

Cluster analysis and two-dimensional quantitative structure–activity relationship (2D-QSAR) of *Pseudomonas aeruginosa* deacetylase LpxC inhibitors

Rameshwar U. Kadam and Nilanjan Roy*

Pharmacoinformatics Division, National Institute of Pharmaceutical Education and Research, Sector 67,
S.A.S Nagar, Punjab 160062, India

Received 20 April 2006; revised 20 June 2006; accepted 7 July 2006

Available online 1 August 2006

Abstract—Compounds from a wide variety of structural classes inhibit *Pseudomonas aeruginosa* deacetylase LpxC. However, a single unified understanding of the relationship between the structures and activities of these compounds still eludes the researchers. We report herein, the development of cluster analysis-based 2D-QSAR models for LpxC inhibition. Principal component analysis (PCA), hierarchical cluster analysis (HCA), and genetic function approximation (GFA) were employed for the development of the QSAR model. The conventional 2D-QSAR model derived for the complete set of three-structural classes had unsatisfactory predictability with a correlation coefficient (r^2) of 0.703 and a cross-validated correlation coefficient (q^2) of 0.584. Descriptor-based cluster analysis indicated that the three-structural classes of LpxC inhibitors studied belonged to two clusters. Separate QSAR models for these two clusters showed substantially improved predictability with r^2 values of 0.904 and 0.944 and q^2 values of 0.805 and 0.906, respectively. Thus, we expect that compared to the conventional model, our two QSAR models can be better used to preliminarily screen molecules from a diverse chemical space while searching for novel LpxC inhibitors.
© 2006 Published by Elsevier Ltd.

Pseudomonas aeruginosa a Gram-negative opportunistic pathogen has been documented as a therapeutic problem because of nosocomial infection and antimicrobial resistance.¹ *P. aeruginosa* is responsible for 16% of nosocomial pneumonia cases, 12% of hospital-acquired urinary tract infections, 8% of surgical wound infections, and 10% of bloodstream infections.² The cell wall biosynthesis enzyme UDP-3-O-[R-3-hydroxymyristoyl]-GlcNAc deacetylase (LpxC) is currently recognized as an attractive target against *Pseudomonas* infection. The reaction catalyzed by LpxC is the first committed step of lipid A biosynthesis, which serves as a permeability barrier that protects the bacterium from many antibiotics, such as erythromycin.³ Till date, the inhibitors developed against LpxC enzymes of different Gram-negative bacteria contain hydroxamate or phosphonate zinc-binding motifs.^{4,5} The highest inhibitory potencies arise from aryl oxazolines with 3,4-disubstituted phenyl

rings containing fluoro or trifluoromethoxy in the *para* or *meta* positions and a two to five atom hydrophobic group in the complementary position. Electronic properties of the phenyl ring, the orientation of an oxygen lone pair, and a certain optimal hydrophobicity are the chief determinants of good inhibitory potency. The *m*-trimethoxy function may be serving some or all of these functions. Most molecules reported to date are not selective for *P. aeruginosa* LpxC.⁴ Thus, predictive models describing the relationship between structure and inhibition, applicable to diverse sets of molecules, could be valuable in the discovery of LpxC inhibitors.

The aim of this work was to develop a predictive QSAR model which will be applicable to diverse sets of molecules and would aid in search for the novel LpxC inhibitors from a diverse chemical space.

The LpxC inhibitors reported by Kline et al., were used for this study.⁴ Since orientation at 4-position in the ring is known to be of special importance, the molecules were removed if they were described as a racemic mixture or if their stereochemistry at the 4-position was not specified. On this basis, the data set of 51 of the 64 LpxC reported

Keywords: LpxC inhibitors; Cluster analysis; PCA; GFA.

* Corresponding author. Tel.: +91 172 2214682; fax: +91 172 2214692; e-mail: nilanjanroy@niper.ac.in

inhibitors was selected for QSAR analysis. The structures and experimental values of activity for the molecules used in this study are shown in Table 1a and 1b.

