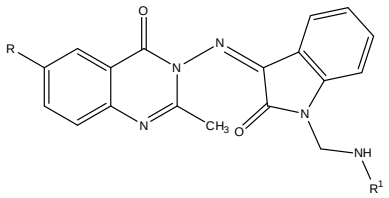
Parent structure for compound 1-8.

Cpd.	R	pMIC	Pred. pMIC	Residual
1	3-OCH_3	1.26	1.71	-0.45
2*	3,4-OCH_3	1.20	1.66	-0.46
3	3,4,5-OCH_3	1.60	1.59	0.01
4	3-OCH_3, 4-OH	0.90	0.90	0.00
5	4-F	1.78	1.72	0.06
6*	4-Cl	1.48	1.72	-0.24
7	4-Br	1.81	1.72	0.09
8	4-I	2.08	1.72	0.36
9*		1.70	2.49	-0.79
10		2.40	2.40	0.00
11		2.30	2.28	0.02
12		2.00	2.00	0.00
13		2.30	2.31	-0.01

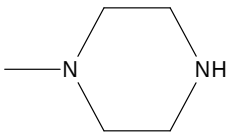
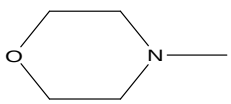
Parent structure for compound 14-29



Cpd.	R	R^1	R^2	R^3	pMIC	Pred. pMIC	residual
14*	H	H	H	Cl	1.34	1.49	-0.15
15	H	H	H	Br	1.28	1.50	-0.22
16*	H	H	H	F	1.34	1.49	-0.15
17	H	H	Cl	F	1.90	1.46	0.44
18	H	H	H	CH_3	1.26	1.50	-0.24
19	H	H	H	OCH_3	1.62	1.50	0.12
20*	H	H	H	NO_2	1.36	1.48	-0.12
21	H	NO_2	H	NO_2	1.28	1.38	-0.10
22*	Br	H	H	Cl	1.58	1.46	0.12
23	Br	H	H	Br	1.32	1.46	-0.14
24*	Br	H	H	F	1.36	1.46	-0.10
25	Br	H	Cl	F	1.32	1.42	-0.10
26*	Br	H	H	CH_3	1.38	1.46	-0.08
27	Br	H	H	OCH_3	1.62	1.47	0.15
28*	Br	H	H	NO_2	1.30	1.45	-0.15
29	Br	NO_2	H	NO_2	1.32	1.35	-0.03



Parent structure for compound 30-35

Cpd	R	R^1	pMIC	Pred. MIC	Residual
30*	H	$N(CH_3)_2$	1.28	1.27	-0.01
31	H	$N(C_2H_5)_2$	1.18	1.15	-0.09
32*	H	$N(C_6H_5)_2$	1.20	1.15	-0.05
33	H	$N(C_6H_{11})_2$	1.23	1.15	0.08
34	H		1.30	1.33	-0.03
35	H		1.36	1.27	0.09

*Test set compound

Model validation and evaluation

To validate the rationality and reliability of QSAR models, the total set of these analogues was randomly divided into a training set (23 compounds) to generate the QSAR models and a test set (11 compounds) to evaluate the predictive power of the resulting model externally. The well-known scheme of "leave-one-out" (LOO) cross-validation was adopted. Usually, the square of cross-validation coefficient (q^2), which was used as a criterion to evaluate both the robustness and the predictive ability of the model, should be > 0.5 for a reliable model (17). Other validation parameters deployed in this study include the square of the correlation coefficient, R^2 (threshold of $\geq$0.6). It describes the fraction of the total variation attributed to the model (18). Also, for the alternative hypothesis that the magnitude of the observed *S. typhi* inhibitory activities of Schiff bases is a direct function of the descriptor of the total chemical structure of the compounds to take preference over the null hypothesis which states otherwise, the P-value of the model at 95% confidence level was determined. Moreover, an external validation is also crucial to obtain QSAR models with more reliable predictive abilities (18). The optimum QSAR model was externally validated using the test set of 11 molecules with proven *S. typhi* inhibitory activity by using the optimum model (model 1) to predict the MIC of the compounds.

