European Journal of Medicinal Chemistry 44 (2009) 4128–4135

Contents lists available at ScienceDirect

European Journal of Medicinal Chemistry

journal homepage: http://www.elsevier.com/locate/ejmech

Original article

Quantitative structure activity relationship (QSAR) of piperine analogs for bacterial NorA efflux pump inhibitors

Amit Nargotra, Sujata Sharma, Jawahir Lal Koul, Pyare Lal Sangwan, Inshad Ali Khan, Ashwani Kumar, Subhash Chander Taneja, Surrinder Koul*

Indian Institute of Integrative Medicine (Formerly known as Regional Research Laboratory), Canal Road, Jammu – 180001, J&K, India

ARTICLE INFO

Article history:
Received 25 February 2009
Received in revised form
4 May 2009
Accepted 7 May 2009
Available online 18 May 2009

Keywords:
QSAR
Efflux pump inhibitors
Potentiators
Piperine analogs
NorA

ABSTRACT

Quantitative structure activity relationship (QSAR) analysis of piperine analogs as inhibitors of efflux pump NorA from *Staphylococcus aureus* has been performed in order to obtain a highly accurate model enabling prediction of inhibition of *S. aureus* NorA of new chemical entities from natural sources as well as synthetic ones. Algorithm based on genetic function approximation method of variable selection in Cerius2 was used to generate the model. Among several types of descriptors viz., topological, spatial, thermodynamic, information content and E-state indices that were considered in generating the QSAR model, three descriptors such as partial negative surface area of the compounds, area of the molecular shadow in the XZ plane and heat of formation of the molecules resulted in a statistically significant model with $r^2 = 0.962$ and cross-validation parameter $q^2 = 0.917$. The validation of the QSAR models was done by cross-validation, leave-25%-out and external test set prediction. The theoretical approach indicates that the increase in the exposed partial negative surface area increases the inhibitory activity of the compound against NorA whereas the area of the molecular shadow in the XZ plane is inversely proportional to the inhibitory activity. This model also explains the relationship of the heat of formation of the compound with the inhibitory activity. The model is not only able to predict the activity of new compounds but also explains the important regions in the molecules in quantitative manner.

1. Introduction

In recent years the emergence of resistance to the antibacterial agents has become serious public health problem [1] and the community acquired infection due to Gram +ve bacteria is one of the major concerns confronted by mankind. In particular, methicillin-resistant *Staphylococcus aureus* (MRSA) is one of the most frequent nosocomial pathogens prevalent in developed countries [2–4]. Among the Gram +ve bacteria, *S. aureus* is responsible for a large number of deaths worldwide [5] (e.g. 19,000 deaths occurred in 2005 in USA and 1200 deaths in UK). For bacteria to become multidrug resistant, it must acquire multiple mechanisms which may not be a requirement to become resistant for one single or a specific class of compounds. The multidrug resistance (MDR) produced by bacteria is due to its antiporter type efflux proteins which derive their energy through ATP hydrolysis or sodium or proton motive force and are responsible for extrusion of drugs/substrates. The removal of the drug from the cell (by efflux pump) leads to the lower levels of drug concentrations at the active site. The primary or main role of efflux pump is to remove toxins that are encountered in the environment or produced during metabolism and this phenomenon is responsible for the bacteria to able to survive in hostile environment such as that provided by the antibiotics. Efflux of drugs from Gram +ve bacteria is mediated by a single cytoplasmic transporter of major facilitator (MF), small multidrug resistance (SMR) or ATP-binding cassette (ABC-type) family. Development of new drugs which may be capable of escaping or bypassing the drug efflux (i.e. are poor pump substrates [6]) is one of the ways to overcome the drug resistance but this approach demands high cost and long duration. Alternatively, the drug activity might be restored by development of potentiator(s) of drugs that may bind to the same site as done by drugs or substrates and inhibit transport by a competitive or noncompetitive mechanism [7] which can perform dual function of increasing potential of the drug as well as induce the emergence of target based resistance [8] thereby expanding the spectrum of the anti-infectives [9]. This combination approach of drug plus potentiator has already been well established for amoxicillin and anti-infective resistance

* Corresponding author. Tel.: +91 191 2569000/01/02/03/04/05/06/07/08/09/10x269; fax: +91 191 2569333.
E-mail addresses: amitnargotra@yahoo.com (A. Nargotra), sujatasharma02@rediffmail.com (S. Sharma), inshad@yahoo.com (I.A. Khan), sc_taneja@yahoo.co.in (S.C. Taneja), surrinderkoul@gmail.com (S. Koul).

0223-5234/$ – see front matter © 2009 Elsevier Masson SAS. All rights reserved.
doi:10.1016/j.ejmech.2009.05.004

inhibitor clavulanic acid [10]. Efforts are on for the therapeutic utility of chemical entities that inhibit bacterial efflux pumps as is evident by the human clinical trials of the aerosolized formulation of ciprofloxacin and EPI [6] as well as in the development of a formulation of ceftazidine (cephalosporin) and beta-lactam inhibitor NXL104 [11]. The most studied chromosomal efflux pump of *S. aureus* is Nor A, a transporter belonging to the major facilitator superfamily (MFS) and extrudes multiple structurally unrelated substrates and some of the other bacterial efflux pumps studied in detail include Tet-K and MRSA transporters [12].