The ligands under study were built using SYBYL6.9 molecular modeling package installed on a Silicon Graphics Fuel Work station running IRIX 6.5 operating system.⁶ Since the crystal structure of the LpxC is not known, the basic skeleton and conformation for the most active molecule **20**, from the series was modeled and minimized using PM3 Hamiltonian using MOPAC interfaced with SYBYL6.9. In order to generate accurate charge information a single-point energy calculation was also performed using the AM1 Hamiltonian⁷ on the PM3 geometry optimized structure.⁸ The rest of the molecules were built by changing the required substitution using **20** as the template and were minimized similarly. Mulliken charges were assigned to all the molecules.⁹

Nearly 480 molecular descriptors including 0D (D = dimensionality) or constitutional, 1D (e.g., empirical descriptors and molecular properties) and 2D (such as: 2D autocorrelations, topological indices, BCUT descriptors, Galvez topological charges indices, molecular walk counts) were calculated using DRAGON5.3.¹⁰ A correlation matrix of the molecular descriptors was prepared and highly correlated descriptors with a correlation value of 0.9 or above were removed from the study and remaining descriptors were used for the GFA^{11–17} study to develop QSAR model in Cerius² 4.10 software¹⁸ (Table 2).

The equation term was set to linear polynomial and the mutation probability was specified as 50%. The length of the equations was set to five terms and a constant. The population size was established as 100. All equations were sorted by a statistical term, the correlation coefficient (r^2). The best equations were saved for subsequent studies.

Cluster analysis is the generic term applied for a wide variety of procedures that can be used for the classification of a heterogeneous data into relatively homogeneous groups. In QSAR modeling, it can be used to examine homogeneity of the data, detect some unusual data points, identify patterns, and to indicate potentially interesting relationships in the data. In the present study, principal component analysis was performed on the molecular descriptors which were initially scaled to variance of 1.0 using the ‘scale to unit variance’ option in Cerius² 4.10 software. Three principal components calculated were used to cluster the complete data set by hierarchical cluster analysis (HCA)–Wards method in the Cerius² 4.10 software.

Initially, we tried to develop a statistically significant model to predict pIC_{50} for the complete data set. The data set of 51 molecules was divided into training set consisting of 37 molecules and test set consisting of 14 molecules such that maximum diversity of structure and activity was maintained. The equation built using the training set molecules was called *Model 1*. The fits of model 1 to the training set and test set are shown in

Figure 1 and the predicted pIC_{50} values are shown in Table 1a,1b.

The five descriptors identified to be related to the inhibitory activity are matslv (a 2D-autocorrelation descriptor) which is Moran autocorrelation lag one weighted by van der Waals volumes; VEA (an eigenvalue-based indices descriptor) which is an Eigenvector coefficient sum from adjacency matrix; SPP (a charge descriptor) which is a sub-molecular polarity parameter; X4V (connectivity descriptor) which is a valence connectivity index chi-4; D/Dr05 (a topological descriptor) which is a distance/detour ring index of order 5.

$$\begin{aligned} \log(1/\text{IC}_{50}) = & -7.04877 \\ & + 43.526 \times \text{matslv}^{\wedge}2 \\ & + 3.48946 \times \text{VEA1} \\ & - 1.75733 \times \text{SPP}^{\wedge}2 \\ & - 0.1086 \times \text{X4V}^{\wedge}2 \\ & - 0.0232 \times (\text{D/Dr05} - 88.418) \end{aligned}$$

