Pergamon

Bioorganic & Medicinal Chemistry 9 (2001) 3153–3160

BIOORGANIC & MEDICINAL CHEMISTRY

Three-Dimensional Quantitative Structure–Activity Relationship (3D-QSAR) of 3-Aryloxazolidin-2-one Antibacterials

Rajeshri G. Karki and Vithal M. Kulkarni*

Pharmaceutical Division, Department of Chemical Technology, University of Mumbai, Matunga, Mumbai 400019, India

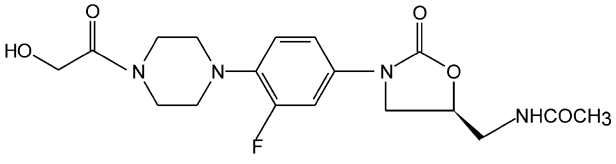
Received 19 March 2001; accepted 14 May 2001

Abstract—Three-dimensional quantitative structure–activity relationship (3D-QSAR) studies for 3-aryloxazolidin-2-one antibacterials were performed using the genetic function approximation algorithm. This study was performed using 60 compounds, in which the QSAR models were developed using a training set of 50 compounds. The in vitro minimum inhibitory concentration (MIC) against *Staphylococcus aureus* SFCO-1a was used for the study. The predictive ability of the QSAR model was evaluated by using a test set of 10 compounds. The statistical quality of the QSAR models was assessed using statistical parameters r^2, r^2_{cv} (cross-validated r^2), r^2_{pred} (predictive r^2) and lack of fit measure (LOF). The results obtained indicate that the antibacterial activity of the 3-aryloxazolidin-2-ones is strongly dependent on electronic factor as expressed by lowest unoccupied molecular orbital energy (LUMO), spatial factor as expressed by density and thermodynamic factors accounted for by molar refractivity and heat of formation. The model is presently being used to design and predict new potent molecules prior to synthesis.

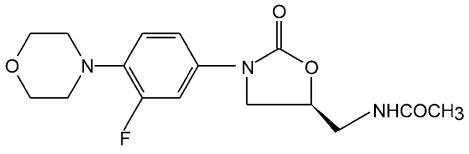
Introduction

The increasing incidence of multidrug resistance among Gram-positive bacterial pathogens represents one of the major challenges.[1–3] One particularly unsettling milestone has been the acquisition of resistance to vancomycin[4] an antibiotic generally regarded as the agent of last resort for serious Gram-positive infections. There is an urgent need for the discovery and development of new agents effective against the emerging and currently problematic Gram-positive pathogens, methicillin resistant *Staphylococcus aureus* (MRSA) and *Staphylococcus epidermidis* (MRSE), methicillin resistant coagulase negative *Staphylococci*, vancomycin resistant enterococci (VRE) and penicillin resistant pneumococci[5] as well as the perceived looming threat of a vancomycin resistant *S. aureus*.

Oxazolidinones typified by eperezolid (**1**) and linezolid (**2**) are one such class of antibacterial agents with potent activity against Gram-positive organisms including MRSA, MRSE and VRE.[6] They have been shown to selectively and uniquely bind to the 50S ribosomal subunit and inhibit bacterial translation at the initiation phase of protein synthesis.[7,8]



(1) Eperezolid



(2) Linezolid

Extensive synthetic and structure–activity studies have been carried out on this class of compounds.[9–15] The antibacterial activity data show that the acetamido group is by far the most active 'B' group (Fig. 1). Compounds with hydrocarbon and halo 'B' groups were found to exhibit weak antibacterial activity. The 'A' groups, when more hydrophilic, were found to be relatively more active with lipophilic 'B' group series and vice versa. In several series it has been observed that the antibacterial activity increases with increasing

*Corresponding author. Tel.: +91-22-414-5616; fax: +91-22-414-5614; e-mail: vithal@biogate.com

0968-0896/01/$ - see front matter
PII: S0968-0896(01)00186-9

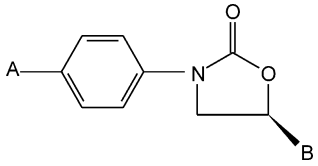


Figure 1. General structure of 3-aryloxazolidin-2-one antibacterials.

lipophilicity of the substituents in a homologous series (alkanes, alkyl ketones, halogens). Several close comparisons have also shown that conjugated, strongly electron withdrawing substituents confer enhanced activity on oxazolidinones relative to corresponding closely related electron donating substituents of comparable lipophilicity (nitro vs amino, ketones vs alcohols).

To gain further insights into the structure–activity relationship of these oxazolidinone antibacterials and to derive a predictive three-dimensional quantitative structure–activity relationship (3D-QSAR) model, we made use of the genetic function approximation algorithm (GFA) in Cerius2.[16] GFA algorithm offers a new approach to building structure–activity models. It automates the search for QSAR models by combining a genetic algorithm with statistical modeling tools. Thousands of candidate models are created and tested during evolution; only the superior models survive, and are then used as 'parents' for the creation of the next generation of candidate models. GFA has been successfully applied for the generation of 3D-QSAR models.[17–20] Such a model provides structure–activity insights, which can be used for designing of newer compounds and activity prediction prior to synthesis.