Results and Discussion

Models 1, 2, 3 give the best three Genetic Function Approximation derived QSAR models for pMIC of anti- *S. typhi* molecules. Based on the model with the least LOF, Model 1 was chosen as the best model for predicting the pMIC of anti- *S. typhi* Schiff bases. Likewise, its validation parameters are in good agreement with the standard validation metrics for a robust QSAR model proposed by Ravinchandra et al. (18).

Model 1:

$$pMIC = -0.0423SPC - 5.381SHsOH + 0.104maxsOH + 5.0MLFER + 1.814$$

LOF = 0.06353, R^2 = 0.802584, R^2_{adj} = 0.758714, Q^2_{LOO} = 0.745292, F-value = 18.294515, P value at 95% confidence level < 0.0001, Min expt. error for non-significant LOF (95%) = 0.16089000.

Model 2:

$$pMIC = -0.0422SPC - 5.392SHsOH + 0.104minssOH + 5.019MLFER - A + 1.813$$

LOF = 0.063532, R^2 = 0.802576, R^2_{adj} = 0.758704, Q^2_{LOO} = 0.745608, F-value = 18.293617, Min expt. error for non-significant LOF (95%) = 0.16089300.

Model 3:

$$pMIC = -0.042SPC - 5.454SHsOH + 1.618maxHsOH + 5.177MLFER + 1.81$$

LOF = 0.06373800, R^2 = 0.801937, R^2_{adj} = 0.757923, Q^2_{LOO} = 0.732588, F-value = 18.220071, Min expt. error for non-significant LOF (95%) for model 1= 0.16115300.

Where ***MinsOH*** is Minimum atom-type electro-topological state (E-State: -OH) descriptor, ***SPC*** the Simple path cluster, ***MaxsOH*** the Maximum atom-type E-State: -OH, ***MLFER*** the Overall or summation solute hydrogen bond acidity, ***MaxHsOH*** the Maximum atom-type E-State: -OH, and ***SHsOH*** the Sum of atom-type E-State: -OHmolecular descriptors.

The predictability of model 1 is evidenced by the low residual values for both test and training set compounds presented in Table 1 which gives the comparison of observed and predicted pMIC of the anti- *S. typhi* Schiff bases. Also, the high linearity (R^2 = 0.8026) of the plot of predicted pMIC against observed pMIC shown in Figure 1 indicates that the model is well trained and it predicts well the pMIC of the compounds.

To ascertain whether there exists a systematic error in the model development, the residual pMIC was plotted against observed pMIC (Figure 2). The propagation of residuals on both sides of zero indicates that there was no systemic error in model development (19).

The P-value of the optimization model at 95% confidence level has α value of < 0.0001 far less than the minimum threshold of < 0.05 at this confidence level. This reveals that the alternative hypothesis that the magnitude of the observed anti-*S. typhi* inhibitory activity of the Schiff bases is a direct function of the descriptors of their total chemical structures takes preference over the null hypothesis which states otherwise.

The result of the QSAR modelling hinted the predominance of an electro-topological descriptor, SHsOH over other descriptors in the model in influencing the anti-*Salmonella typhi* bioactivity of the studied Schiff bases owing to its relatively high numerical percentage contribution of 51.11% to the model. The percentage contribution of other descriptors include 0.40% (SPC), 0.99 (MaxsOH), 47.50% (MLFER).

The negative coefficient of SHsOH descriptor in the model implies that the minimum inhibitory concentration (MIC) of the Schiff bases that fall within the model's applicability domain increases with decrease in the value of this descriptor. Since activity varies inversely with MIC, it can be inferred that the *S. typhi* inhibitory potential of these molecules is in direct proportion with the value of this descriptor in them. This is in consonance with the result of QSAR modelling of anti-bacterial activities of simple 2`-hydroxy chalcones by Basic et al. (20) in which they found that increase in value of SHsOH descriptor in the chalcone derivatives also enhances their anti-bacterial activities.