One of the essential requirements for the development of a combination drug involving EPI should be such that the EPI concentration used *in vivo* should match concentrations of the *in vitro* derived EPI so that c_{max} of the drug is achieved effectively (otherwise the EPIs developed shall have academic importance due to limited preclinical and clinical capabilities). But this activity (for robust EPI) demands screening of a large number of chemical compounds by *in vitro* studies involving labor and cost besides environmental hazard. Alternatively, *in silico* approach can be used as a strategy for designing rational and robust EPIs. Quantitative structure activity relationship (QSAR) is one such methodology to design a rational EPI that meets the above-said requirement and shall involve much lower effort, time and lesser issues of environmental pollution. A QSAR equation correlates the biological activity to a wide variety of physical or chemical parameters [13–16] and is one of the most important methods in chemometrics where biological activities are quantitatively related to physical and chemical characteristics [17,18]. There are many examples available in literature in which QSAR models have been used successfully for the screening of compounds for the biological activity [19–22].

In our recent studies, we have demonstrated the efflux pump inhibitory activity of piperine, its mimics and analogs when used in combination with anti-infectives [23–26]. Based on the data generated from the wet lab studies on piperine and its analogs including their structure activity relationship, we report in this communication a QSAR model developed for determining the efflux pump inhibitory activity of these piperine analogs using Cerius2 software (Accelrys Inc.). The model developed in the present study is the first of its kind for efflux pump inhibitors of NorA overexpressing *S. aureus* strain (*S. aureus* 1199B).

2. Experimental

2.1. In vitro evaluation of EPIs

The EPIs were tested in combination with ciprofloxacin by a broth checkerboard method in a 96-well microdilution plate [27]. The detailed methodology is described in our earlier paper [23]. Briefly the MIC of ciprofloxacin was determined against *S. aureus* 1199B (Nor A overexpressing) in Muller–Hinton Broth in the presence of EPIs. Seven serial dilutions of EPIs (50–0.8 μg/ml) were tested in combination with ten serial dilutions of ciprofloxacin (64–0.12 μg/ml). The plates were incubated at 37 °C for 24 h. The plates were read visually after incubation at 37 °C for 24 h. The lowest concentration well in each row showing no turbidity was recorded as MIC (Tables 1 and 2).

2.2. Efflux mechanism studies

The mechanism of inhibition of bacterial efflux pumps by these EPIs was determined by the measure of the levels of ethidium bromide accumulation and efflux in *S. aureus* 1199B using the method described by Brenwald and co-workers [28].

Table 1
MIC (μg/ml) of ciprofloxacin against *S. aureus* in combination with synthesized molecules (active).

Compound no.	Conc. (μg/ml)								Potentiation factor (PF)	log(PF)
	50	25	12.5	6.25	3.12	1.56	0.8	0.0		
1	4	4	8	8	8	8	8	8	60	1.78
2	4	4	8	8	8	8	8	8	60	1.78
3	4	4	8	8	8	8	8	8	60	1.78
4	4	4	8	8	8	8	8	8	60	1.78
5	4	4	8	8	8	8	8	8	60	1.78
6	4	4	8	8	8	8	8	8	60	1.78
7	2	4	4	8	8	8	8	8	140	2.15
8	2	4	4	8	8	8	8	8	140	2.15
9	4	4	8	8	8	8	8	8	60	1.78
10	4	4	8	8	8	8	8	8	60	1.78
11	4	4	8	8	8	8	8	8	60	1.78
12	4	4	4	4	8	8	8	8	200	2.30
13	1	2	2	2	4	4	8	8	660	2.82
14	4	4	8	8	8	8	8	8	60	1.78
15	2	2	4	4	8	8	8	8	260	2.42
16	2	2	2	4	8	8	8	8	320	2.51
17	2	2	4	4	8	8	8	8	260	2.42
18	4	4	4	4	8	8	8	8	200	2.30
19	4	4	4	4	4	8	8	8	300	2.48
20	2	2	4	4	4	8	8	8	360	2.56
21	4	4	4	4	4	8	8	8	300	2.48
22	4	8	8	8	8	8	8	8	20	1.30
23	4	8	8	8	8	8	8	8	20	1.30
24	4	8	8	8	8	8	8	8	20	1.30
25	4	4	8	8	8	8	8	8	60	1.78

2.3. In silico studies

2.3.1. Data set

The potentiation of activity of ciprofloxacin in presence of piperine analogs based efflux pump inhibitors was reflected in the reduced MIC of combination compared to that of ciprofloxacin alone (taken from the experimental results of *in vitro* studies above). The reduction in the MIC values of ciprofloxacin was used to measure the extent of potentiation of the molecules and was expressed in terms of potentiation factor (PF) of each compound as described below:

To calculate the potentiation factor (PF) of the compounds (efflux pump inhibitors – EPI), we gave weightages to each column of the MIC table (Tables 1 and 2) and higher weightage was given to a column where the compound potentiated the ciprofloxacin activity even at a lower conc. of say 1.56 μg/mg. It is important because of the fact that if a compound can potentiate ciprofloxacin activity at a lower concentration than any other compound, its PF has to be higher, moreover, we have considered the fact that if a compound inhibits the efflux pump (EP) and increases potentiation at double diluted concentration, i.e., if it increases the potency of ciprofloxacin at 25 μg/ml, and continues to do so at 12.5 μg/ml, the weightage for diluted concentration has to be more than the original and hence we have taken increased weightage for each

Table 2
MIC (μg/ml) of ciprofloxacin against *S. aureus* in presence of synthesized molecules (not active).