$N = 37$, LOF = 0.314, $r^2 = 0.703$, $r_{\text{adj}}^2 = 0.655$, F -test = 14.698, LSE = 0.155, $r = 0.839$, $q^2 = 0.584$, $\text{BS}r^2 \pm \text{SD} = 0.704 \pm .006$, $r_{\text{pred}}^2 = 0.107$; where, N is the number of molecules in training set, LOF is lack of fit score that resists overfitting, r^2 is squared correlation coefficient, and r_{adj}^2 is square of adjusted correlation coefficient; F -test is a variance-related statistic that compares two models differing by one or more variable 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 square of the correlation coefficient of the cross validation; $\text{BS}r^2$ is the r^2 value calculated by bootstrap method. A high bootstrap r^2 with a low standard deviation indicates the robustness of the model. r_{pred}^2 is the predicted correlation coefficient which indicates that the model can predict well the activity of molecules not considered in the training set. However, in model 1 neither the correlation coefficient ($r^2 = 0.703$) nor the cross-validated correlation coefficient ($q^2 = 0.584$) was satisfactory (Fig. 2). The generalized r_{pred}^2 value increased to 0.107 after the removal of four outliers (molecules 35, 53, 65, and 69, which had the highest residual values) from test set, yet the prediction could not be considered to be satisfactory.

The assumption was that the molecules have diverse structures, so the activities of these molecules should be more accurately reflected by separate QSAR models. Hence cluster analysis was employed to classify inhibitors resulting in their division into two clusters. QSAR modeling was performed for each cluster separately using GFA. As we mentioned earlier, GFA gives a set of equations. Although, the correlation coefficients for the top several models were nearly equal, we only discuss the best model for each cluster in this paper. This is justifiable because, the models with similar correlation coefficients (at least three to five models) differ by only one out of five variables that too belonging to the same sub-class of descriptors and therefore reflecting similar

Table 1a. Structures and activities of training set molecules used for QSAR study

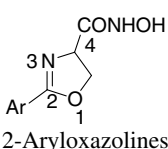
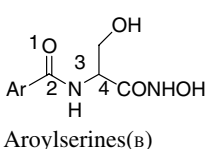
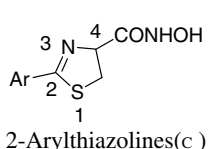
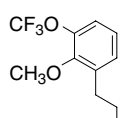
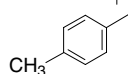
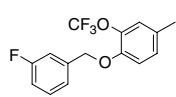
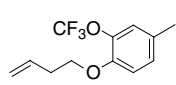
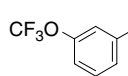
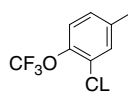
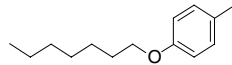
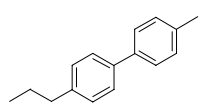
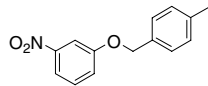
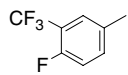
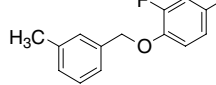
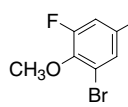
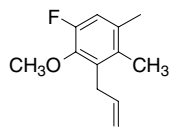
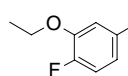
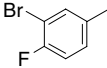
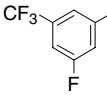
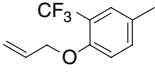
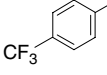
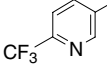
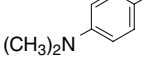
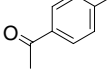
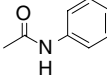
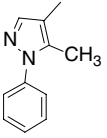
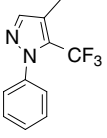
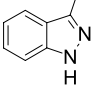
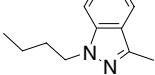
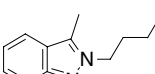
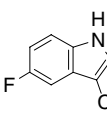
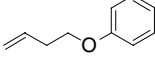
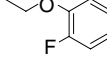
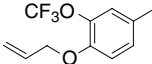
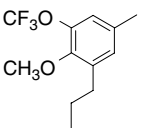
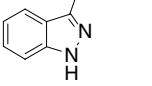
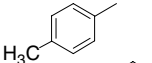
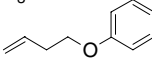
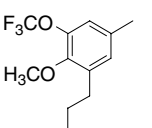
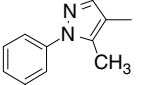
Mol ID	Ar	pIC ₅₀	
		Expt.	Predicted
	 2-Aryloxazolines	 Arylserines(B)	 2-Arylthiazolines(c)
5 ^a		6.30	6.11
19b ^b		4.69	5.00
21 ^b		6.60	6.43
22 ^b		6.45	6.71
25 ^b		5.22	4.99
27 ^b		4.52	4.86
30 ^a		5.25	4.74
32 ^a		6.55	6.14
34 ^b		6.25	6.37
36 ^b		5.88	5.91
40 ^b		6.07	6.14
42 ^a		5.04	4.78
43 ^a		5.48	5.64
44 ^b		6.00	5.67