Figure 1. Plot of actual pMIC against predicted pMIC

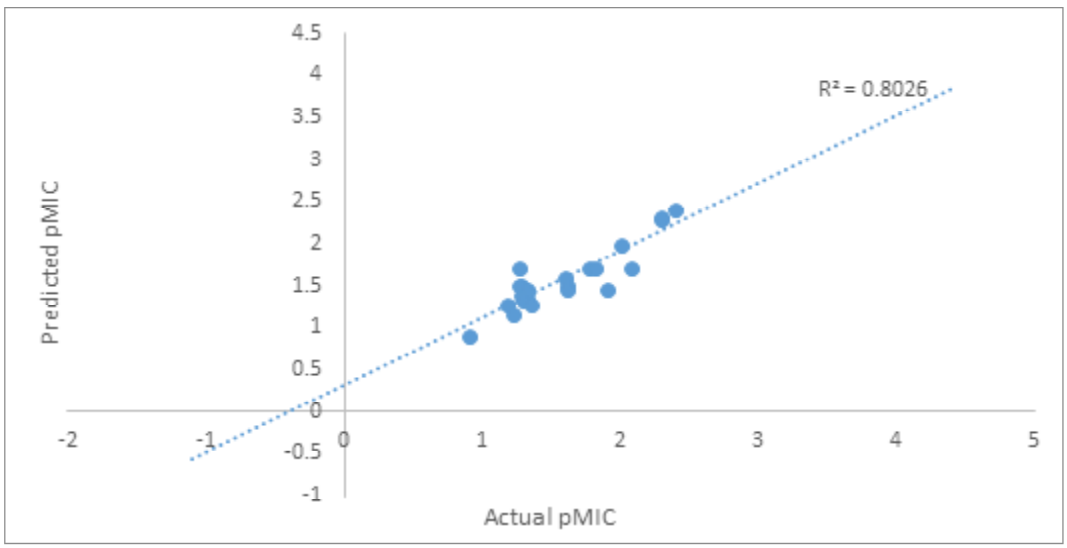
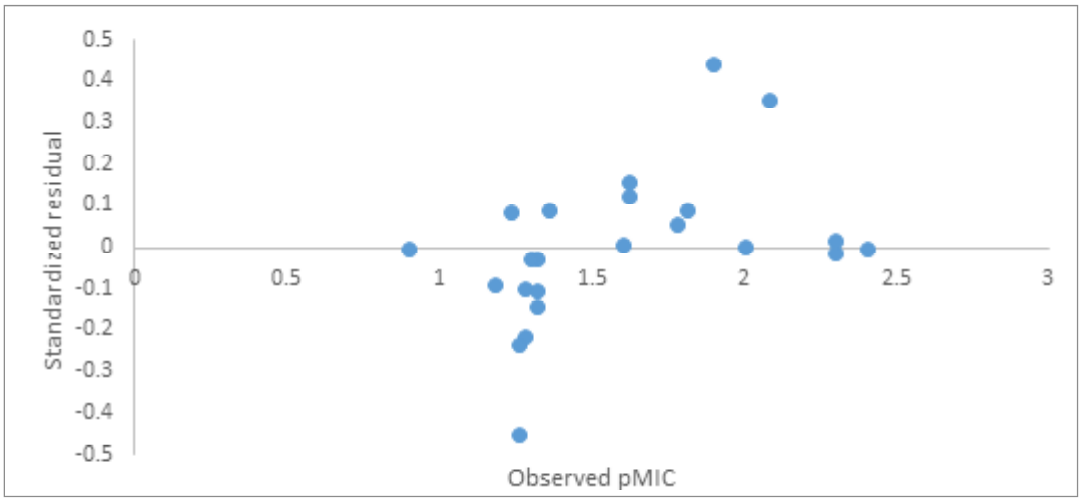