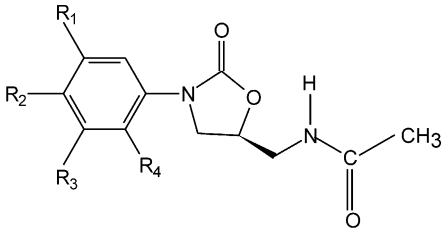
Results and Discussion

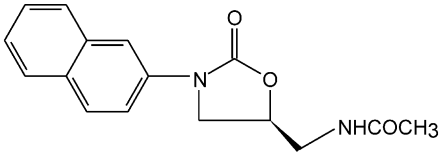
Different QSAR equations were generated using the GFA algorithm in Cerius2 for a series of 3-aryloxazolidin-2-one antibacterials. A total of 50 compounds were used for QSAR model generation (**1**–**50,** Table 1). It is essential to assess the predictive power of models by using a test set of compounds. This was achieved by setting aside 10 compounds as a test set (compounds **51**–**60,** Table 1) such that it represented the various functional groups included in the training set and had a regularly distributed biological data. The mean of the biological activity of the training and test set was 1.68 and 1.66, respectively.

Conformational sampling

For each of the compounds a conformational database was generated starting from the local minimized structure by random sampling method. The conformational models were generated using a rotation increment of 10° for all the torsional angles. Since the number of conformations generated was too large, only conformers within an energy threshold value of 10 kcal/mol of the local minimized structure were saved.[21] The torsion angle value between the aryl and oxazolidinone rings was calculated for the lowest energy conformations. All

Table 1. Structures of training (**1**–**50**) and test set compounds (**51**–**60**)



Compound	R_1	R_2	R_3	R_4	Log (1/C)[a]
1	H	H	H	H	0.263
2	H	CH_3	H	H	0.893
3	H	$-CH_2CH_3$	H	H	1.817
4	H	$-n$-$CH_2CH_2CH_3$	H	H	2.140
5	H	–CHO	H	H	1.516
6	H	$-COCH_3$	H	H	2.742
7	H	$-COCH_3$	Cl	H	2.191
8	H	$-COCH_3$	Br	H	1.948
9	H	$-COCH_3$	I	H	1.099
10	CH_3	$-COCH_3$	H	H	2.463
11	H	$-COCH_3$	F	H	2.469
12	H	$-COCH_3$	OH	H	2.165
13	H	$-COCH_2CH_3$	H	H	2.463
14	H	$-CH_2COCH_3$	H	H	1.861
15	H	$-COCH_2CH_2CH_3$	H	H	3.085
16	H	$-COCH(CH_3)_2$	H	H	2.483
17	H	$-COCH_2F$	H	H	2.168
18	H	$-COCH_2Cl$	H	H	2.793
19	H	$-COCH_2C{\equiv}N$	H	H	1.576
20	H	$-COCH_2NHCOCH_3$	H	H	1.620
21	H	$-COCH_2OCOCH_3$	H	H	2.825
22	H	$-COOCH_3$	H	H	2.165
23	H	–CH(=NOH)	H	H	1.239
24	H	$-C(=NOH)CH_3$	H	H	1.862
25	H	$-CH(OH)CH_2CH_3$	H	H	0.961
26	H	$-CH(OH)CH_2CH_2CH_3$	H	H	1.282
27	H	$-C(OH)(CH_3)_2$	H	H	2.165
28	H	$-CH(OH)CH(CH_3)_2$	H	H	1.884
29	H	$-NO_2$	H	H	2.145
30	H	$-NH_2$	H	H	0.289
31	H	$-N(C_2H_5)_2$	H	H	1.582
32	H	$-OCH_2CH_2CH_2CH_3$	H	H	1.583
33	H	$-CF_3$	H	H	1.577
34	H	$-SCH_3$	H	H	1.846
35	H	$-SCH_2CH_3$	H	H	2.469
36	H	$-SOCH_3$	H	H	1.870
37	H	$-SOCH_3$	NO_2	H	1.630
38	H	$-SOCH_2CH_3$	H	H	1.288
39	H	$-SOCH_2CH_2CH_3$	H	H	1.006
40	H	H	F	H	0.897
41	H	$-SOCH_3$	CH_3	H	1.589
42	H	CH_3	CH_3	H	1.516
43	H	$-SO_2NH_2$	H	H	0.991
44	H	H	CH_3	H	0.890
45	H	H	Cl	H	0.623
46	H	H	Br	H	1.292
47	H	H	$-CH_2CH_3$	H	1.215
48	H	H	$-OCH_3$	H	0.616
49	H	H	$-SCH_3$	H	1.244
50					1.852
51	H	$-CH{=}CH_2$	H	H	1.813
52	H	$-C{\equiv}CH$	H	H	1.208
53	H	$-CH(OH)CH_3$	H	H	1.843
54	H	$-SCH_2CH_2CH_3$	H	H	1.586
55	H	$-SO_2CH_3$	H	H	1.893
56	H	$-COCH_2Br$	H	H	1.948
57	H	$-N(CH_3)_2$	H	H	1.239
58	H	$-COCH_3$	CH_3	H	2.463
59	H	$-COCH_3$	NH_2	H	1.862
60	H	H	I	H	0.750

[a]C, minimum concentration in mM/l required to inhibit growth of *Staphylococcus aureus* SFCO-1a.

the molecules showed a negative torsional value. The mean torsional value of the lowest energy conformations of all the molecules was calculated to be −33.88°. This is in agreement with an earlier molecular modeling study carried out on DuP 721, wherein it was observed that the torsional angle value between the aryl and oxazolidinone rings was negative (−29° calculated for average low energy conformations).[22] Thus, the present approach of generating the conformations was found to adequately cover the conformation space of the molecules.