Table 1a (continued)

Mol ID	Ar	pIC ₅₀	
		Expt.	Predicted
45 ^b		5.30	5.23
46 ^b		4.60	4.80
48 ^b		5.65	5.62
49 ^b		5.10	5.27
50 ^b		5.00	5.04
52 ^a		5.30	5.06
56 ^a		4.74	4.70
57 ^a		4.52	4.56
58 ^a		4.52	4.68
59 ^b		4.52	4.41
60 ^b		5.25	5.32
61 ^a		5.00	4.83
62 ^a		4.92	4.89
64 ^b		6.52	6.43
66 ^{B a}		5.40	5.51
68 ^{B a}		4.33	4.51

(continued on next page)

Table 1a (continued)

Mol ID	Ar	pIC ₅₀	
		Expt.	Predicted
70 _B ^a		5.12	4.69
71 _B ^a		4.30	4.38
72 _B ^a		4.88	5.01
76 _C ^a		4.41	4.25
77 _C ^a		4.23	4.55
80 _C ^b		4.74	4.78
81 _C ^a		4.22	4.00

Note: (B) indicates aroylserines series and (C) indicates 2-arylthiazolines series and remaining are 2-aryloxazolines series.

^a Structures in cluster 1.

^b Structures in cluster 2.

aspects of the structures. Thus the top several models are minor variants of each other.

Model 2: Is the equation built for the training set of the cluster 1, the fits of model 2 to the training set and test set of cluster 1 are shown in Figure 2 and the predicted pIC₅₀ values are shown in Table 1a and 1b. The model relates four descriptors to inhibition. The first one is MAXDP, maximal electrotopological positive variation, which is connectivity indices descriptor. The second one is JHETE, Balaban-type index from electronegativity weighted distance matrix, which is eigenvalue-based indices descriptor. The third descriptor is GATS5m, Geary autocorrelation –lag 5/weighted by atomic masses, which is 2D-autocorrelation descriptor. The fourth one is vez1, Eigenvector coefficient sum from z-weighted distance matrix, which is eigenvalue-based descriptor. A translation showing the equivalence between our codes and the commonly accepted symbols has been shown in Table 3.

$$\begin{aligned} \log(1/IC_{50}) = & 5.3512 \\ & + 0.0795 \times \text{MAXDP}^2 \\ & + 16.6973 \times \langle 1.961 - \text{JHETE} \rangle \\ & - 1.755 \times \text{GATS5m}^2 \\ & - 2.8359 \times \langle \text{vez1} - 4.509 \rangle \end{aligned}$$

$N = 20$, $\text{LOF} = 0.170$, $r^2 = 0.904$, $r_{\text{adj}}^2 = 0.878$, $F\text{-test} = 35.126$, $\text{LSE} = 0.042$, $r = 0.951$, $q^2 = 0.805$, $\text{BS}r^2 \pm \text{SD} = 0.904 \pm 0.003$, $r_{\text{pred}}^2 = 0.598$.