Compound no.	Conc. (μg/ml)								Potentiation factor (PF)	log(PF)
	50	25	12.5	6.25	3.12	1.56	0.8	0.0		
26	8	8	8	8	8	8	8	8	Not active	
27	8	8	8	8	8	8	8	8		
28	8	8	8	8	8	8	8	8		
29	8	8	8	8	8	8	8	8		
30	8	8	8	8	8	8	8	8		

column and finally summed up the result; as otherwise it would have led to an error and might have erroneously rated two compounds with varying potentiating activity as same. Hence, we needed to quantify the inhibitory activity of the compounds by some index value in such a manner that the difference of the activity is reflected in the index value. Hence, we calculated the potentiation factor (PF) of all the compounds. Each column of the MIC table (Tables 1 and 2) was given weightage in ascending order, starting with the base value of 10 from column 2 (compound concentration at 50 μg/ml) and adding 10 for each subsequent column, up to maximum of 70 for column 7 (compound concentration at 0.8 μg/ml). This value was multiplied by the number of fold decrease in MIC for that particular dilution and PF was determined by adding up the values from each column, e.g., for compound **13**,

$$\mathrm{PF} = (8 \times 10) + (4 \times 20) + (4 \times 30) + (4 \times 40) + (2 \times 50) + (2 \times 60) + (0 \times 70) = 660$$

The PF was finally converted to log of PF to get the linear relationship in the equation in order to develop the QSAR model.

From the data of Tables 1 and 2, we could derive Table 3 that gives the concentration of the compound (EPI) which shows specific decrease in MIC of ciprofloxacin at a given concentration of ciprofloxacin. Interestingly, we got the same order for the activity of the compounds, as obtained by using the PF method (Table 1). While Table 1 shows several levels of concentration of EPIs and corresponding fold reductions in the MIC of ciprofloxacin, Table 3 depicts only maximum effective concentrations of EPIs and the fold reduction in MIC of ciprofloxacin at that particular concentration of EPI only. For example compound no. **13** in Table 3 gives the single most effective concentration (50 μg/ml) showing 8-fold reduction in MIC of ciprofloxacin, whereas Table 1 shows three levels of potentiations of this compound viz., 8-fold (at 50 μg/ml), 4-fold (upto 6.25 μg/ml) and 2-fold (upto 1.56 μg/ml).

For the preparation of the QSAR model, the total set of 25 compounds was divided into 20 molecules in training set and 5 molecules in test set by arranging the molecules in ascending order of their PF and selecting every fifth molecule to be included in the training set.

The structures of the compounds are given in Tables 4–7. The potentiation factor (PF) and log of PF of all the active compounds are shown in Table 1 and the inactive ones are shown in Table 2.

Table 3
MEC (μg/ml) of EPI and fold reduction of ciprofloxacin (at a specified concentration) against *S. aureus*.

Compound	MIC of EPI (μg/ml)	MEC of EPI (μg/ml)	MIC of ciprofloxacin (μg/ml)		Reduction (*n*-fold) in MIC of ciprofloxacin
			Without EPI	With EPI	
1	>100	25	8	4	2
2	>100	25	8	4	2
3	>100	25	8	4	2
4	>100	25	8	4	2
5	>100	25	8	4	2
6	>100	25	8	4	2
7	>100	50	8	2	4
8	>100	50	8	2	4
9	>100	25	8	4	2
10	>100	25	8	4	2
11	>100	25	8	4	2
12	>100	6.25	8	4	2
13	>100	50	8	1	8
14	>100	25	8	4	2
15	>100	25	8	2	4
16	>100	12.5	8	2	4
17	>100	25	8	2	4
18	>100	6.25	8	4	2
19	>100	3.12	8	4	2
20	>100	3.12	8	4	2
21	>100	3.12	8	4	2
22	>100	50	8	4	2
23	>100	50	8	4	2
24	>100	50	8	4	2
25	>100	25	8	4	2
26	>100	50	8	8	0
27	>100	50	8	8	0
28	>100	50	8	8	0
29	>100	50	8	8	0
30	>100	50	8	8	0

2.3.2. Molecular modelling

The structures of the compounds taken in the data set were modeled using the 3D sketcher of the software Cerius2 from Accelrys Inc. All the molecules were energy minimized using the Cerius2.OFF module with default parameters of Smart Minimizer, which reads as follows:

Maximum displacement = 2 Å
ABNR termination criterion = 50 kcal/mol
QN termination criterion = 0.1 kcal/mol
Atom limit (asymmetric unit) for QN = 200
Atomic limit (total) for TN = 100

The convergence criteria adopted for the above method of minimization were standard with the following parameters:

Atom RMS force = 0.1 kcal/mol Å
Overall energy difference = 1×10^{-3} kcal/mol
Overall RMS displacement = 3×10^{-3} Å
Cell RMS stress = 0.1 GPa.

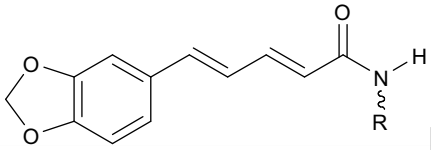
2.3.3. Descriptor calculation

E-state indices [29–31], electronic, information content [32], spatial, structural, thermodynamic and topological descriptors [33] were calculated using the Cerius2 4.10 software package. Topological indices are 2D descriptors based on graph theory concepts [34–36], widely used in QSPR and in QSAR studies [37] and these indices help to differentiate the molecules mostly according to their size, degree of branching, flexibility, and overall shape. Some of the descriptors included in the study are listed and described in Table 8.