Computation of molecular descriptors

A total of 34 descriptors were calculated using the Cerius2 molecular modeling package. A list of the descriptors calculated is summarized in Table 2.

QSAR and GFA

A preliminary study was undertaken to study the effect of a number of crossovers and the value of the smoothing parameter '*d*'. GFA crossover of 100,000 and *d* value of 1 gave reasonable convergence. Hence, all further studies were carried out using these values.

A distinctive feature of GFA is that instead of generating a single model, as do most other statistical methods, it produces a population of models (e.g., 100). The range of variation in this population gives added information on the quality of fit and importance of descriptors. For example, the frequency of use of a particular descriptor in the population of equations may indicate how relevant the descriptor is to the prediction of activity.

Different sets of equations were generated by altering the chain length of the equations. The generated equations were evolved by repeating the GFA runs to check the stability of the GFA models. The final sets of GFA models were analyzed statistically to select the best model. The selection of the best model was based on the values of r^2 (square of the correlation coefficient for the training set of compounds), r^2_{cv} (cross-validated r^2), LOF (Friedman's Lack of Fit), r^2_{pred} (predictive r^2 for the test set of compounds).

An initial population of 100 equations was generated without restricting the number of descriptors (infinite chain length). The best model (Model A) had seven descriptors including a constant.

Model A

$$\log\ 1/C = 1.535 + 0.024\ \mathrm{CosV} + 0.361\ \mathrm{HOMO} + 1.239\ \mathrm{Shape\ RMS} - 0.315\ \mathrm{LUMO} - 0.0031\ \mathrm{Hf} + 0.000438\ \mathrm{Foct}$$

Table 2. Table of all descriptors used in this study

No.	Descriptor	Family	Description
1	Energy	Conformational	Energy of the currently selected conformation
2	Charge	Electronic	Sum of Partial charges
3	Fcharge	Electronic	Sum of formal charges
4	Apol	Electronic	Sum of atomic polarizabilities
5	Dipole mag	Electronic	Dipole moment-magnitude
6	Dipole X	Electronic	Dipole moment along *X* axis
7	Dipole Y	Electronic	Dipole moment along *Y* axis
8	Dipole Z	Electronic	Dipole moment along *Z* axis
9	HOMO	Electronic	Highest occupied molecular orbital energy
10	LUMO	Electronic	Lowest unoccupied molecular orbital energy
11	Sr	Electronic	Superdelocalizability
12	DIFFV	Molecular shape analysis (MSA)	Difference volume
13	Fo	MSA	Common overlap volume (ratio)
14	NCOSV	MSA	Non-common overlap steric volume
15	ShapeRMS	MSA	RMS to shape reference
16	COSV	MSA	Common overlap steric volume
17	SRVOL	MSA	Volume of shape reference compound
18	RadOfGyration	Spatial	Radius of gyration
19	Area	Spatial	Molecular surface area
20	Density	Spatial	Density
21	PMImag	Spatial	Principle moment of inertia—magnitude
22	PMIX	Spatial	Principle moment of inertia—along *X* axis
23	PMIY	Spatial	Principle moment of inertia—along *Y* axis
24	PMIZ	Spatial	Principle moment of inertia—along *Z* axis
25	Vm	Spatial	Molecular volume
26	MW	Structural	Molecular weight
27	Rotlbonds	Structural	Number of rotatable bonds
28	Hbond acceptor	Structural	Number of hydrogen bond acceptors
29	Hbond donor	Structural	Number of hydrogen bond donors
30	AlogP	Thermodynamic	Log of the partition coefficient
31	Fh2o	Thermodynamic	Desolvation free energy for water
32	Foct	Thermodynamic	Desolvation free energy for octanol
33	Hf	Thermodynamic	Heat of formation
34	MolRef	Thermodynamic	Molar refractivity

$r^2 = 0.732, \quad r^2_{cv} = 0.608, \quad r^2_{pred} = -0.0228,$

$LOF = 0.224$

Though the model showed a good internal predictivity, it exhibited poor external predictivity.

Interpretation of QSAR models with more terms becomes difficult in designing newer molecules. Moreover all the terms may not be relevant. Hence, to determine the more relevant descriptors, equations were generated with the option that they should have no more than four terms including a constant. The best model (Model B) yielded r^2 of 0.603, r^2_{cv} of 0.529, r^2_{pred} of 0.657 and LOF of 0.223.