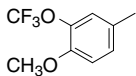
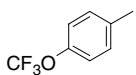
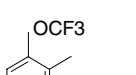
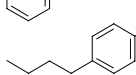
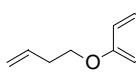
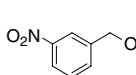
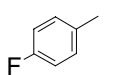
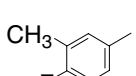
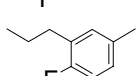
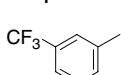
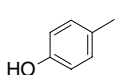
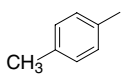
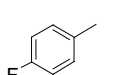
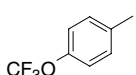
Model 3: Is the equation built for the training set of the cluster 2, the fits of model 3 to the training set and test set of cluster 2 are shown in Figure 3 and the predicted pIC₅₀ values are shown in Table 1a and 1b. The model relates four descriptors to inhibition, the first one is MATS1P, Moran autocorrelation –lag 1/weighted by atomic polarizability, which is 2D-autocorrelation descriptor. The second one is MDDD which is connectivity based indices descriptor. The third descriptor is MATS8m, Moran autocorrelation –lag 8/weighted by atomic masses, which is 2D-autocorrelation descriptor. The fourth one is VEP2, average eigenvector coefficient sum from polarizability weighted distance matrix, which is eigenvalue-based descriptor.

$$\begin{aligned} \log(1/IC_{50}) = & 3.9227 \\ & + 165.06 \times \text{MATS1P}^2 \\ & + 0.0048 \times \text{MDDD}^2 \\ & + 2.1208 \times \langle -0.023 - \text{MATS8m} \rangle \\ & - 53.057 \times \langle 0.223 - \text{VEP2} \rangle \end{aligned}$$

$N = 17$, $\text{LOF} = 0.163$, $r^2 = 0.944$, $r_{\text{adj}}^2 = 0.925$, $F\text{-test} = 50.359$, $\text{LSE} = 0.028$, $r = 0.944$, $q^2 = 0.906$, $\text{BS}r^2 \pm \text{SD} = 0.904 \pm 0.00$, $r_{\text{pred}}^2 = 0.608$.

The descriptors appearing in two models provide some insight into the nature of the binding sites and structural requirement as potent LpxC inhibitors. In case of model 2 the positive correlation of topological descriptor

Table 1b. Structures and activities of test set molecules used for QSAR study

Mol ID	Ar	pIC ₅₀	
		Expt.	Predicted
20 ^b		6.79	7.18
26 ^b		5.30	5.30
28 ^a		5.00	5.74
29 ^a		5.13	5.20
31 ^a		5.25	4.65
33 ^a		6.00	6.15
35 ^b		5.22	6.07
37 ^a		5.92	5.19
38 ^b		6.01	5.78
47 ^b		5.60	5.81
53 ^b		4.52	7.87
65 ^B ^a		4.30	5.46
67 ^B ^a		4.30	5.05
69 ^B ^a		5.81	4.82

Note: (B) indicates aroylserines series and (C) indicate 2-arylthiazolines series and remaining are 2-aryloxazolines series.

^a Structures in cluster 1.

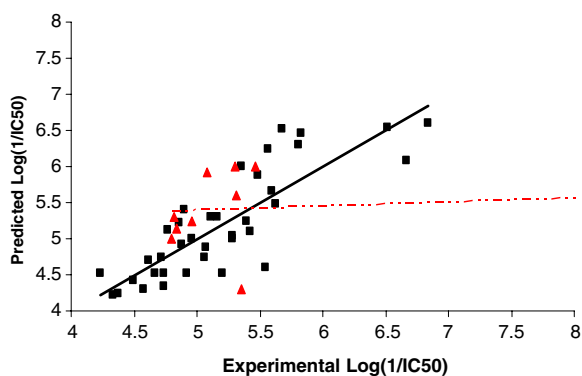
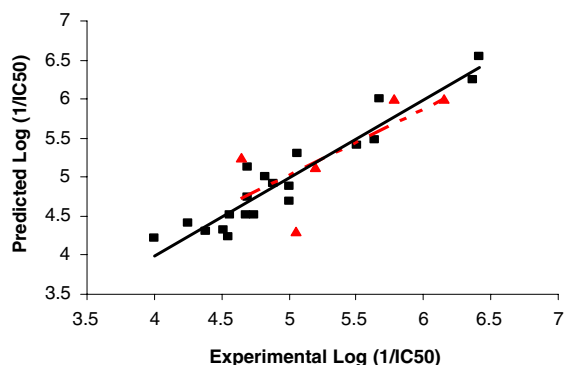
^b Structures in cluster 2.

