Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 12 (2004) 3027–3035
Bioorganic & Medicinal Chemistry

QSPR of 3-aryloxazolidin-2-one antibacterials

Alan R. Katritzky,[a,*] Dan C. Fara[a,b] and Mati Karelson[b]

[a]*Center for Heterocyclic Compounds, Department of Chemistry, University of Florida, PO Box 117200, Gainesville, FL 32611, USA*
[b]*Department of Chemistry, University of Tartu, 2 Jakobi Street, Tartu 51014, Estonia*

Received 24 October 2003; revised 12 February 2004; accepted 8 March 2004
Available online 20 April 2004

Abstract—A QSPR treatment has been applied to a data set consisting of 60 3-aryloxazolidin-2-one antibacterials to relate the in vitro minimum inhibitory concentration (MIC) (required to inhibiting growth of *S. aureus*) with theoretical molecular and fragment descriptors. The treatment using CODESSA PRO descriptors leads to a seven-parameter model with $r^2 = 0.820$ and $r^2_{cv} = 0.758$.

1. Introduction

The increase during the last decade of bacterial resistance to antibiotics poses a serious concern for medical professionals.[1] Multidrug resistant gram-positive bacteria including methicillin-resistant *Staphylococcus aureus* (MRSA) and *S. epidermitisi* (MRSE) and vancomycin resistant *Enterococci* (VRE) are of major concern. The most worrisome bacteria, the MRSA, are particularly virulent organisms that cause a broad array of problems starting from pimples to life-threatening bacteremia, osteomyelitis, etc. MRSA are resistant to most of the common antibiotics except Vancomycin; if they develop or acquire Vancomycin resistance, physicians have presently little in their arsenal to fight them.[2]

Oxazolidinones, a new class of synthetic antibacterials with activity against gram-positive pathogenic and anaerobic bacteria, have been shown to selectively bind to the 50S ribosomal subunit and inhibit bacterial translation at the initiation phase of protein synthesis. Extensive studies on the syntheses of oxazolidinone antibacterial agents and their structure–activity relationship (SAR) have been reported,[2–12] but few QSAR studies have been published. Linezolid (Zyvox®), recently approved in the USA and Europe as the first drug of this class is efficacious in treating skin and soft tissue infections, pneumonia and bacteremia.

Gregory and co-workers[3,4] studied the effect of varying the 'B' and 'A' groups in a series of oxazolidinone antibacterials and suggested that: (i) oxazolidinone antibacterials bearing acetamide as the 'B' group are the most active analogues among a large number of compounds bearing a variety of different 'B' groups, and (ii) 'A' group must be compact or relatively flat (see Fig. 1).

The first QSAR study,[7] reported in 1999, comprises a 3D-QSAR study on two novel series of oxazolidinone antibacterial agents using 'Comparative Molecular Field Analysis' (CoMFA). The authors used a training set of 17 with 2 reference compounds. The CoMFA steric and electrostatic fields and ClogP were used as descriptors, and the activities against methicillin-resistant *S. aureus* 88 (MRSA 88) as dependents. The cross-validated r^2_{cv} (0.653) and conventional r^2 (0.984) from the PLS and CoMFA analyses indicated considerable reliability for predicting the antibacterial activities of oxazolidinone antibacterial agents, but this study utilized a rather small data set.

The second 3D-QSAR study, reported by Karki and Kulkarni in 2001, used a genetic function algorithm[13] and 60 compounds. The QSAR models were developed using a training set of 50 compounds and the in vitro minimum inhibitory concentration (MIC) against

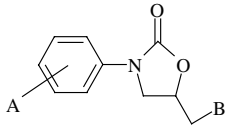

Figure 1. General structure of 3-aryloxazolidin-2-one antibacterials.

Keywords: QSPR/QSAR; Antibacterials; Molecular descriptors; CODESSA PRO.
* Corresponding author. Tel.: +1-352-392-0554; fax: +1-352-392-9199; e-mail: katritzky@chem.ufl.edu
URL: http://ark.chem.ufl.edu/

0968-0896/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2004.03.013

S. aureus SFCO-1; the r^2 values reported[13] for the models range from 0.629 to 0.732. The predictive ability of the QSAR model was evaluated with a test set of 10 compounds. The authors concluded 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.

Gopalakrishnan et al.[14] performed a three-dimensional quantitative structure–activity relationship (3D-QSAR) studies for a series of tricyclic oxazolidinones, using CoMFA and Comparative Molecular Similarity Indices Analysis (CoMSIA) procedures. The QSAR model was developed using a training set of 33 compounds and the predictive ability of this model was assessed using a test set of 9 compounds. The predictive 3D-QSAR models were found to have conventional r^2 values of 0.975 and 0.940 for CoMFA and CoMSIA, respectively (the respective cross-validated coefficient q^2 values were 0.523 and 0.557 for CoMFA and CoMSIA). Thus, it was concluded that CoMFA 3D-QSAR model performs better than the CoMSIA model.

In the present work, our aim is to develop robust QSPR/QSAR models and to explain the antibacterial activity of oxazolidinones using theoretical molecular and fragmental descriptors.

2. Data set and methodology

The data set used of 60 different organic compounds (see Table 1) is identical with that used by Karki and Kulkarni.[13] The in vitro minimum inhibitory concentration (MIC) in mM/L required for inhibiting growth of *S. aureus* SFCO-1 was the property studied.

The basic skeleton of all 3-aryloxazolidin-2-one antibacterials studied is given in Figure 2; the substituents R^1, R^2, and R^3 for all 60 compounds are listed in Table 1.