Figure 2. Residual plot of model 1

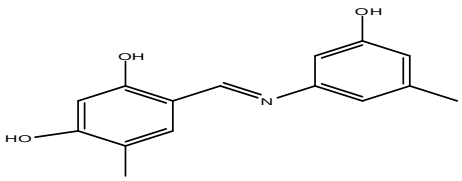
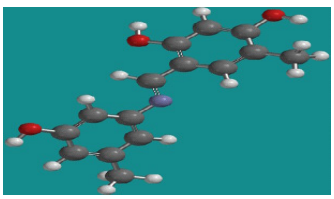


SHsOH is a descriptor of hydrogen-bond donor capabilities of molecules (20). The high contribution of the descriptor to model 1 may be as a result of the strong influence of the acidity of the functional groups of the Schiff bases. Thus, attaching functional groups (such as hydroxyl group) with hydrogen-bond donor capabilities may enhance the *S. typhi* inhibitory activities of the Schiff bases.

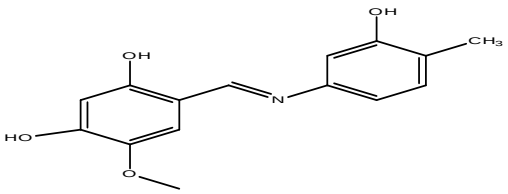
Design of Potent Anti-S. typhi Schiff bases

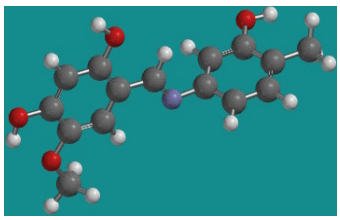
Based on the information from the optimum QSAR model (Model 1), molecules a, b, c and d (Table 2) were designed insilico as potential inhibitor of the growth of *S. typhi* strain. The molecules exhibit significant inhibitory activity against the pathogenic organism. The pMIC of the newly designed molecules presented in the Table indicates that compounds b and c exhibit high activity even more than the compound with the highest inhibitory activity against *S. typhi* (C4) in the data set.

Table 2. Optimized structures of the newly designed anti-S.typhi Schiff bases.

S/n	Molecule	Optimized Structure	Pred. pMIC
a.	 (E)-4-(((3-hydroxy-5-methylphenyl)imino)methyl)-6-methylbenzene-1,3-diol		1.64

b.

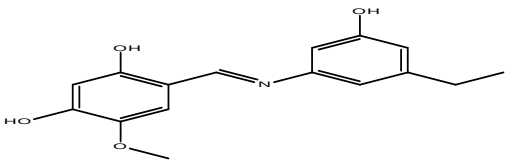


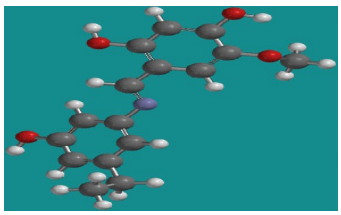


0.31

(E)-4-(((3-hydroxy-4-methylphenyl)imino)methyl)-6-methoxybenzene-1,3-diol

c.





0.45

(E)-4-(((3-ethyl-5-hydroxyphenyl)imino)methyl)-6-methoxybenzene-1,3-diol

Conclusion

The generated QSAR models, performed to explore the structural requirements controlling the observed *Salmonella typhi* inhibitory properties of some selected Schiff bases, hinted that the MIC of the molecules was predominantly influenced by an electro-topological descriptor (SHsOH), a descriptor of hydrogen bond donor capability of the molecules. The robustness and applicability of QSAR equation has been established by internal and external validation techniques. Based on the information obtained from the optimum QSAR model four new compounds with enormous inhibitory activities against *S. typhi* were designed and their bioactivities were further predicted using the optimum QSAR. These compounds are therefore recommended for further studies as far as search for newer antibiotics that could curb the dangerous trend of multi-drug resistance by *S.typhi* is concerned.

Conflicts of interest: None declared.