(MAXDP, JHETE) toward the pIC₅₀ value indicated that increase in the substitution at C-2 and C-5 in phenyl ring is favorable for activity and the negative value from eigenvalue-based indices (VEZ1) contains contribution from all atoms and thus reflects topology of whole molecule. 2D-autocorrelation descriptors (GATS5m, MATS1P, and MATS8m) played different role in the

models 2 and 3. These descriptors indicate the role of physicochemical properties such as mass, volume, and polarizability of compound in deciding the activity. The positive correlation of the GATS5m, MATS8m to pIC₅₀ value indicated that structures weighted by atomic masses, that is, addition of bulky substitution like fluorine, trimethoxy groups at *meta* or *para* in phenyl

Table 2. Correlation matrix of molecular descriptors used in QSAR analysis

	mats8m	Mats1v	mats1p	gats5m	Spp	X4v	Veal	Ve1	Maxdp	mddd	D/Dr05	jhete	vep2	pIC ₅₀
mats8m	1													
Mats1v	0.032	1.00												
mats1p	0.014	0.663	1.00											
gats5m	0.208	0.059	0.048	1.00										
Spp	0.223	-0.197	-0.202	-0.137	1.00									
X4v	0.093	0.080	-0.063	0.120	-0.144	1.00								
Veal	0.217	-0.377	-0.516	-0.089	0.492	0.46	1.00							
Ve1	0.298	-0.242	-0.334	-0.227	0.520	0.452	0.812	1.00						
Maxdp	0.205	-0.027	-0.083	0.034	0.064	0.124	0.342	0.347	1.00					
mddd	0.199	-0.279	-0.437	-0.189	0.339	0.524	0.761	0.854	0.150	1.00				
D/Dr05	-0.049	-0.165	-0.223	0.076	-0.353	0.638	0.336	0.191	-0.031	0.302	1.00			
jhete	0.102	-0.031	-0.074	-0.108	0.606	-0.427	-0.031	0.088	0.249	-0.176	-0.798	1.00		
vep2	-0.274	0.305	0.451	0.129	-0.437	-0.566	-0.859	-0.886	-0.269	-0.877	-0.382	0.081	1.00	
pIC ₅₀	0.108	-0.409	-0.363	-0.094	0.030	0.070	0.515	0.289	0.242	0.377	0.211	-0.236	-0.388	1.00

**Figure 1.** Scatter-plots of actual versus predicted activity for both training (■) and test set (▲) molecules. (QSAR-without clustering) Solid line indicates training set point and dot line indicates test set point.**Figure 2.** Scatter-plots of actual versus predicted activity for both training (■) and test set (▲) molecules. (cluster-1) Solid line indicates training set point and dot line indicates test set point.

ring favourable for maximum activity. The positive contribution to activity of MATS1P, descriptors indicate that the polarizability, useful for maximum activity. Comparatively very less predictivity of model 1 is due to the negative contribution from SPP, X4V, D/Dr05 representing charge, connectivity indices, and topology of molecules. Negative contribution of electronic charge indicates that the less electronic densities

at ligand–receptor site led to decrease in affinity at receptor site.