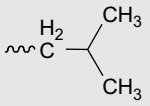
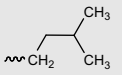
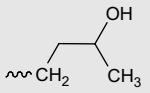
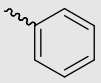
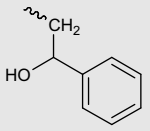
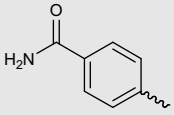
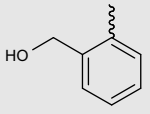
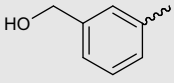
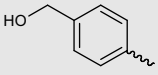
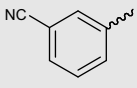
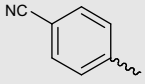
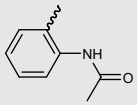
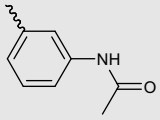
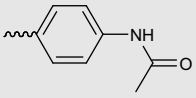
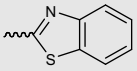
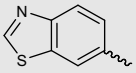
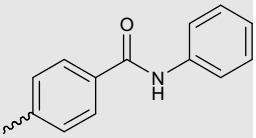
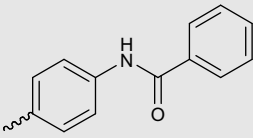
2.3.4. Regression analysis

The total number of descriptors calculated were about 250, and those that contained zero values for all the compounds were rejected. Based on the intercorrelation of the remaining descriptors, the highly correlated descriptors were grouped together and the only one descriptor among the group, that had the highest correlation with the biological activity, was selected and the rest ignored [38]. From descriptors that remained thus the selection of variables to obtain the QSAR models was carried out using genetic function approximation (GFA) method. GFA is genetics based method of variable selection, which combines Holland's genetic algorithm (GA) with Friedman's multivariate adaptive regression splines (MARS) [19,39]. The GFA method works in the following way: first of all a particular number of equations (set at 100 by default in the Cerius2 software) are generated randomly. Then pairs of "parent" equations are chosen randomly from this set of 100 equations and "crossover" operations are performed at random. The number of crossing over was set by default at 5000. The goodness of each progeny equation is assessed by Friedman's lack of fit (LOF) score, which is given by the following formula

Table 4
Structures of compounds of general formula 1 taken for present study.



Formula 1

Compound no.	R
1	
2	
5	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	

$$\mathrm{LOF} = \mathrm{LSE}/\{1 - (c + dp)/m\}^2$$

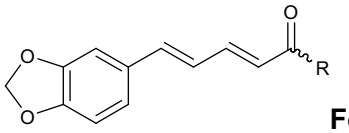
where LSE is the least-squares error, c is the number of basis functions in the model, d is smoothing parameter, p is the number of descriptors and m is the number of observations in the training set. The smoothing parameter, which controls the scoring bias between equations of different sizes, was set at default value of 1.0 and the new term was added with a probability of 50%. Only the linear equation terms were used for model building, which is set by default in the software. The best equation out of the 100 equations was taken based on the statistical parameters such as regression coefficient, adjusted regression coefficient, regression coefficient cross-validation and *F*-test values.

2.3.5. Validation test

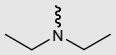
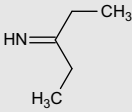
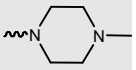
Further statistical significance of the relationship between the efflux pump inhibitory activity and chemical structure descriptors was obtained by randomization procedure. The test was done by repeatedly permuting the activity values of the data set and using the permuted values to generate QSAR models and then comparing the resulting scores with the score of the original QSAR model generated from non-randomized activity values. If the original QSAR model is statistically significant, its score should be significantly better than those from permuted data [40]. The randomized test was performed at 90%, 95%, 98% and 99% confidence intervals. The higher the confidence level, the more randomization tests are run. In this direction, nine trials were run at 90% confidence level, 19 trials at 95%, 49 trials at 98% and 99 trials at 99% confidence level.

To further check the intercorrelation of descriptors, variance inflation factor (VIF) analysis was performed. VIF value is calculated

Table 5
Structures of compounds of general formula 2 taken for present study.



Formula 2

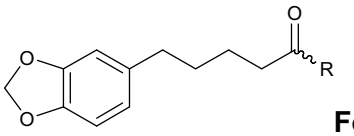
Compound no.	R
3	
4	
6	
25	

from $1/1 - r^2$, where r^2 is the multiple correlation coefficient of one descriptor's effect regressed on the remaining molecular descriptors. If VIF value is larger than 10, information of descriptors can be hidden by correlation of descriptors [41,42].

It has been observed that a high value of statistical characteristic need not be the proof of a highly predictive model [43,44]. Hence, in order to evaluate the predictive ability of our QSAR model, we used the method described by Golbraikh et al. [43], and Roy and Roy [44]. The values of correlation coefficient of predicted and actual activities and correlation coefficient for regressions through the origin (predicted versus observed activities) were calculated using the regression of analysis toolpak option of excel sheet and other parameters were calculated as reported by the above authors [43,44].

To arrive at the predictive R^2 (R^2_{Pred}) the following equation was used [44]:

Table 6
Structures of compounds of general formula 3 taken for present study.



Formula 3

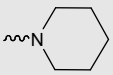
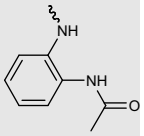
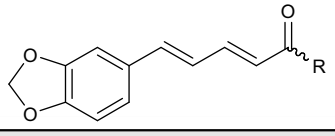
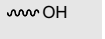
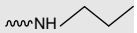
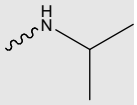
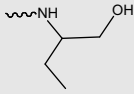
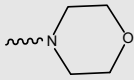
Compound no.	R
22	
23	
24	

Table 7
Structures of compounds of some non-active compounds.