The structures of the compounds were drawn using ISIS/Draw as implemented in the ISIS 2.4 package and pre-optimized using Molecular Mechanics Force Fields (MM+). The molecular geometries were refined using AM1 Hamiltonian (Austin Method 1) calculations together with eigenvector following a geometry optimization procedure available in the quantum-chemical program MOPAC 7.05 and implemented in the CODESSA PRO package.[15] The gradient norm limit 0.01 kcal/Å was applied in the geometry optimization.

CODESSA PRO and CODESSA version 2.0[16] packages were used to calculate up to 1627 different molecular (888) and fragmental (739) descriptors, derived solely from molecular structure: (i) constitutional, (ii) geometrical, (iii) topological, (iv) electrostatic, (v) quantum chemical, and (vi) thermodynamic, based on the geometric, electronic, energetic, and thermodynamic characteristics from the MOPAC calculations.

Table 1. List of compounds used in this study—obtained by varying the substituents R^1, R^2, and R^3

Compound	R^1	R^2	R^3
S1	H	H	H
S2	H	$-CH_3$	H
S3	H	$-CH_2CH_3$	H
S4	H	$-CH_2CH_2CH_3$	H
S5	H	–CHO	H
S6	H	$-COCH_3$	H
S7	H	$-COCH_3$	Cl
S8	H	$-COCH_3$	Br
S9	H	$-COCH_3$	I
S10	$-CH_3$	$-COCH_3$	H
S11	H	$-COCH_3$	F
S12	H	$-COCH_3$	–OH
S13	H	$-COCH_2CH_3$	H
S14	H	$-CH_2COCH_3$	H
S15	H	$-COCH_2CH_2CH_3$	H
S16	H	$-COCH(CH_3)_2$	H
S17	H	$-COCH_2F$	H
S18	H	$-COCH_2Cl$	H
S19	H	$-COCH_2CN$	H
S20	H	$-COCH_2NHCOCH_3$	H
S21	H	$-COCH_2OCOCH_3$	H
S22	H	$-COOCH_3$	H
S23	H	–CH(=NOH)	H
S24	H	$-C(=NOH)CH_3$	H
S25	H	$-CH(OH)CH_2CH_3$	H
S26	H	$-CH(OH)CH_2CH_2CH_3$	H
S27	H	$-C(OH)(CH_3)_2$	H
S28	H	$-CH(OH)CH(CH_3)_2$	H
S29	H	$-NO_2$	H
S30	H	$-NH_2$	H
S31	H	$-N(C_2H_5)_2$	H
S32	H	$-OCH_2CH_2CH_2CH_3$	H
S33	H	$-CF_3$	H
S34	H	$-SCH_3$	H
S35	H	$-SCH_2CH_3$	H
S36	H	$-SOCH_3$	H
S37	H	$-SOCH_3$	$-NO_2$
S38	H	$-SOCH_2CH_3$	H
S39	H	$-SOCH_2CH_2CH_3$	H
S40	H	H	F
S41	H	$-SOCH_3$	$-CH_3$
S42	H	$-CH_3$	$-CH_3$
S43	H	$-SO_2NH_2$	H
S44	H	H	$-CH_3$
S45	H	H	Cl
S46	H	H	Br
S47	H	H	$-CH_2CH_3$
S48	H	H	$-OCH_3$
S49	H	H	$-SCH_3$
S50	–(–CH=CH–)$_2$–		H
S51	H	$-CH=CH_2$	H
S52	H	–C≡CH	H
S53	H	$-CH(OH)CH_3$	H
S54	H	$-SCH_2CH_2CH_3$	H
S55	H	$-SO_2CH_3$	H
S56	H	$-COCH_2Br$	H
S57	H	$-N(CH_3)_2$	H
S58	H	$-COCH_3$	$-CH_3$
S59	H	$-COCH_3$	$-NH_2$
S60	H	H	I

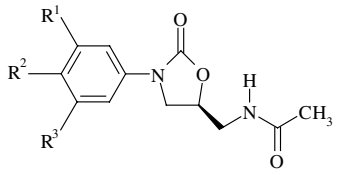

Figure 2. The basic structure of all 3-aryloxazolidin-2-one antibacterials considered.

The best multilinear regression (BMLR) procedure[18] was used to find the best correlation models from selected noncollinear descriptors. The BMLR selects the best two-parameter regression equation, the best three-parameter regression equation, etc., based on the highest r^2 value in the stepwise regression procedure. During the BMLR procedure the descriptor scales are normalized, centered automatically and the final result is given in natural scales. This result has the best representation of the property in the given descriptors pool.

A major decision in developing successive QSPRs is when to stop adding descriptors to the model during the stepwise regression procedure. The lack of an adequate control leads to over-correlated equations, which contain an excess of descriptors and are difficult to interpret in terms of interaction mechanisms. A simple procedure to control the model expansion is the so-called 'break point' in improvement of the statistical quality of the model. By analysis of the plot of the number of descriptors involved in the obtained models versus squared correlation coefficient (and cross-validated squared correlation coefficient) values corresponding to those models, it appears that the statistical parameters of the model improve (steeper ascent of the relationship) up to a certain point ('break point') and after that the improvement is negligible (low ascent of the relationship). Consequently, the model corresponding to the break point is considered the best/optimum model.

The QSPR models obtained were validated (i) by the leave-one-out method, (ii) by internal correlation whereby 1/3 of the compounds is predicted with the model fitted with 2/3 of the compounds (see description below), and (iii) by external validation.

3. Results and discussion

3.1. Full data set of compounds

3.1.1. Correlations using the full set of molecular descriptors. In the general treatment of data, the pool of descriptors included both the whole molecule and the molecular fragment descriptors. Based on the conclusions of Gregory and co-workers,[3,4] the fragment descriptors were defined for two distinct parts of the molecules of the given series: (i) the aryl group A, and (ii) the acetamido group B (see Figs. 1 and 2). The corresponding fragment descriptors (altogether 739) were calculated using CODESSA version 2.0, and imported into CODESSA PRO as external descriptors. As a result, the full initial pool included 1627 molecular descriptors.