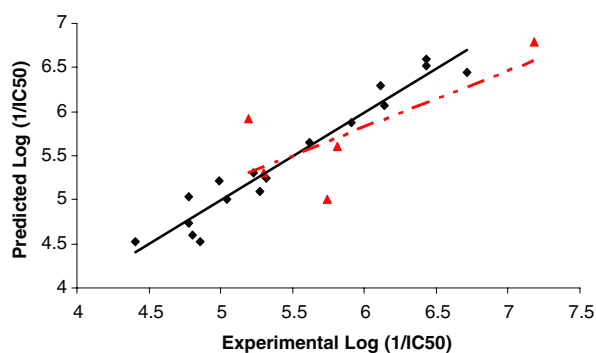
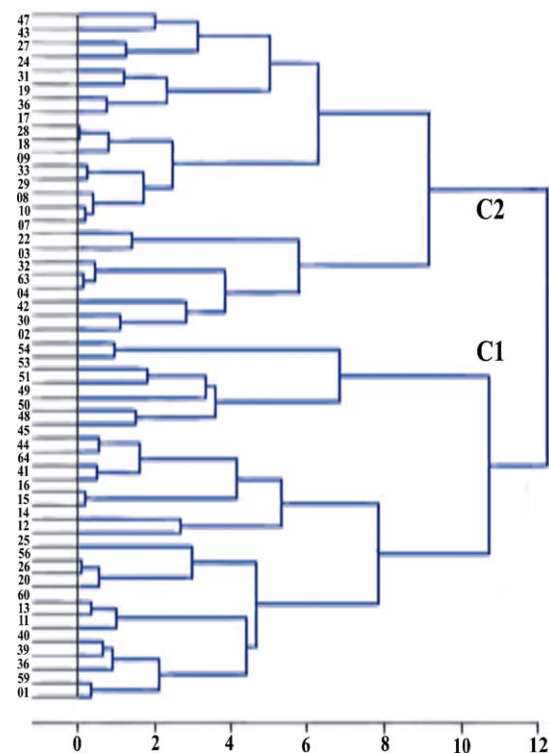
An outlier to a QSAR is identified normally by having a large standard residual and can indicate the limits of applicability of QSAR models.¹⁹ There are several reasons for their occurrence in QSAR studies, for example, an incorrectly measured experimental value that might be significant when analyzing the large data sets, a significant difference in the physicochemical properties, or structural uniqueness. Chemicals might be acting by a mechanism different from that of the majority of the data set. Although it is acceptable to remove a small number of outliers from QSAR, it is noted that it is not acceptable to remove the outlier repeatedly from a QSAR analysis simply to improve a correlation.

In the present work, r_{pred}^2 values obtained by two cluster models and one conventional model are listed in Table 1a and 1b. Initially, model 2 exhibited an r_{pred}^2 value of -0.33 which increased to 0.598 after removing two outliers from seven test set molecules. Model 3 exhibited an r_{pred}^2 value of -0.399 which increased to 0.608 after removing two outliers from seven test set molecules, whereas, the conventional model exhibited an r_{pred}^2 value of 0.104 which increased to 0.107 after removing same four molecules from fourteen test set molecules in model 1. Molecules **35** and **53** in cluster 1 presented large residuals and were removed as outliers, also, compound **53** has been found to be inactive experimentally.⁴ In cluster 2, molecules **65** and **69** presented large residuals and were removed as outliers. Both molecules **65** and **69** are acylserines and less potent than corresponding oxazolines. Thus the high bootstrap r^2 with a low standard deviation, along with good r_{pred}^2 , indicates the robustness of the model developed using cluster analysis-based approach, as compared to conventional model developed (Fig. 4).

In the present study, conventional 2D-QSAR modeling of LpxC inhibition, a single model developed for all classes of inhibitors, could not adequately describe the relationship between structures and activities. Accordingly, descriptor-based cluster analysis was performed, two clusters of inhibitors have been identified and predictive