Compound no.	R
26	
27	
28	
29	
30	

$$R^2_{\text{pred}} = 1 - \frac{\sum\left(Y_{\text{pred (Test)}} - Y_{\text{Test}}\right)^2}{\sum\left(Y_{\text{Test}} - \overline{Y}_{\text{training}}\right)^2}$$

where $Y_{\text{pred(Test)}}$ and Y_{Test} are the predicted and observed activity values respectively, of the test set compounds and $\overline{Y}_{\text{training}}$ is the mean activity values of the training set. Further evaluation of the predictive ability of the model was done by determining the value of rm^2 by the equation [44]:

$$rm^2 = R^2\left(1 - \left|\sqrt{R^2 - R_0^2}\right|\right)$$

where r^2 is the squared correlation coefficient between observed and predicted values and r_0^2 is the squared correlation coefficient between observed and predicted values without intercept.

Table 8
List of descriptors used in the study.

Type	Descriptors
E-state indices	Electro topological-state indices
Electronic	Sum of partial charges, sum of formal charges, dipole moment, energy of the highest occupied orbital, energy of the lowest unoccupied orbital, superdelocalizability
Information content	Information of atomic composition index, information indices based on the A-matrix, information indices based on the D-matrix, multigraph information content indices
Spatial	Radius of gyration, Jurs descriptors, shadow indices, area, density, PMI, V_m
Structural	Number of chiral centers, molecular weight, number of rotatable bonds, number of hydrogen-bond acceptors, number of hydrogen-bond donors
Thermodynamic	log of the partition coefficient, log of the partition coefficient atom-type value, desolvation free energy of water, desolvation free energy of octanol, heat of formation, molar refractivity
Topological	Wiener index, Zagreb index, Hosoya index, Kier and Hall molecular connectivity index, Balaban indices

Table 9
Statistical assessment of equations with varying number of descriptor (before removing the outliers).

No. of descriptors	r^2	SD	PRESS	q^2
1	0.553	3.56	1.87	0.476
2	0.681	3.56	1.44	0.594
3	**0.775**	**3.56**	**1.24**	**0.652**
4	0.790	3.56	1.58	0.557

The values of k and k', slopes of the regression line of the predicted activity vs. actual activity were calculated using the following equations [44]:

$$k = \frac{\sum y_i \tilde{y}_i}{\sum \tilde{y}_i^2} \quad \text{and} \quad k' = \frac{\sum y_i \tilde{y}_i}{\sum y_i^2}$$

where $\tilde{y}_i$ and y_i are the predicted and actual activities, respectively of the test set.

The leave-25%-out cross-validation for the training set was also carried out. The compounds were deleted (from the training set) in four cycles in the following manner: (1,5,9…), (2,6,10…), (3,7,12…) and (4,8,13…).

3. Results and discussion

The 25 active compounds identified by wet lab experiments as potentiators of anti-infective drug ciprofloxacin (Table 1) were divided into 20 molecules in training set and 5 in test set.

The first operational step involved the determination of the number of descriptors necessary and sufficient for the QSAR equation followed by a brute force approach to increase the number of terms in the QSAR equation one by one and evaluation of the effect of addition of new term on the statistical quality of the model. As the r^2 correlation coefficient can be easily increased by number of terms in the QSAR equation, the cross-validation correlation coefficient, q^2 was taken as the limiting factor for number of descriptors to be used in the model. It was observed that the q^2 value increased till the number of descriptors in the equation was 3 and decreased as the number of descriptors in the equation reached 4, as shown in Table 9. So the number of descriptors was restricted to 3. After analyzing the equation with three descriptors the following equation was arrived at:

$$\log(\mathrm{PF}) = -2.32 + 0.006 \times \text{Jurs-PNSA-1} - 0.0084 \times \text{Shadow} - \text{XZ} - 0.002 \times \text{Hf} \quad (1)$$

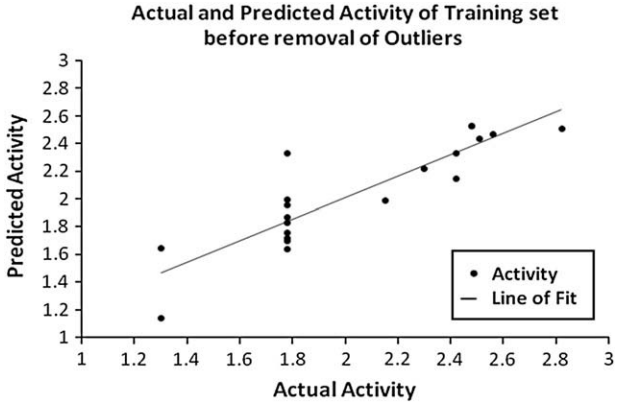


Fig. 1. Actual and predicted activities of training set before removal of outliers.