The analysis of the plot of correlation coefficients versus number of descriptors (see Fig. 3) indicates the presence of a distinct break point at the 7th descriptor. No significant improvement (<0.02) was registered in the r^2 value after this point.

The corresponding best regression equation with seven parameters is given as follows.

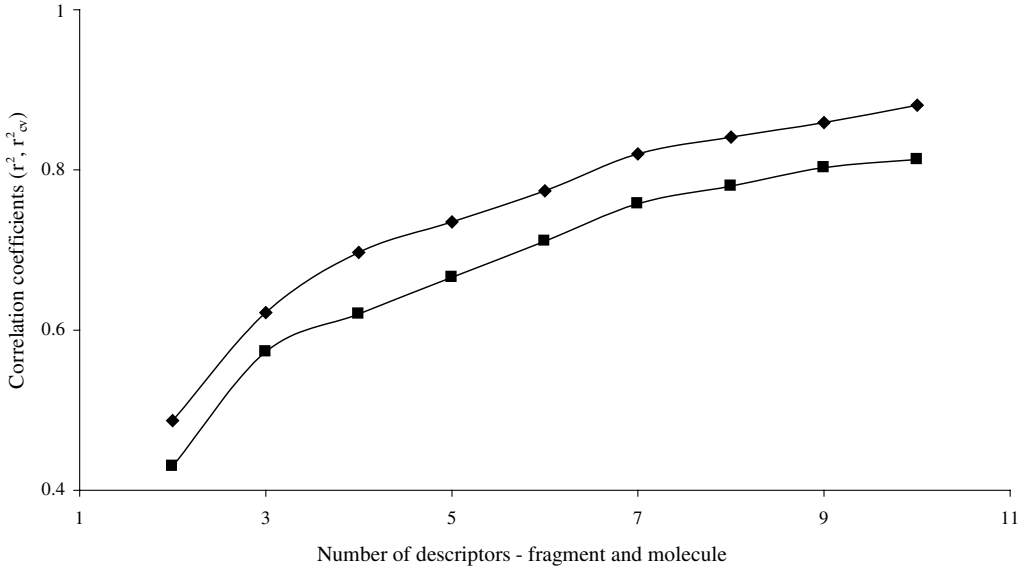

Figure 3. Correlation coefficients r^2 (◆) and r^2_{cv} (■) versus number of descriptors (both the fragment and the whole molecule) for the multilinear QSPR equations developed on the basis of the full data set of 60 compounds.

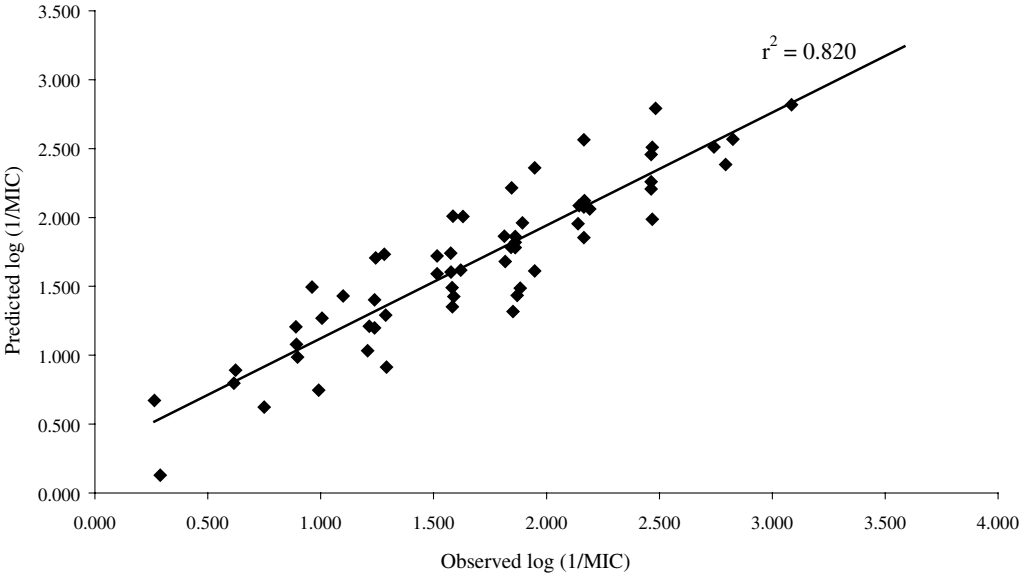

Figure 4. Predicted versus observed log(1/MIC) based on the seven-parameter correlation equation for the full data set and involving both the fragment and the whole molecule descriptors.

$$\begin{aligned}\mathrm{Log}(1/\mathrm{MIC}) &= -164.88 + 151.63\,\mathrm{D}_1 - 123.84\,\mathrm{D}_8 \\ &\quad + 9.76\,\mathrm{D}_{18} + 224.37\,\mathrm{D}_6 - 18.16\,\mathrm{D}_{13} \\ &\quad - 0.65\,\mathrm{D}_4 + 0.03\,\mathrm{D}_{10}, \\ N &= 60,\ \ n = 7,\ \ r^2 = 0.820,\ \ r^2{}_{\mathrm{cv}} = 0.758, \\ F &= 33.77,\ \ s^2 = 0.082 \end{aligned} \quad (1)$$

The respective plot of calculated versus experimental log(1/MIC) data is presented in Figure 4.