References

1. Sehra D, Sehra S, Relia P, Sehra ST. An Altered Drug Resistance Pattern in Salmonella Typhi. Am J Infec Dis Microb 2013;1:84-5.
2. Chungathu R, Jayavardhana A. Current pattern of Salmonella typhi antimicrobial susceptibility in the era of antibiotic abuse. IJBAMR 2015;5:400-4.
3. Sharif HM, Ahmed D, Mir H. Antimicrobial salicylaldehyde Schiff bases: Synthesis, characterization and evaluation. Pak J Pharm Sci 2015;28:449-55.
4. Kumar BS, Parthiban KG. Synthesis of Schiff bases of some novel n-nitrosoisatin derivatives as potential antimicrobial agents. Asian J Pharm Clin Res 2011;4:137-40.

5. Rajasekaran S, Rao G. Synthesis, antibacterial and antioxidant activity of some 2, 3-susbtituted quinazolin-4(3H)-ones. Pharm Lett 2012;4:470-4.

6. Sprous DG, Zhang J, Zhang L, Wang Z, Tepper MA. Kinase inhibitor recognition by use of a multivariable QSAR model. J Mol Graph Model 2006;24:278-95.

7. Rathod AK. Antifungal and Antibacterial activities of Imidazole pyrimidines derivatives and their QSAR Studies under Conventional and Microwave-assisted. Int J Pharm Tech Res 2011;3:1942-51.

8. Jhanwar B, Sharma V, Singla RK, Shrivastava B. QSAR-Hansch analysis and related approaches in drug design. PhOL 2011;1:306-44.

9. Ndaghiya AS, Ogadimma AI, Sani S. Modelling of some Schiff bases as anti-Salmonella typhi drugs: A QSAR approach. JCMMD 2015;5:147-57.

10. Hansch C, Fujita TP. A Method for the correlation of biological activity and chemical structure. J Am Chem Soc 1964;84:1616-26.

11. Sahu SK, Afzal Azam MD, Banerjee M, Acharrya S, Behera CC, Si S. Synthesis, characterization and biological activity of 2-methyl-3-aminoquinazolin-4(3H)-ones Schiff bases. J Braz Chem Soc 2008;19:5-10.

12. Hehre WJ. Practical Strategies for Electronic Structure Calculations, Wavefunction, Inc., Irvine, CA; 1995.

13. Nantasenamat C, Isarankura-Na-Ayudhya C, Naenna T, Prachayasittikul V. A Practical Overview of Quantitative Structure-Activity Relationship. EXCLI J 2009;8:74-88.

14. Wu W, Zhang C, Lin W, Chen Q, Guo X, Qian Y. Quantitative structure-property relationship (qspr) modeling of drug-loaded polymeric micelles via genetic function approximation. PLoS One 2015;10:3-16.

15. Friedman JF. Multivariate Adaptive Regression Splines, Technical Report No. 102, Stanford University; 1990.

16. Khaled KF, Abdel-Shafi NS. Quantitative Structure and Activity Relationship Modeling Study of Corrosion Inhibitors: Genetic Function Approximation and Molecular Dynamics Simulation Methods. Int J Electrochem Sci 2011;6:4077-94.

17. Tropsha A. Best Practices for QSAR Model Development, Validation, and Exploitation. Mol Inf 2010;29:476-88.

18. Veerasamy R, Rajak H, Jain A, Sivadasan S, Varghese CP, Agrawal RK. Validation of QSAR Models - Strategies and Importance. IJDDD 2011;3:511-9.

19. Jalali-Heravi M, Kyani A. Use of Computer-Assisted Methods for the Modeling of the Retention Time of a Variety of Volatile Organic Compounds: A PCA-MLR-ANN Approach. J Chem Inf Comput Sci 2004;44:1328-35.

20. Basic J, Kalinic M, Ivkovic B, Eric S, Milenkovic M, Vladimirov S, et al. Synthesis, QSAR analysis and mechanism of anti-bacterial activity of simple 2'-hydroxy chalcones. Dig J Nanomater Biostruct 2014;9:1537-46.