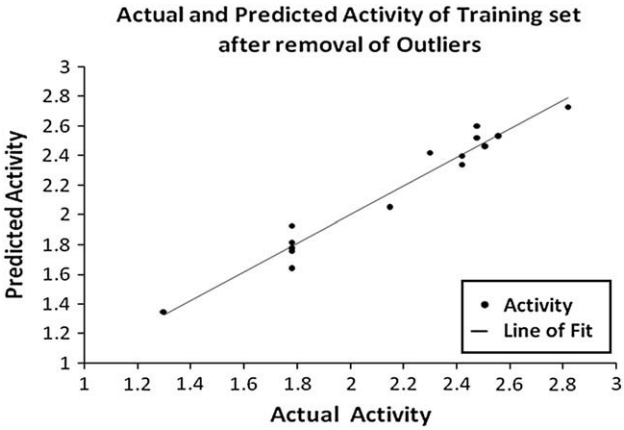


Fig. 2. Actual and predicted activities of training set after removal of outliers.

$$(N = 20;\ \mathrm{LOF} = 0.082;\ r^2 = 0.775;\ r^2_{\mathrm{adj}} = 0.733;\ F\text{-test} = 18.38;\ \mathrm{LSE} = 0.040;\ r = 0.880;\ q^2 = 0.652).$$

It was found that the compound numbers **9**, **14** and **24** were outliers. It was also observed that the activities of these compounds are very low and the GFA residual value very high. The probable reason for these compounds to be outliers could be that their mode of action may be different from the rest, which could be the probable reason for them to be outliers. Hence, these compounds were removed from the training set and a new equation was generated as given under:

$$\log(\mathrm{PF}) = 0.84 - 0.002 \times \mathrm{Hf} + 0.01 \times \text{Jurs-PNSA-1} - 0.01 \times \text{Shadow} - \text{XZ} \quad (2)$$

$$(N = 17;\ \mathrm{LOF} = 0.015;\ r^2 = 0.962;\ r^2_{\mathrm{adj}} = 0.953;\ F\text{-test} = 110.08;\ \mathrm{LSE} = 0.006;\ r = 0.981;\ q^2 = 0.917).$$

where N is number of compounds in training set, LOF is lack of fit score, r^2 is squared correlation coefficient, r^2_{adj} is square of adjusted correlation coefficient, F-test is a variance-related static which compares two models differing by one or more variables to see if the more complex model is more reliable than the less complex one, the model is supposed to be good if the F-test is above a threshold value, LSE is least-square error, r is correlation coefficient, q^2 is the square of the correlation coefficient of the cross-validation.

The graph of predicted and actual activities before and after the removal of the outliers is shown in Figs. 1 and 2 respectively.

Table 10
Results of randomization test performed to check the validation of the model.

Confidence level	90%	95%	98%	99%
Total trials	9	19	49	100
R from non-random	**0.98**	**0.98**	**0.98**	**0.98**
Random r's > non-random	0	0	0	0
Random r's < non-random	9	19	49	100
Mean value of r from random trial	**0.38**	**0.38**	**0.41**	**0.39**
Std. deviation of random trials	0.11	0.14	0.15	0.15
Std. deviation from non-random to mean	5.16	4.37	3.68	3.92

Table 11
Correlation matrix of the descriptors used in equation.

	PNSA-1	Shadow_XZ	Hf
PNSA-1	1		
Shadow_XZ	0.622	1	
Hf	0.424	0.147	1

Jurs_PNSA-1 is the partial negative surface area; this descriptor uses the sum of the solvent accessible surface area of all negatively charged atoms in the compound. Shadow_XZ is the area of the molecular shadow in the XZ plane. Hf is the heat of formation descriptor.

Randomization test was carried out in order to ascertain further statistical significance of the relationship between the anti-efflux pump activity and chemical structure descriptors. It was found that the r value of the original model was much higher than any of the trials using permuted data, showing thereby, that the model developed is statistically significant and robust as shown in Table 10.

It was also found that the intercorrelation of the descriptors used in the final model (Eq. (2)) was very low as shown in Table 11, thereby making the model statistically significant [45]. To further check the intercorrelation of descriptors variance inflation factor (VIF) analysis was performed (as described in Section 2.3.5). In this model, the VIF values of these descriptors are 2.42 (Jurs_PNSA-1), 2.20 (Shadow_XZ), and 1.15 (Hf). Therefore, from VIF analysis also it is clear that the descriptors used in the final model have very low intercorrelation.

The predicted activities of test set compounds were found to be very close to their actual activity as shown in Table 12, which indicates the robustness of the model and also indicates that it can be used confidently for predicting the anti-efflux pump activity of similar compounds. In order to further validate this, the external validation test values for predictive ability evaluation of the model were evaluated and found to be within the acceptable range [43] with the value of R^2 (squared correlation coefficient between the predicted and the actual activities of the test set) = 0.95, and the values of R_0^2 and $R_0'^2$ equal to 0.89 and 0.925. Therefore, the value of $[(R^2 - R_0^2)/R^2] = (0.95 - 0.89)/0.95 = 0.06/0.95 = \mathbf{0.063}$, which is less than 0.1 (stipulated value) [43]. Also, the values of k and k' were 0.91 and 1.09, respectively which are well within the specified range of 0.85 and 1.15. Highly predictive ability of the model is further justified with the value of q^2 being 0.917 which is an acceptable value [43].

Also the values of $R_{\text{Pred}}^2 = 0.78$ and $rm^2 = 0.72$ were found to be in the acceptable range [44], indicating thereby the good external predictability of the QSAR model.

Leave-25%-out cross-validation for the training set was also performed. The value of q^2 in all the four cycles was greater than 0.5 (Table 13), hence further validating the model internally.