The seven descriptors involved in Eq. 1 can be classified as follows: (i) one constitutional (*Relative number of double bonds*, D_{18}), and (ii) six quantum-chemical (*Average bond order of a H atom*, D_1, *HOMO–LUMO energy gap*, D_4, *Minimum* (>0.1) *bond order for atom N*, D_6, *Minimum one-electron reactivity index for atom O*, D_8, *Total molecular one-center electron–electron repulsion*, D_{10}, *Maximum sigma–pi bond order*, D_{13}).

According to the *t*-test values ($|t|$), the importance of the descriptors involved in the model decreases in the following order: $D_1 > D_8 > D_{18} > D_6 > D_{13} > D_4 > D_{10}$.

The most important descriptor in Eq. 1 is *Average bond order of a H atom*, D_1. The quantum-chemical descriptors related to the H atoms indicate the importance of electrostatic interactions in determining the activity of compounds. Three out of seven parameters are quantum chemically calculated bond orders (D_1, D_6, and D_{13}) and contain useful information about the most stable and the weakest bond in the molecule, also characterizing the molecule as a whole using σ and π molecular orbital coefficients.[17] Both D_1 and D_6 have significant positive contributions to the antibacterial activity, expressed as log(1/MIC).

The second descriptor by importance in the model is the *Minimum one-electron reactivity index for atom O*, D_8 that is defined by Eq. 2:

$$R_O = \sum_{i\in O}\sum_{j\in O} c_{i\mathrm{HOMO}} c_{j\mathrm{LUMO}} / (\varepsilon_{\mathrm{LUMO}} - \varepsilon_{\mathrm{HOMO}}) \quad (2)$$

where the summations are performed over all atomic orbitals (AO) *i*, *j* of oxygen atom, $c_{i\mathrm{HOMO}}$ and $c_{j\mathrm{LUMO}}$ denote the *i*th and *j*th AO coefficients for the HOMO (highest occupied molecular orbital) and LUMO (lowest occupied molecular orbital), respectively, and $\varepsilon_{\mathrm{HOMO}}$ and $\varepsilon_{\mathrm{LUMO}}$ are the energies of these orbitals, respectively. This index gives an estimate of the relative reactivity of oxygen atoms in the molecules of antibacterials and can be related to the activation energy of the inhibiting growth of *S. aureus*[17]—the increase of the relative reactivity of oxygen atoms will cause a substantial decrease of log(1/MIC).

The HOMO energy characterizes the susceptibility of the molecule toward the attack by electrophiles whereas the LUMO energy characterizes the susceptibility of the molecule toward the attack by nucleophiles. It is well known that: (i) hard nucleophiles have a low-energy HOMO, (ii) soft nucleophiles have a high-energy HOMO, (iii) hard electrophiles have a high-energy LUMO, and (iv) soft electrophiles have a low-energy LUMO. The *HOMO–LUMO energy gap*, D_4, which is the difference in energy between the HOMO and LUMO, can be related to the chemical stability of compounds. Thus, a large *HOMO–LUMO energy gap* can be considered as an indicator of higher stability of a compound toward a chemical reaction, especially in the case of radical reactions. The nucleophilic and electrophilic reactivities depend on the respective individual HOMO and LUMO energies. The sign of the regression coefficient indicates that the biological activity increases with the decrease of the *HOMO–LUMO energy gap*. It may thus be speculated that the biological counterpart in the antibacterial activity is a soft nucleophile (e.g., –SH or double bond).

The *Total molecular one-center electron–electron repulsion*, D_{10}, describes the electron repulsion-driven processes in the molecule and relates to the conformational changes (rotation, inversion) in the molecule that may have importance in the biological transport of the molecules.

The predicted values of log(1/MIC) using Eq. 1 are given in Table 3.

3.1.2. Correlations using the whole molecule descriptors. The data treatment was also carried out for two limited sets of molecular descriptors. The first of them included only the whole molecule descriptors whereas the second involved only the fragment descriptors. A QSPR model using only the whole molecule descriptors was developed again for the full set of 60 compounds.

The seven-parameter model, comparable to Eq. 1, is presented by Eq. 3 and plotted in Figure 5.

$$\begin{aligned}\mathrm{Log}(1/\mathrm{MIC}) &= 225.81 + 147.45\,D_1 - 1.90\,D_7 \\ &\quad - 0.92\,D_4 - 0.67\,D_3 - 0.27\,D_9 \\ &\quad - 71.55\,D_8 - 2.43\,D_5, \\ N = 60,\ \ n = 7,\ \ r^2 &= 0.795,\ \ r^2_{cv} = 0.727, \\ F = 28.79,\ \ s^2 &= 0.094 \qquad (3)\end{aligned}$$

All seven descriptors involved in Eq. 3 have a quantum-chemical origin (*The average bond order for atom H*, D_1, *HOMO-1 energy*, D_3, *HOMO–LUMO energy gap*, D_4, *Maximum atomic state energy for atom H*, D_5, *Minimum electron–electron repulsion for bond C–O*, D_7, *Minimum one-electron reactivity index for atom O*, D_8, *Total hybridization component of the molecular dipole*, D_9). According to the *t*-test values, the importance of the descriptors involved in the model decreases in the following order: $D_1 > D_7 > D_4 > D_3 > D_9 > D_8 > D_5$. Further analysis of Eq. 3 indicates that the descriptor D_1 has the biggest contribution to the model (a positive one). Also, a significant contribution to the model is given by descriptor D_8, but in this case the biological