Model B

$$\log\ 1/C = 11.610 - 0.841\ \text{LUMO} - 7.42\ \text{Density} - 0.0072\ \text{Hf}$$

Thus, it was observed that by restricting the number of descriptors in the QSAR model, there was not much change in the values of r^2, r^2_{cv} and LOF. On the contrary, the model had very good external predictivity. The analysis of the population of equations revealed that, equations with descriptors LUMO (lowest unoccupied molecular orbital energy) and Density had the best external and internal predictivity. The frequency of use of the 10 commonly used descriptors in the population of equations generated is summarized in Table 3. It gives an indication of how relevant the descriptor is to the prediction of activity. From Table 3, it is evident that descriptors LUMO and Density have been used the maximum irrespective of the chain length. Thus it indicates that these two descriptors have higher relevance to the prediction of activity.

To further confirm this, multiple linear regression analysis was done on different combinations of the 10 most frequently used descriptors taking four descriptors at a time. The 10 best equations are summarized in Table 4. It was observed that equations with descriptors LUMO, Density and Hf (heat of formation) exhibited the best correlation with biological activity. This was also observed in Model B.

To confirm the results of the multiple linear regression analysis, a population of equation was generated with the option that the equations should have five terms including a constant. Twenty-seven of the 33 best equations had descriptors LUMO, Density and Hf. The 5 best GFA models are described in Table 5. The best model (Model C) has r^2 of 0.651, r^2_{cv} of 0.592, r^2_{pred} of 0.657 and LOF of 0.215.

Model C

$$\log\ 1/C = 8.1857 - 0.00728\ \text{Hf} - 0.7374\ \text{LUMO} - 6.33\ \text{Density} + 0.026\ \text{MR}$$

Table 3. Frequency distribution of the variables used

Equations of infinite length		Equations with four terms		Equations with five terms	
Descriptors	Frequency	Descriptors	Frequency	Descriptors	Frequency
LUMO	80–90	LUMO	60	LUMO	95
Hf	85	Hf	35	Hf	80
Density	50	COSV	30	Density	30
COSV	30–40	Density	25	COSV	28
HOMO	25–30	RadOfGyration	15	Apol	18
Hbond donor	20	Hbond donor	15	HOMO	15
MolRef	20	HOMO	10	Area	15
Foct	20	Foct	10	RadOfGyration	12
ShapeRMS	20	Fh2o	10	MolRef	10
Area	5—10	Area	8	Hbond acceptor	10

Table 4. QSAR equations generated by multiple linear regression analysis

No.	Equation	LOF	r^2	r^2_{cv}	r^2_{pred}
1	log 1/C = 11.785−0.844 LUMO−0.00018 COSV−0.0072 Hf−7.539 Density	0.244	0.603	0.513	0.657
2	log 1/C = 8.995−0.762 LUMO + 0.288 RadOfGyration−0.0067 Hf−6.356 Density	0.230	0.625	0.543	0.654
3	log 1/C = 11.985−0.0071 Hf−7.717 Density−0.882 LUMO + 0.097 AlogP	0.238	0.613	0.519	0.546
4	log 1/C = 8.428−6.326 Density + 0.0052 Area−0.759 LUMO−0.0069 Hf	0.219	0.644	0.577	0.593
5	log 1/C = 11.275−0.0071 Hf−7.239 Density−0.828 LUMO + 0.0184 Rotlbonds	0.244	0.604	0.509	0.656
6	log 1/C = 12.194−0.889 LUMO−7.471 Density−0.0655 Hbond acceptor−0.0073 Hf	0.240	0.610	0.517	0.638
7	log 1/C = 11.608−0.0039 Sr−7.417 Density−0.840 LUMO−0.0072 Hf	0.244	0.603	0.508	0.659
8	log 1/C = 8.129−6.263 Density + 0.495 RadOfGyration + 0.126 AlogP−0.635 LUMO	0.339	0.449	0.328	0.359
9	log 1/C = 8.785−0.655 LUMO−6.756 Density + 0.0063 Area + 0.111 AlogP	0.345	0.440	0.325	0.428
10	log 1/C = 4.825 + 0.464 RadOfGyration + 0.00305 COSV−3.909 Density−0.536 LUMO	0.337	0.453	0.338	0.409

Table 5. Summary of the five best QSAR models generated

No.	Equation	LOF	r^2	r^2_{cv}	r^2_{pred}	F
1	log 1/C = 0.954−0.724 LUMO−0.009Hf + 0.124 MR−0.0253 MW	0.208	0.662	0.599	0.265	22
2	log 1/C = 8.1857−0.00728 Hf−0.7374 LUMO−6.33 Density + 0.026 MR	0.215	0.651	0.592	0.657	21
3	log 1/C = 10.213−0.00687 Hf−0.75 LUMO−7.865 Density + 0.005796 MW	0.221	0.640	0.574	0.595	20
4	log 1/C = 10.238−0.00684Hf + 0.00645 DIFFV−0.749 LUMO−6.35 Density	0.222	0.639	0.571	0.611	19.88
5	log 1/C = 9.93−0.006Hf + 0.0329 DIFFV−0.723 LUMO−0.0231 MW	0.226	0.629	0.554	0.637	19.07

There was no further improvement in the quality of the models generated, by altering the number of descriptors. Thus considering the diversity of the substituents on the aromatic ring model C with a conventional r^2 of 0.651 was selected as the QSAR model for the series of 3-aryloxazolidin-2-one antibacterials, as it exhibited good internal and external predictivity. The descriptors show low correlation among themselves indicating low probability of chance correlation. The estimated activity of the molecules of the training and test set using model C is shown in Table 6. A graph depicting the plot of predicted versus observed activities of the molecules of the training and test set using Model C is shown in Figures 2 and 3, respectively.