Based on the developed QSAR model, it is observed that Jurs_PNSA-1 and Shadow_XZ are the important parameters that contribute to the potentiating activity. Also the parameter Hf (heat of formation) has emerged to be another important parameter for the design of bacterial NorA efflux pump inhibitors. The parameter Shadow_XZ is very well supported by the compounds set **10**, **11**, **12** and **15**, **16**, **17**, as also shown in Fig. 3. In these compounds, there is only the difference between the placement of the substituent group either in the *ortho*, *meta* or *para* positions; but it makes a significant difference in the area of the molecule made on the XZ plane and hence a difference in the activity value. Hence the Shadow_XZ parameter also contributes significantly in the efflux pump inhibitory activity and this parameter must be taken into consideration while designing the efflux pump inhibitors. The other parameter, which is the partial negative surface area is also very significant and is very well supported when we compare the molecule piperine (compound **25**) with its analogs (compounds **8**, **15**, **16**, **17**), where we notice that the introduction of polar groups to the original structure of piperine has increased the partial negative surface area of the compound and also the overall activity (i.e. inhibition of NorA) thereby justifying the developed model. Also when we compare the compounds **14** and **19**, we notice that with the introduction of S–C═N in place of CN, there is an increase in the overall activity of the compound. Hence, partial negative surface area of the compound plays an important role in the inhibition of NorA.

Table 12
The values of predicted and actual activities [(log(PF)] of test set compounds.

Compound no.	Actual activity	Predicted activity
3	1.78	1.79
5	1.78	1.89
7	2.15	2.32
18	2.30	2.74

Table 13
The values of q^2 and r^2 of the 4 cross-validation cycles leave-25%-out for the training set.

No. of cycles	Left out molecules	Molecules in training set	r^2	q^2
I	3,8,10,17,20,22	Rest 16 molecules	0.91	0.813
II	4,11,12,13,19,23	Rest 16 molecules	0.95	0.905
III	1,5,18,21,25	Rest 17 molecules	0.94	0.9
IV	2,6,7,15,16	Rest 17 molecules	0.939	0.904

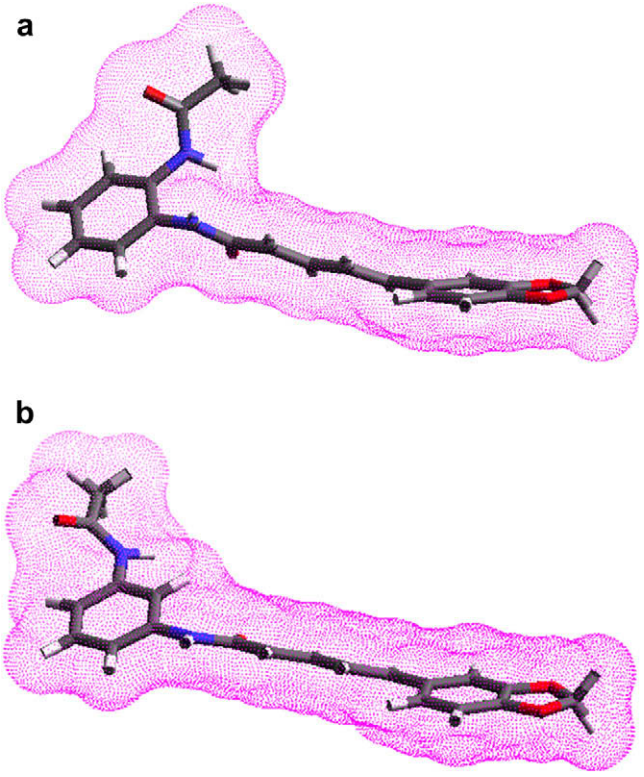


Fig. 3. (a) Surface area projection on XZ plane of compound **16**. The Shadow_XZ descriptor value for compound **16** is 82.5 Å^2. (b) Surface area projection on XZ plane of compound **17**. The Shadow_XZ descriptor value for compound **17** is 76.9 Å^2. The smaller surface area of compound **17** on XZ plane accounts for its higher activity in comparison to compound **16**.

4. Conclusion

The QSAR study shows that Jurs_PNSA-1, Shadow_XZ and heat of formation are important descriptors responsible for describing the activity of *S. aureus* NorA efflux pump inhibitors. The QSAR model is statistically and chemically sound and explains more than 95% of the variance in the experimental activity with excellent predictive power as is evidenced from the predicted activity of test set compounds.

Acknowledgement

The present work forms a part of the network project entitled "Drug target development using in silico biology" (CMM 0017). We are also thankful to the Department of Biotechnology (under the project GAP 0141) for their financial support.