Table 2. Descriptors involved in the QSPR models

#	Descriptors	
	Name	Symbol
Whole molecules		
1	Average bond order for atom H	D_1
2	Average nucleophile reactivity index for atom N	D_2
3	HOMO-1 energy	D_3
4	HOMO–LUMO energy gap	D_4
5	Maximum atomic state energy for atom H	D_5
6	Minimum (>0.1) bond order for atom N	D_6
7	Minimum electron–electron repulsion for bond C–O	D_7
8	Minimum one-electron reactivity index for atom O	D_8
9	Total hybridization component of the molecular dipole	D_9
10	Total molecular one-center electron–electron repulsion	D_{10}
Fragments		
11	Average bond order of a C atom (f-aryl)	D_{11}
12	FHBSA Fractional HBSA (HBSA/TFSA) (f-amide)	D_{12}
13	Maximum sigma–pi bond order (f-aryl)	D_{13}
14	Minimum exchange energy for a C–H bond (f-aryl)	D_{14}
15	Minimum nuclear-nuclear repulsion for a C–H bond (f-amide)	D_{15}
16	Moment of inertia A (f-aryl)	D_{16}
17	Number of triple bonds (f-aryl)	D_{17}
18	Relative number of double bonds (f-aryl)	D_{18}

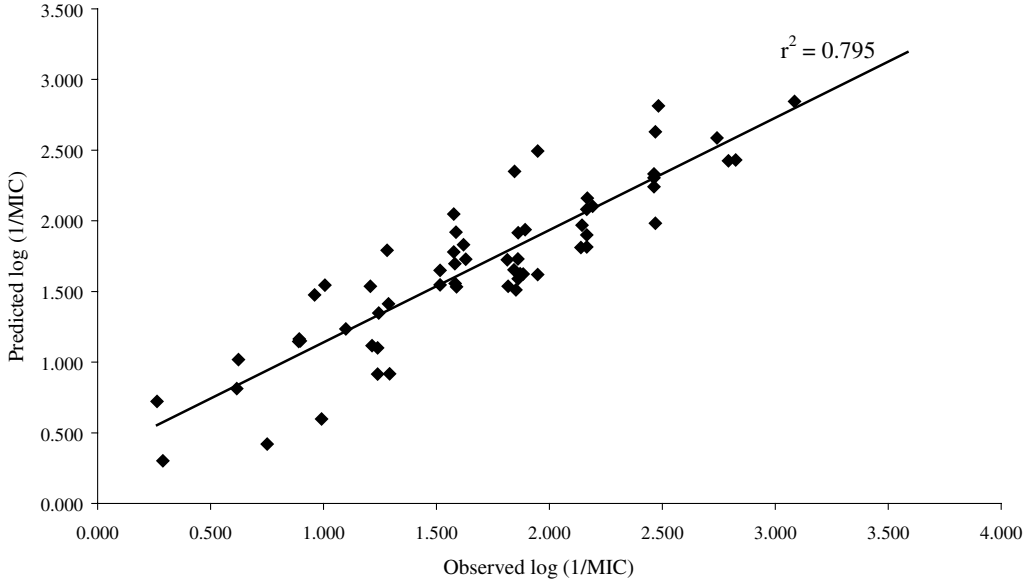

Figure 5. Predicted versus observed log(1/MIC) based on the seven-parameter correlation equation for the full data set and involving only the whole molecule descriptors.

Table 3. The comparison between the experimental and calculated log(1/MIC) values of the molecules—Eqs. 1 and 5—using both the fragment and the whole molecule descriptors

Compound[a]	Log(1/MIC) (exp)	Log(1/MIC) (calc)		Δ (residual)	
		Eq. 1	Eq. 5	Eq. 1	Eq. 5
S1	0.263	0.672	0.365	−0.409	−0.102
S2	0.893	1.078	1.168	−0.185	−0.275
S3	1.817	1.680	1.398	0.137	0.419
S4	2.140	1.955	2.102	0.185	0.038
S5	1.516	1.721	1.741	−0.205	−0.225
S6	2.742	2.511	2.708	0.231	0.034
S7	2.191	2.062	2.055	0.129	0.136
S8	1.948	1.612	1.333	0.336	0.615
S9	1.099	1.431	1.127	−0.332	−0.028
S10	2.463	2.260	2.240	0.203	0.223
S11	2.469	2.509	2.430	−0.040	0.039
S12	2.165	2.564	2.502	−0.399	−0.337
S13	2.463	2.206	2.202	0.257	0.261
S14	1.861	1.780	1.623	0.081	0.238
S15	3.085	2.817	2.849	0.268	0.236
S16	2.483	2.793	2.922	−0.310	−0.439
S17	2.168	2.123	2.168	0.045	0.000
S18	2.793	2.384	2.644	0.409	0.149
S19	1.576	1.741	1.800	−0.165	−0.224
S20	1.620	1.617	1.479	0.003	0.141
S21	2.825	2.568	2.726	0.257	0.099
S22	2.165	1.855	2.223	0.310	−0.058
S23	1.239	1.401	1.452	−0.162	−0.213
S24	1.862	1.861	1.529	0.001	0.333
S25	0.961	1.495	1.505	−0.534	−0.544
S26	1.282	1.733	1.748	−0.451	−0.466
S27	2.165	2.075	1.731	0.090	0.434
S28	1.884	1.486	1.617	0.398	0.267
S29	2.145	2.087	1.988	0.058	0.157
S30	0.289	0.130	0.388	0.159	−0.099
S31	1.582	1.490	2.035	0.092	−0.453
S32	1.583	1.352	1.284	0.231	0.299
S33	1.577	1.604	1.481	−0.027	0.096
S34	1.846	2.215	1.963	−0.369	−0.117
S35	2.469	1.986	1.842	0.483	0.627
S36	1.870	1.434	1.856	0.436	0.014
S37	1.630	2.007	1.734	−0.377	−0.104
S38	1.288	1.291	1.686	−0.003	−0.398
S39	1.006	1.268	1.599	−0.262	−0.593
S40	0.897	0.986	0.825	−0.089	0.072
S41	1.589	1.425	1.814	0.164	−0.225
S42	1.516	1.591	1.527	−0.075	−0.011
S43	0.991	0.747	0.993	0.244	−0.002
S44	0.890	1.206	1.129	−0.315	−0.239
S45	0.623	0.892	0.975	−0.269	−0.352
S46	1.292	0.914	0.831	0.378	0.461
S47	1.215	1.211	1.108	0.004	0.107
S48	0.616	0.797	0.897	−0.181	−0.281
S49	1.244	1.706	1.154	−0.462	0.090
S50	1.852	1.316	1.651	0.536	0.201
S51	1.813	1.864	1.559	−0.051	0.254
S52	1.208	1.032	1.570	0.176	−0.362
S53	1.843	1.783	2.157	0.060	−0.314
S54	1.586	2.009	1.994	−0.423	−0.408
S55	1.893	1.960	2.815	−0.067	−0.922
S56	1.948	2.360	2.134	−0.412	−0.186
S57	1.239	1.199	1.800	0.040	−0.561
S58	2.463	2.458	2.573	0.005	−0.110
S59	1.862	1.819	2.910	0.043	−1.048
S60	0.750	0.623	0.641	0.127	0.109