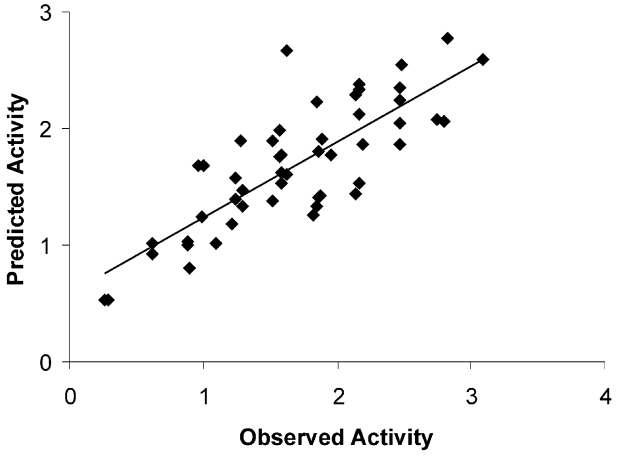


Figure 2. Calculated versus observed activity of the molecules of the training set using Model C.

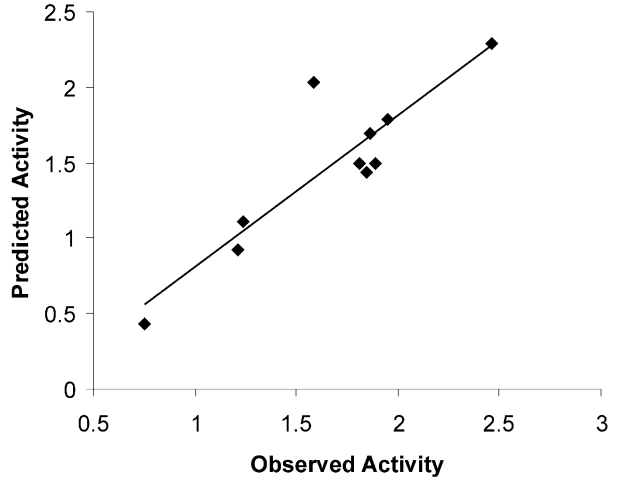


Figure 3. Calculated versus observed activity of the molecules of the test set using Model C.

Randomization tests

The statistical significance of the relationship between the antibacterial activity and chemical structure descriptors was further demonstrated by randomization procedure. The test was done by (1) repeatedly permuting the activity values of the data set, (2) using the permuted values to generate QSAR models and (3) 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. The r^2 values for 50 trials based on permuted data are shown in Fig. 4. The r^2 value of the original model was much higher than any of the trials using permuted data. Hence, model C is statistically significant and robust.

Interpretation of the QSAR model

The antibacterial activity of the series of 3-aryloxazolidin-2-one antibacterials is thus a function of LUMO (Lowest Unoccupied Molecular Orbital) energy, Density, Hf (Heat of Formation) and MolRef (Molar Refractivity). Descriptor LUMO is an electronic parameter. It measures the electrophilicity of the molecules. When a molecule acts as a lewis acid (an electron pair acceptor) in bond formation, incoming electrons are received in its LUMO. Molecules with low-lying LUMO are more able to accept electrons than those with high energy LUMO. Density is a 3-D-spatial descriptor that is defined as the ratio of molecular

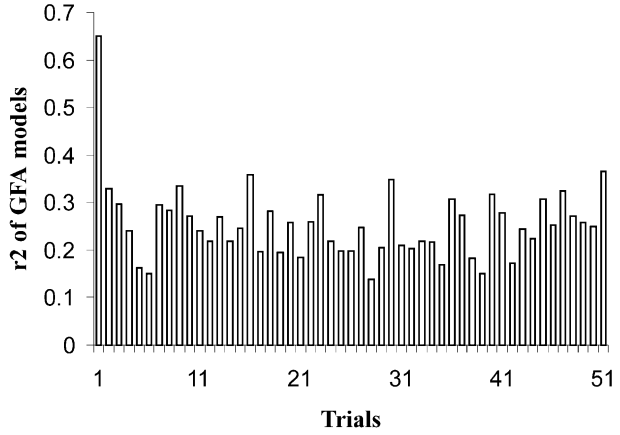


Figure 4. GFA randomization tests. The first bar shows the r^2 value for the model based on the actual data. The other 50 bars show the r^2 values for 50 models based on permuted data. The figure demonstrates that the model for actual data was statistically robust and reliable (see text for details).