References

[1] J.S. Francis, M.C. Doherty, U. Lopatin, C.P. Johnston, G. Sinha, T. Ross, M. Cai, N.N. Hansel, T. Perl, J.R. Ticehurst, K. Carroll, D.L. Thomas, E. Nuermberger, J.G. Bartlett, Clin. Infect. Dis. 40 (2005) 100–107.
[2] A. Opar, Nat. Rev. Drug Discov. 6 (2007) 943–944.
[3] A.S. Haddadin, S.A. Fappiano, P.A. Lipsett, Postgrad. Med. J. 78 (2002) 385–392.
[4] M.A. Martin, Curr. Clin. Top. Infect. Dis. 14 (1994) 170–191.
[5] S. Sabatini, G.W. Kaatz, M. Rossolini, D. Brandini, A. Fravolini, J. Med. Chem. 51 (2008) 4321–4330.
[6] A.S. Lynch, Biochem. Pharmacol. 71 (2006) 949–956.
[7] A. Mahamond, J. Chevalier, S. Alibert-Franco, W.V. Kern, J.M. Pages, J. Antimicrob. Chemother. 59 (2007) 1223–1229.
[8] O. Lomovska yan, A. Lee, K. Hoshino, H. Ishida, A. Mistry, M.S. Warren, E. Boyer, S. Chamberlemd, V.J. Lee, Antimicrob. Agents Chemother. 43 (1999) 1340–1346.
[9] G.D. Wright, Chem. Biol. 7 (2000) R127.
[10] K. Poole, O. Lomovskaya, Drug Discov. Today 3 (2006) 145.
[11] www.medicalnewstoday.com/articles/129160.php.
[12] K. Poole, J. Antimicrob. Chemother. 56 (2005) 20–51.
[13] C. Hansch, A. Kurup, R. Garg, H. Gao, Bioorg. Med. Chem. 101 (2001) 619–672.
[14] P.P. Maloney, C. Hansch, T. Fujita, R.M. Muir, Nature 194 (1962) 178–180.
[15] T. Fujita, J. Iwasa, C. Hansch, J. Am. Chem. Soc. 86 (1964) 5175–5180.
[16] C. Hansch, Acc. Chem. Res. 2 (1969) 232–239.
[17] V.J. Marder, Blood Rev. 15 (2001) 143–157.
[18] K. Tuppurainen, Chemosphere 38 (1999) 3015–3030.
[19] L.M. Shi, Y. Fan, T.G. Myers, K.D. Paull, J.N. Weinstein, J. Chem. Inf. Comput. Sci. 38 (1998) 189–199.
[20] S. Oloff, R.B. Mailman, A. Tropsha, J. Med. Chem. 48 (2005) 7322–7332.
[21] Y. Meneses-Marcel, Y. Marrero-Ponce, A. Machado-Tugores, D.M. Montero-Torres, J.A. Pereira, J.J. Escario, C. Nogal-Ruiz, V.J. Ochoa, A.R. Aran, R.N. Martinez-Fernandez, Garcia Sanchez, Bioorg. Med. Chem. Lett. 15 (2005) 3838–3843.
[22] L. Santana, E. Uriarte, H. Gonzalez-Diaz, G. Zagotto, R. Soto-Otero, E. Mendez-Alvarez, J. Med. Chem. 49 (2006) 1149–1156.
[23] I.A. Khan, Z.M. Mirza, A. Kumar, V. Verma, G.N. Qazi, Antimicrob. Agents Chemother. 50 (2006) 810–812.
[24] S. Koul, J.L. Koul, S.C. Taneja, I.A. Khan, A. Kumar, A. Tikoo, M. Tikoo, R.K. Johri, V. Verma, G.N. Qazi, Aromatic substituted pentadienoic acid amides, their di- and tetrahydro derivatives for combination with anti-infective drugs, US Patent 2007/0004645, dt. 04.01.2007.
[25] A. Kumar, I.A. Khan, S. Koul, J.L. Koul, S.C. Taneja, I. Ali, F. Ali, S. Sharma, Z.M. Mirza, M. Kumar, P.L. Sangwan, P. Gupta, N. Thota, G.N. Qazi, J. Antimicrob. Chemother. 6 (2008) 1270–1276.
[26] P.L. Sangwan, J.L. Koul, S. Koul, M.V. Reddy, N. Thota, I.A. Khan, A. Kumar, N.P. Kalia, G.N. Qazi, Bioorg. Med. Chem. 16 (2008) 9847–9857.
[27] G.M. Eliopoulus, R.C.J. Moellering, Antimicrobial combinations, in: V. Lorian (Ed.), Antibiotics in Laboratory Medicine, fourth ed. The Williams & Wilkins Co., Baltimore, MD, 1996, pp. 330–396.
[28] N.P. Brenwald, M.J. Gill, R. Wise, Antimicrob. Agents Chemother. 42 (1998) 2032–2035.
[29] C. de Gregorio, L.B. Kier, L.H. Hall, J. Comput.-Aided Mol. Des. 12 (1998) 557–561.
[30] L.B. Kier, L.H. Hall, Pharm. Res. 7 (1990) 801–807.
[31] K. Rose, L.H. Hall, L.B. Kier, J. Chem. Inf. Comput. Sci. 42 (2002) 651–666.
[32] G. Melagraki, A. Afantitis, H. Sarimveis, O. Igglessi-Markopoulou, C.T. Supuran, Bioorg. Med. Chem. 14 (4) (2006) 1108–1114.
[33] E. Estrada, E. Uriarte, Curr. Med. Chem. 8 (2001) 1573–1588.
[34] L.B. Kier, L.H. Hall, Molecular connectivity, Chemistry and Drug Research, Academic Press, New York, 1976.
[35] B. Kier, L.H. Hall, Molecular Connectivity in Structure–Activity Analysis, John Wiley Publ., London, 1986.
[36] A.R. Katritzky, Ekaterina Gordeeva, J. Chem. Inf. Comp. Sci. 33 (6) (1993) 835–857.
[37] www.accelrys.com.
[38] W. Shi, Q. Shen, W. Kong, Y. Bao-xian, Eur. J. Med. Chem. 42 (1) (2007) 81–86.
[39] D. Rogers, A.J. Hopfinger, J. Chem. Inf. Comput. Sci. 34 (1994) 854–866.
[40] S. Deswal, N. Roy, Eur. J. Med. Chem. 41 (11) (2006) 1339–1346.
[41] M. Jaiswal, P.V. Khadikar, A. Scozzafava, C.T. Supuran, Bioorg. Med. Chem. Lett. 14 (2004) 3283–3290.
[42] S. Shapiro, B. Guggenheim, Quant. Struct.-Act. Relat. 17 (1998) 327–337.
[43] A. Golbraikh, A. Tropsha, J. Mol. Graph. Model. 20 (2002) 269–276.
[44] P. Roy, K. Roy, QSAR Comb. Sci. 26 (2007).http://dx.doi.org/10.1002/qsar.200710043.
[45] S. Deswal, N. Roy, Eur. J. Med. Chem. 42 (4) (2007) 463–470.