[a] The first 50 compounds belongs to the training set, and the last 10 compounds to the test set (the case of Eq. 5).

activity decreases when the value of this descriptor is increased.

3.1.3. Correlations using fragment descriptors. The QSPR modeling using only fragment descriptors also

involved all 60 compounds. The seven-parameter model obtained is given by Eq. 4 and shows a significantly poorer correlation in comparison with Eq. 3 obtained using the whole molecule descriptors alone ($\Delta r^2 = 0.064$, $\Delta r^2_{cv} = 0.055$). The corresponding plot of the correlation between predicted and observed values of log(1/MIC) is presented in Figure 6.

$$\begin{aligned}\mathrm{Log}(1/\mathrm{MIC}) &= 4431.37 + 5.59\,D_{11} + 23.95\,D_{12} \\ &\quad + 36.91\,D_{13} - 39.52\,D_{16} - 113.51\,D_{15} \\ &\quad + 4.83\,D_{14} - 0.69\,D_{17}, \\ N &= 60,\ \ n = 7,\ \ r^2 = 0.731,\ \ r^2_{cv} = 0.672, \\ F &= 20.19,\ \ s^2 = 0.123 \end{aligned} \quad (4)$$

Four of the seven descriptors involved in the model of Eq. 4 are quantum chemical (*Average bond order of a C atom*, D_{11}, *Maximum sigma–pi bond order*, D_{13}, *Minimum exchange energy for a C–H bond*, D_{14}, *Minimum nuclear–nuclear repulsion for a C–H bond,* D_{15}). One of the other descriptors is constitutional (*Number of triple bonds,* D_{17}), one is geometrical (*Moment of inertia A*, D_{16}), and another charge-distribution-related (*FHBSA Fractional HBSA (HBSA/TFSA)*, D_{12}). The inspection of the descriptors involved in Eq. 4 reveals that five out of the six fragmental descriptors (D_{11}, D_{13}, D_{14}, D_{16}, D_{17}; see Table 2) relate to the aryl group A in the compounds. This indicates the structural importance of group A (Fig. 1) in determining the variance in the antibacterial activity of oxazolidinones.

The influence of the acetamido group B, considered as the active center of oxazolidinones, is accounted in this model for by two descriptors: D_{12}, which is defined as the difference between the respective hydrogen-bonding donor and hydrogen-bonding acceptor descriptors, and D_{15}, which describes the nuclear-nuclear repulsion energy between two atoms, C and H.

The values of these two fragment descriptors, although they relate to a structurally constant fragment of the molecules throughout the series (acetamido group B), are not constant. This is due to (i) the nature/type of these two parameters, and (ii) the way in which CODESSA PRO software calculates these fragment descriptors. Descriptor D_{12} is charge related and descriptor D_{15} is calculated quantum chemically. The CODESSA PRO program calculates fragment descriptors in real (whole) molecules, and therefore, the descriptors for a constant fragment can depend on the other part of the molecule. In particular, descriptors D_{12} and D_{15} vary with the structure of the compound and even model the property of interest (the antibacterial activity of oxazolidinones in our case).

The poorer correlation using only the fragmental descriptors is not unexpected, because in the multilinear QSAR the interfragmental interactions are not accounted for. In the first approximation, the additional cross-terms involving descriptors from different molecular fragments might reflect these interactions. Such a possibility can be examined in future.

3.2. The model validation

The cross-validated (leave-one-out) correlation of the seven-parameter model (Eq. 1) gives r^2_{cv} value 0.758. For further internal validation, the whole data set was divided into three subsets (the 1st, 4th, 7th, etc. entries go into the first subset, the 2nd, 5th, 8th, etc. into the second subset, and the 3rd, 6th, 9th, etc. into the third subset). In each of three combinations, two of the subsets were combined into one and the correlation equation was derived with the same descriptors. The obtained equation was used to predict data for the remaining subset. The results show that the predicted r^2 values are in good agreement with our original QSPR