Table 6. The comparison between the measured and estimated activity of the molecules used in the present study

Compound	Log (1/C)		
	Measured[a]	Estimated[b]	Residual[c]
1	0.263	0.533	−0.27
2	0.893	1.032	−0.139
3	1.817	1.263	0.554
4	2.140	1.444	0.696
5	1.516	1.890	−0.374
6	2.742	2.070	0.672
7	2.191	1.865	0.326
8	1.948	1.779	0.169
9	1.099	1.010	0.089
10	2.463	2.045	0.418
11	2.469	2.246	0.223
12	2.165	2.121	0.044
13	2.463	2.345	0.118
14	1.861	1.408	0.453
15	3.085	2.595	0.490
16	2.483	2.544	−0.061
17	2.168	2.373	−0.205
18	2.793	2.063	0.730
19	1.576	1.990	−0.414
20	1.620	2.672	−1.052
21	2.825	2.771	0.054
22	2.165	2.330	−0.165
23	1.239	1.574	−0.335
24	1.862	1.805	0.057
25	0.961	1.682	−0.721
26	1.282	1.893	−0.611
27	2.165	1.532	0.633
28	1.884	1.906	−0.022
29	2.145	2.281	−0.136
30	0.289	0.531	−0.242
31	1.582	1.524	0.058
32	1.583	1.779	−0.196
33	1.577	1.751	−0.174
34	1.846	1.340	0.506
35	2.469	1.864	0.605
36	1.870	1.431	0.439
37	1.630	1.599	0.031
38	1.288	1.465	−0.177
39	1.006	1.681	−0.675
40	0.897	0.801	0.096
41	1.589	1.622	−0.033
42	1.516	1.383	0.133
43	0.991	1.246	−0.255
44	0.890	0.998	−0.108
45	0.623	0.917	−0.294
46	1.292	1.327	−0.035
47	1.215	1.182	0.033
48	0.616	1.021	−0.405
49	1.244	1.393	−0.149
50	1.852	2.225	−0.373
51	1.813	1.501	0.312
52	1.208	0.923	0.285
53	1.843	1.442	0.401
54	1.586	2.036	−0.45
55	1.893	1.497	0.396
56	1.948	1.785	0.163
57	1.239	1.109	0.130
58	2.463	2.284	0.179
59	1.862	1.698	0.164
60	0.750	0.435	0.315

[a]Actual activity expressed as log 1/C; C, minimum concentration in mM/l required to inhibit growth of *Staphylococcus aureus* SFCO-1a.
[b]Predicted activity using Model C.
[c]Residuals calculated as difference between measured and estimated activity.

weight to molecular volume. Density reflects the types of atoms and how tightly they are packed in a molecule. Hf is the enthalpy gained in forming a molecule from its constituent atoms. It is a measure of the relative thermal stability of a molecule. MolRef is also a thermodynamic descriptor. MolRef index of a substituent is a combined measure of its size and polarizability.

Thus, from the 3D-QSAR model, one can conclude that the antibacterial activity of 3-aryloxazolidin-2-one will be higher if an electron withdrawing substituent is present on the aromatic ring. This will decrease the LUMO energy, increase the electrophilicity of the molecule and thereby elicit better activity. Compound **29** with a nitro substituent is more active compared to compound **30** with an amino substituent. Also compounds **13**, **15** and **16** with electron withdrawing ketone functionality are more active than compounds **25**, **26** and **28,** respectively, which have an electron donating alcoholic functionality.

In the QSAR model, Density is negatively correlated with activity. This means that substituent on the aromatic ring if compact will reduce the density and thereby increase the antibacterial activity (**25** vs **27**; **26** vs **28**). However this does not hold true for compound **15**, which is more active than compound **16** inspite of the substituent on compound **16** being more compact.

The molecules should have low heat of formation, as Hf is negatively correlated with activity. Hf represents the chemical stability and reactivity of the molecules. Since no information is available about chemical reactivity of the molecules with the target, usage of Hf in the QSAR model indicates conformational stability. This will probably favor better binding at the molecular level and thereby elicit better activity. Polarizability of the substituent has a positive effect on the biological activity. Thus, the overall activity will depend on the combination of the four chemical structure descriptors. The results are in accordance with the preliminary structure–activity relationship (SAR) reported by Gregory et al.[10]

Conclusion

A 3D-QSAR model has been derived using the genetic function approximation in Cerius2 for a series of 3-aryloxazolidin-2-one antibacterials. The best model generated correlates the antibacterial activity with LUMO, Density, Hf and MolRef. The model has good internal and external predictivity as shown by r^2_{cv} of 0.592 and r^2_{pred} of 0.657, respectively. The statistical significance and robustness of the model has been confirmed by doing a randomization study. Thus, in the series of 3-aryloxazolidin-2-ones, molecules with electron withdrawing substituent in conjugation with the aromatic ring, polarizable and compact enhance the antibacterial activity. Based on the QSAR model, some compounds have been synthesized and screened for antibacterial activity in our laboratory.[23]

Experimental

Data set

The chemical structures along with the biological activity data of the compounds used in this study are listed in Table 1. A total of 60 compounds with variously substituted aryl ring were selected for the study.[9–11] The biological activity of each of these compounds were evaluated under identical conditions and expressed as log (1/MIC) for the study, where MIC is minimum concentration in mM/L required to inhibit growth of *S. aureus* SFCO-1a. It was determined by a microtiter broth dilution assay.[24,25] To assess the predictive power of the models, a set of 10 compounds was arbitrarily set aside as test set such that it represented the various functional groups included in the training set (**51–60,** Table 1).

Modeling

Model building. Three-dimensional structures of each of the compounds were constructed using the 3-D-sketcher module in Cerius2. Each structure was energy-minimized using the Universal forcefield 1.02[26] and partial atomic charges were computed using the charge equilibration approach.[27]

Conformational sampling. The local minimized geometry was used as the initial structure for conformational analysis. Conformational ensembles were generated by random sampling method using a rotation increment of 10° for all the torsional angles. In order to restrict the number of conformers being generated to a maximum of 500, conformers with an energy threshold value of greater than 10 kcal/mol from the local minimized structure were rejected, thus selecting only the energetically stable conformers.