Table 3. Symbols and definitions of the descriptors

S. No.	Descriptors	Symbol	Descriptors' meaning
1	Mats1v	$(I(d))$	Moran autocorrelation lag one weighted by van der Waals volume
2	Mats1p	$(I(d))$	Moran autocorrelation lag one weighted by atomic polarizability
3	Mats8m	$(I(d))$	Moran autocorrelation lag one weighted by atomic masses
4	GATS5m	$(C(d))$	Geary autocorrelation lag one weighted by atomic masses
5	SPP	Δ	Submolecular polarity parameter
6	X4V	${}^m\chi_q^v$	Valence connectivity index chi-4
7	VEA1	λ_i^A	Eigenvector coefficient sum from adjacency matrix
8	VEZ1	λ_i^D	Eigenvector coefficient sum from z weighted distance matrix (Barysz matrix)
9	MAXDP	T_α^E	Maximal electrotopological positive variation
10	MDDD	$\Delta\sigma$	Mean distance degree deviation
11	D/Dr05	$[D/\Delta]_{ij}$	Distance/detour ring index of order 5
12	JHETE	J^x	Balaban type index from electronegativity weighted distance matrix
13	VEP2	λ_i^z	Average coefficient sum from polarizability weighted distance matrix

**Figure 3.** Scatter-plots of actual versus predicted activity for both training (■) and test set (▲) molecules. (cluster-2) Solid line indicates training set point and dot line indicates test set point.**Figure 4.** Cluster analysis of the data set of 51 of the 64 LpxC inhibitors, using HCA–Ward method in Cerius² 4.10 software.

QSAR models have been developed for each cluster using GFA. Chemometrics data as well as test set data indicate that the models derived by this approach have better predictability as compared to conventional 2D-QSAR. These models can be used in search for novel LpxC inhibitors from a diverse chemical space and designing potent LpxC inhibitors.

References and notes

- Nicoll, L. *N. Eng. J. Med.* **1995**, 332, 616.
- Van Delden, C.; Iglewski, B. H. *Emerg. Infect. Dis.* **1998**, 4, 551.
- Whittington, D. A.; Rusche, K. M.; Shin, H.; Fierke, C. A.; Christianson, D. W. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, 100, 8146, Epub 2003, June 20.
- Kline, T.; Andersen, N. H.; Harwood, E. A.; Bowman, J.; Malanda, A.; Endsley, S.; Erwin, A. L.; Doyle, M.; Fong, S.; Harris, A. L.; Mendelsohn, B.; Mdluli, K.; Raetz, C. R.; Stover, C. K.; Witte, P. R.; Yabannavar, A.; Zhu, S. *J. Med. Chem.* **2002**, 45, 3112.
- Pirrung, M. C.; Tumej, L. N.; Raetz, C. R.; Jackman, J. E.; Snehalatha, K.; McClerren, A. L.; Fierke, C. A.; Gantt, S. L.; Rusche, K. M. *J. Med. Chem.* **2002**, 45, 4359.
- SYBYL version 6.9*; Tripos Associates Inc.; 1699, S Hanley Road, St. Louis, MO631444; USA.
- Dewar, M. J.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. P. *J. Am. Chem. Soc.* **1985**, 107, 3902.
- He, L.; Jurs, P. C. *J. Mol. Graph.* **2005**, 23, 503.
- Mulliken, R. S. *J. Chem. Phys.* **1955**, 23, 1833.
- Dragon version 5.3*; Milano Chemometrics and QSAR Research Groups Inc., 2002.
- Rogers, D.; Hopfinger, A. J. *J. Chem. Inform. Comput. Sci.* **1994**, 34, 854.
- Hahn, M. R. *J. Med. Chem.* **1995**, 38, 2091.
- Tokarski, J. S. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 779.
- Tokarski, J. S. *J. Chem. Inf. Comput. Sci.* **1997**, 37, 792.
- Shi, L. M.; Yi, F.; Myers, T. G.; O'Connor, P. M.; Paull, K. D.; Friend, S. H.; Weinstein, J. N. *J. Chem. Inform. Comput. Sci.* **1998**, 38, 189.
- Venkataraman, P.; Hopfinger, A. J. *J. Med. Chem.* **1999**, 42, 2169.
- Gokhale, V. M.; Kulkarni, V. M. *Bioorg. Med. Chem.* **2000**, 8, 2487.
- Cerius² Version 4.10*, Accelrys Inc, 6985 Scranton Road, San Diego, CA, USA.
- Lipnick, R. L. *Sci. Total Environ.* **1991**, 109, 131.