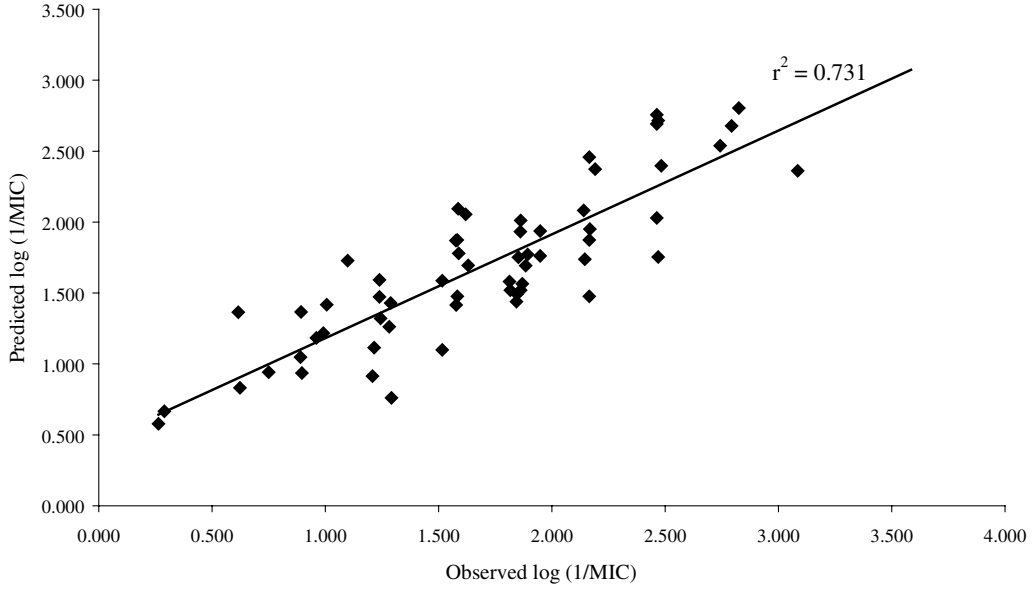

Figure 6. Predicted versus observed log(1/MIC) based on the seven-parameter correlation equation for the full data set and involving only the fragment descriptors.

model with the average correlation coefficients of 0.825 and 0.700 for the fitted and predicted sets, respectively (see Table 4).

For external validation, we selected the same dissection for our training (50 compounds) and test (10 compounds) sets, as did Karki and Kulkarni,[13] to allow a better comparison of the methods.

For these 50 compounds, the best model has six parameters, and is given by Eq. 5, and plotted in Figure 7.

$$\begin{aligned}\mathrm{Log}(1/\mathrm{MIC}) &= 249.58 + 123.28\,D_1 - 1.53\,D_4 \\ &\quad - 1.99\,D_7 + 125.81\,D_2 + 19.43\,D_{12} \\ &\quad - 82.43\,D_8 \\ N &= 50,\ \ n = 6,\ \ r^2 = 0.809,\ \ r^2_{cv} = 0.737, \\ F &= 30.40,\ \ s^2 = 0.096 \qquad (5)\end{aligned}$$

The descriptors involved in this model are: *Average bond order for H atom*, D_1 *HOMO–LUMO energy gap*, D_4, *Minimum electron–electron repulsion for bond C–O*, D_7, *Average nucleophile reactivity index for atom N*, D_2, *FHBSA Fractional HBSA (HBSA/TFSA)*, D_{12}, *Minimum one-electron reactivity index for atom O*, D_8. Here again, the most important descriptor is the *Average bond order for H atom*, followed closely by the *HOMO–LUMO energy gap* and the *Minimum electron–electron repulsion for bond C–O*, according to the *t*-test values. The predicted property values for the training and test sets using Eq. 5 are given in Table 3.

According to Table 3, Eq. 5 shows a poor predictive power for two compounds: S55 and S59, in comparison with Eq. 1. A reason for that could be that compounds S55 and S59 are somehow unique. In fact, none of the compounds from the training set contain as substituent the group $–SO_2$-alkyl (S55), or have amino as the R^3 substituent (S59). This 'lack of information' in the training set could explain the observed discrepancies.

The cross-validated correlation of the model of Eq. 5 gives r^2_{cv} value 0.737. The test set, 10 compounds (not involved to develop the training model) by external validation of Eq. 5 gives an acceptable external correlation coefficient, $r^2 = 0.640$ (Fig. 7).

4. Conclusions

QSPR modeling of in vitro minimum inhibitory concentration (MIC)—required for inhibiting growth of *S. aureus*—with 60 3-aryloxazolidin-2-one antibacterials has been performed using the so-called 'CODESSA PRO

Table 4. Internal validation of the seven-parameter QSPR model for 60 antibacterials, Eq. 1

Set to fit	r^2 (fit)	s^2 (fit)	Set to predict	r^2 (pred)	s^2 (pred)
#1 and #2	0.756	0.056	#3	0.855	0.066
#1 and #3	0.865	0.061	#2	0.514	0.087
#2 and #3	0.854	0.053	#1	0.731	0.117
Average	0.825	0.057		0.700	0.090