Computation of molecular descriptors. The molecular descriptors were computed using the Cerius2 molecular modeling package. A total of 34 descriptors categorized as: (1) conformational (2) electronic (3) spatial (4) structural (5) thermodynamic and (6) molecular shape analysis (MSA) descriptors were calculated. The lowest energy structure of compound **15**, the most active compound, was taken as the reference for calculation of MSA descriptors.

QSAR and GFA. The QSAR analysis[28–31] relates chemical descriptors (x) to a response variable (y) by a mathematical equation, most often in the form

$$y = f(x) = a_0 + \Sigma a_{ij} B_j(x_i)$$

where a is a (regression) coefficient, i ranges over the set of descriptors, and $B_j(x_i)$ is the jth basis function type for the ith descriptor. The function (operator) B may, for example, be a logarithmic transform, a quadratic transform, a half-space spline, or unity. To fit the equations and thereby generate QSAR models we used the recently developed GFA method of Rogers.[17,32] GFA combines Friedman's multivariate adaptive regression splines (MARS) algorithm[33,34] with Holland's genetic algorithm (GA).[35]

A training set of 50 compounds with their associated biological activity was used for the QSAR study. A preliminary study was undertaken to study the effect of the number of genetic crossover operations and the value of the smoothing parameter '*d*'. The smoothing parameter controls the number of terms in the model equation. The default value of d is 1.0, and larger values of d lead to smoother models (that is models with fewer terms). Following the analysis, the number of GFA crossover was set to 100,000 to achieve reasonable convergence and the value of d was set to 1.

A population of 100 equations was generated in each run by using a set of 34 descriptors and linear basis functions. The models were scored using Friedman's lack of fit (LOF) measure,[32,33] which is given by:

$$\mathrm{LOF} = \mathrm{LSE} \,/\left\{1 - (c + d^{*}p)/m\right\}^2$$

where LSE is the least-squares error (calculated from the difference between actual and calculated values for the activity index over data set), c is the number of basis function in the model. d is a smoothing parameter that controls the number of terms in the model equation, p is the total number of features contained in all terms of the model (some basis functions contain more than one feature), and m is the number of samples (compounds) in the training set. The LOF measure penalizes appropriately for the addition of terms to the equation (and consequent loss of degrees of freedom) in such a way to resist overfitting. The generated population of equations was repeatedly subjected to GFA runs to evolve the population and to check the stability of the models. Whichever type of algorithm is selected, the QSAR model should be predictive as well as descriptive. Hence, cross-validation and randomization of the dependent variable was used to test the predictive ability and robustness of the QSAR models.

Predictive r^2 values (r^2_{pred}). The predictive ability of each analysis was determined from a set of 10 compounds that were not included in the training set. The molecular descriptors were computed, and their activities predicted using the generated QSAR models. The predictive r^2 (r^2_{pred}) value will be based on molecules of the test set only and is defined as:[36]

$$r^2_{pred} = (\mathrm{SD} - \mathrm{PRESS})/\mathrm{SD}$$

where SD is the sum of the squared deviations between the biological activities of the test set and mean activity of the training set molecules and PRESS is sum of the squared deviation between predicted and actual activity values for every molecule in the test set.

Acknowledgements

The authors thank University Grants Commission (UGC) and Department of Science and Technology

(DST), Ministry of Science and Technology, Govt. of India and Cipla Ltd for the research supports. R.G.K. thanks Mr. Vijay M. Gokhale and Mrs. Varsha A. Sadekar for the helpful discussion.