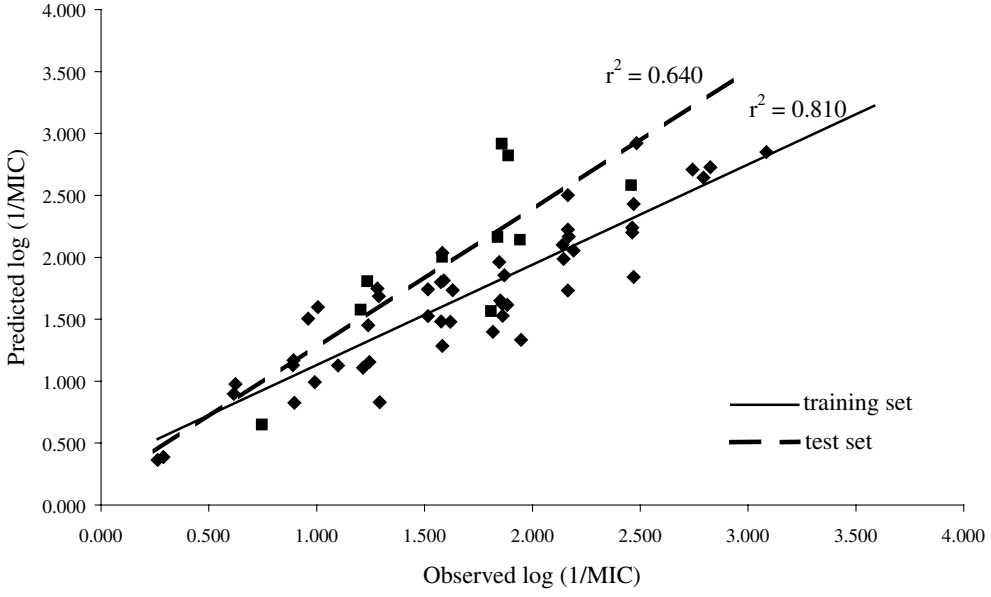

Figure 7. Predicted versus observed log(1/MIC) based on the six-parameter model (Eq. 5) for training set (◆) and test set (■), and involving both the fragment and the whole molecule descriptors.

technique'. Two different programs—CODESSA PRO and CODESSA version 2.0—were used to calculate various molecular (888) and fragmental (739) descriptors providing a total of 1627 descriptors, derived solely from molecular structure. These include constitutional, geometrical, topological, electrostatic, quantum chemical, and thermodynamic descriptors, based on the geometric, electronic, energetic, and thermodynamic characteristics from MOPAC calculations. The multilinear QSPR models have been obtained using both molecular and fragmental descriptors.

The best models selected at the training stage for the full set of compounds (Eq. 1) were verified by using different validation methods that confirmed the correct predictions of minimum inhibitory concentration (MIC) for the antibacterials.

In conclusion, our results also demonstrate that in characterizing a complex biological property (antibacterial activity) by multilinear QSPR equation, the descriptors related to the whole molecule can provide superior model (cf. Eq. 3) versus that (Eq. 4) obtained using just fragmental descriptors. However, the best model of all (Eq. 1) utilized both fragmental and molecular descriptors.

Acknowledgements

The authors thanks to Prof. Alexandre Varnek for providing valuable comments that enhanced the quality of this study.

Estonian Science Foundation (grant no 4548) is gratefully acknowledged for the partial support of this work.

References and notes

1. Service, R. F. *Science* **1995**, *270*, 724.
2. Selvakumar, N.; Srinivas, D.; Khera, M. K.; Kumar, M. S.; Mamidi, R. N. V. S.; Sarnaik, H.; Charavaryamath, C.; Rao, B. S.; Raheem, M. A.; Das, J.; Iqbal, J.; Rajagopalan, R. *J. Med. Chem.* **2002**, *45*, 3953.
3. Gregory, W. A.; Brittelli, 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.
4. Gregory, W. A.; Brittelli, D. R.; Wang, C.-L. J.; Kezar, 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.
5. Park, C.-H.; Brittelli, 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.
6. Pae, A. N.; Kim, H. Y.; Joo, H. J.; Kim, B. H.; Cho, Y. S.; Choi, K. I.; Choi, J. H.; Koh, H. Y. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2679.
7. Pae, A. N.; Kim, S. Y.; Kim, H. Y.; Joo, H. J.; Cho, Y. S.; Choi, K. I.; Choi, J. H.; Koh, H. Y. *Bioorg. Med. Chem. Lett.* **1999**, *9*, 2685.
8. Tokuyama, R.; Takahashi, Y.; Tomita, Y.; Suzuki, T.; Yoshida, T.; Iwasaki, N.; Kado, N.; Okezaki, E.; Nagata, O. *Chem. Pharm. Bull.* **2001**, *49*, 347.
9. Tokuyama, R.; Takahashi, Y.; Tomita, Y.; Tsubouchi, M.; Yoshida, T.; Iwasaki, N.; Kado, N.; Okezaki, E.; Nagata, O. *Chem. Pharm. Bull.* **2001**, *49*, 353.
10. Tokuyama, R.; Takahashi, Y.; Tomita, Y.; Tsubouchi, M.; Iwasaki, N.; Kado, N.; Okezaki, E.; Nagata, O. *Chem. Pharm. Bull.* **2001**, *49*, 361.
11. Lee, C. S.; Allwine, D. A.; Barbachyn, M. R.; Grega, K. C.; Dolak, L. A.; Ford, C. W.; Jensen, R. M.; Seest, E. P.; Hamel, J. C.; Schaadt, R. D.; Stapert, D.; Yagi, B. H.; Zurenko, G. E.; Genin, M. J. *Bioorg. Med. Chem.* **2001**, *9*, 3243.
12. Phillips, O. A.; Udo, E. E.; Ali, A. A. M.; Al-Hassawi, N. *Bioorg. Med. Chem.* **2003**, *11*, 35.
13. Karki, R. G.; Kulkarni, V. M. *Bioorg. Med. Chem.* **2001**, *9*, 3153.
14. Gopalakrishnan, B.; Khandelwal, A.; Rajjak, S. A.; Selvakumar, N.; Das, J.; Trehan, S.; Iqbalc, J.; Kumarb, M. S. *Bioorg. Med. Chem.* **2003**, *11*, 2569.
15. www.codessa-pro.com.
16. www.semichem.com/codessa.html.
17. Karelson, M. *Molecular Descriptors in QSAR/QSPR*; J. Wiley & Sons: New York, 2000.
18. Katritzky, A. R.; Ignatchenko, E. S.; Barcock, R. A.; Lobanov, V. S.; Karelson, M. *Anal. Chem.* **1994**, *11*, 1799.