References and Notes

1. Swartz, M. N. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 2420.
2. Tomasz, A. N. *Engl. J. Med.* **1994**, *330*, 1247.
3. Fekete, T. *Perspect. Biol. Med.* **1995**, *38*, 363.
4. Courvalin, P. *Antimicrob. Agents. Chemother.* **1994**, 1447.
5. Appelbaum, P. C. *Clin. Infect. Dis.* **1992**, *15*, 77.
6. Brickner, S. J. *Curr. Pharmaceut. Des.* **1996**, *2*, 175.
7. Lin, A. H.; Murray, R. W.; Vidmar, T. J.; Marotti, K. R. *Antimicrob. Agents. Chemother.* **1997**, *41*, 2127.
8. Shinabarger, D. L.; Marotti, K. R.; Murray, R. W.; Lin, A. H.; Melchior, E. P.; Swaney, S. M.; Dunyak, D. S.; Demyan, W. F.; Buysse, J. M. *Antimicrob. Agents Chemother.* **1997**, *41*, 2132.
9. Gregory, W. A.; Britelli, D. R.; Wang, C.-L. J.; Wuonola, M. A.; McRipley, R. J.; Eustice, D. C.; Eberly, V. S.; Bartholomew, P. T.; Slee, A. M.; Forbes, M. *J. Med. Chem.* **1989**, *32*, 1673.
10. Gregory, W. A.; Britelli, D. R.; Wang, C.-L. J.; Kejar, H. S.; Carlson, R. K.; Park, C.-H.; Corless, P. F.; Miller, S. J.; Rajagopalan, P.; Wuonola, M. A.; McRipley, R. J.; Eberly, V. S.; Slee, A. M.; Forbes, M. *J. Med. Chem.* **1990**, *33*, 2569.
11. Park, C.-H.; Britelli, D. R.; Wang, C. L.-J.; Marsh, F. D.; Gregory, W. A.; Wuonola, M. A.; McRipley, R. J.; Eberly, V. S.; Slee, A. M.; Forbes, M. *J. Med. Chem.* **1992**, *35*, 1156.
12. Barbachyn, M. R.; Toops, D. S.; Ulanowicz, D. A.; Grega, K. C.; Brickner, S. J.; Ford, C. W.; Zurenko, G. E.; Hamel, J. C.; Schaadt, R. D.; Stapert, D.; Yagi, B. H.; Buysse, J. M.; Demyan, W. F.; Kilburn, J. O.; Glickman, S. E. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 1003.
13. Barbachyn, M. R.; Toops, D. S.; Grega, K. C.; Hendges, S. K.; Ford, C. W.; Zurenko, G. E.; Hamel, J. C.; Schaadt, R. D.; Stapert, D.; Yagi, B. H.; Buysse, J. M.; Demyan, W. F.; Kilburn, J. O.; Glickman, S. E. *Bioorg. Med. Chem. Lett.* **1996**, *6*, 1009.
14. Tucker, J. A.; Allwine, D. A.; Grega, K. C.; Barbachyn, M. R.; Klock, J. L.; Adamski, J. L.; Brickner, S. J.; Hutchinson, D. K.; Ford, C. W.; Zurenko, G. E.; Conradi, R. A.; Burton, P. S.; Jensen, R. M. *J. Med. Chem.* **1998**, *41*, 3727.
15. Pae, A. N.; Kim, H. Y.; Joo, H. J.; Kim, B. H.; Cho, Y. S.; Choi, K.; Choi, J. H.; Koh, H. Y. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2679.
16. Computational results obtained using software program Cerius2 version 3.5, Molecular Simulations Inc., 9685, Scranton road, San Diego, CA 92121, USA.
17. Rogers, D.; Hopfinger, A. J. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 854.
18. Shi, M. L.; Fan, Y.; Myers, T. G.; Connor, P. M.; Paul, K. D.; Friend, S. H.; Weinstein, J. N. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 189.
19. Kiyoshi, H.; Toshiro, K.; Funatsu, K. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 112.
20. Gokhale, V. M.; Kulkarni, V. M. *Bioorg. Med. Chem.* **2000**, *8*, 2487.
21. Tanaji, T. T.; Kulkarni, V. M. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 204.
22. Gleave, D. M.; Brickner, S. J.; Manninen, P. R.; Allwinw, D. A.; Lovasz, K. D.; Rohrer, D. C.; Tucker, J. A.; Zurenko, G. E.; Ford, C. W. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 1231.
23. Karki, R. G.; Kulkarni, V. M. *Novel Oxazolidinone Derivatives as Antibacterial Agents* (patent application to be filed).
24. Slee, A. M.; Wunonola, M. A.; Mclipley, R. J.; Zajac, I.; Zawada, M. J.; Bartholomew, P. T.; Gregory, W. A.; Forbes, M. *Antimicrob. Agents Chemother.* **1987**, *31*, 1791.
25. National Committee for Clinical Laboratory Standards. *Tentative Standard M7-T*; National Committee for Clinical Laboratory Standards: Villanova, PA, 1982.
26. Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. *J. Am. Chem. Soc.* **1992**, *114*, 10024.
27. Rappe, A. K.; Goddard, W. A. *J. Phys. Chem.* **1991**, *95*, 3358.
28. Hansch, C.; Leo, A.; Hockman, D. *Exploring QSAR, Vol. 1. Fundamentals and Applications in Chemistry and Biology: Vol. 2. Hydrophobic, Electronic, and Steric Constants*; American Chemical Society: Washington, DC, 1995.
29. van de Waterbeemd, H., Ed. *Chemometric Methods in Molecular Design*: VCH: Weinheim, 1995; Vol. 2.
30. van de Waterbeemd, H., Ed. *Advanced Computer-Assisted Techniques in Drug Discovery*; VCH: Weinheim 1995; Vol. 3.
31. van de Waterbeemd. H., Ed. *Structure–Property Correlations in Drug Research*; Academic and R. G. Landes: San Diego, CA, 1996.
32. Rogers, D. In *Genetic Algorithms in Molecular Modeling*; Devillers, J., Ed.; Academic Press, 1996.
33. Friedman, J. H. *Multivariate Adaptive Regression Splines* (Technical Report No. 102, Nov./Rev. Aug 1990); Laboratory of Computational Statistics, Department of Statistics, Stanford University: Stanford, CA., 1988.
34. Friedman, J. H. *Ann. Stat.* **1991**, *19*, 1.
35. Holland, J. *Adaptation in Artificial and Natural Systems*; University of Michigan Press; Ann Arbor, MI, 1975.
36. Kulkarni, S. S.; Kulkarni, V. M. *J. Med. Chem.* **1999**, *42*, 373.