CROATICA CHEMICA ACTA
CCACAA **81** (4) 579–592 (2008)
ISSN-0011-1643
CCA-3277
Original Scientific Paper

A Priori Descriptors in QSAR: a Case of Gram-Negative Bacterial Multidrug Resistance to β-Lactams

Rudolf Kiralj* and Márcia M. C. Ferreira

Instituto de Química, Universidade Estadual de Campinas, Campinas, SP 13083-970, Brazil

RECEIVED DECEMBER 15, 2006; REVISED OCTOBER 5, 2007; ACCEPTED JANUARY 31, 2008

Keywords
• gram-negative bacteria
• multidrug resistance
• β-lactam antibiotics
• quantitative structure-activity relationships
• exploratory analysis
• chemometrics

Over hundred new *a priori* global/local molecular descriptors that encoded steric, topological, electronic, hydrogen bonding, compositional and hydrophobic properties were generated for 16 β-lactams, and two partial least squares regression models were constructed and cross-validated. These *a priori* models ($Q^2 > 0.80$, $R^2 > 0.95$, SEV < 0.50) are comparable with the previously obtained computed models. β-Lactam intramolecular and β-lactam-receptor intermolecular interactions are also discussed in terms of molecular descriptors.

INTRODUCTION

Quantitative Structure-Activity Relationships (QSAR),[1–3] Quantitative Structure-Property Relationships (QSPR)[4,5] and Linear Free Energy Relationships (LFER)[6,7] are quantitative relationships between a dependent variable and independent variables called molecular descriptors. Representations of molecules are sources of information for generation of the descriptors at different structural levels: 1D (chemical composition), 2D (chemical formula), 3D (atomic coordinates) or higher structural levels[8,9] such as those in ND-QSARs.[10,11]

A simple approach for chemical interpretation of descriptors is the *a priori* approach with *a priori* descriptors, as postulated by present authors a few years ago.[8,11] *A priori* descriptors are known-before-computer-assistance since they can be easily made, by employing 1D or 2D chemical formula, chemical knowledge and a minimum amount of literature data. The *a priori* approach does not mean using several tabulated descriptors (as in the Hansch-Fujita approach[12]), codes for molecular fragments (as in Free-Wilson approach[13]) or exclusively topological descriptors. The emphasis of the *a priori* approach is on intuitive generation of molecular descriptors.

A priori molecular descriptors have been successfully applied to QSAR of HIV-1 protease inhibitors[8,11] with PLS (Partial Least Squares)[1] regression, and have been combined with computed descriptors in QSAR studies of β-lactam antibiotics[14] and progestogens.[3,15] *A priori* chemical bond descriptors have been used in PLS studies of planar benzenoid hydrocarbons[16] and nucleobases.[17] Encouraged by these results, we present a new application of the *a priori* approach to 16 β-lactam antibiotics (Figure 1) and correlate *a priori* molecular descriptors with biological activities by establishing PLS models comparable to earlier results.[14] The emphasis of this work is in extending the *a priori* approach to microbial resistance phenomena and to represent the advantages of this ap-

* Author to whom correspondence should be addressed. (E-mail: rudolf@iqm.unicamp.br)

proach (especially in terms of model interpretability) while not necessarily obtaining QSAR models better than in the literature. Exploratory analysis of *a priori* data by means of Principal Component Analysis (PCA)[1] and Hierarchical Cluster Analysis (HCA)[1] is performed to give more insight into the chemical background of the descriptors. Biological activities are Minimal Inhibitory Concentration (MIC) values[18] for β-lactams that are extruded from Gram-negative bacterial cells by means of a three-component multidrug resistance efflux pump AcrAB-TolC[19–21](Figure 2). Efflux pumps[22–25] are one of the most important Multidrug Resistance (MDR)[26,27] intrinsic mechanisms in all cellular microbes and cancer cells as effective defense systems against a large variety of drugs and other structurally non-related xenobiotics. Recent studies on bacterial MDR efflux pumps[22–25,28–31] and crystal structures of AcrB,[32–36] AcrA[37] and TolC[38,39] point out non-specific interactions between drugs and AcrAB-TolC as a good example for drug-pump interactions in bacteria. β-Lactam molecules in bacterial periplasm or cytoplasm interact with the inner membrane or AcrB pump and are transferred to the openings called vestibules. AcrAB-TolC is trimeric (the dimer is shown in Figure 2 for clarity), where the cylinder-like TolC is docked to the jelly-fish shaped AcrB, and AcrA additionally connects AcrB and TolC together and with the membranes. AcrB has three vestibules with characteristic BRAMLA shape (BRAzil Map-Like Area[14,40]) that meet in the central cavity where drug molecules are collected. The nearest cellular ATP hydrolysis causes proton influx that provokes conformational changes in the pump and its components like the opening of the AcrB-TolC channel through which the drugs are excreted. Drug molecular properties are important for drug interaction with this efflux system.

Penicillins and cephalosporins are the most important β-lactam antibiotics.[41] Physico-chemical[42–45] and theoretical[14,46,47] studies have pointed out the amphiphi-

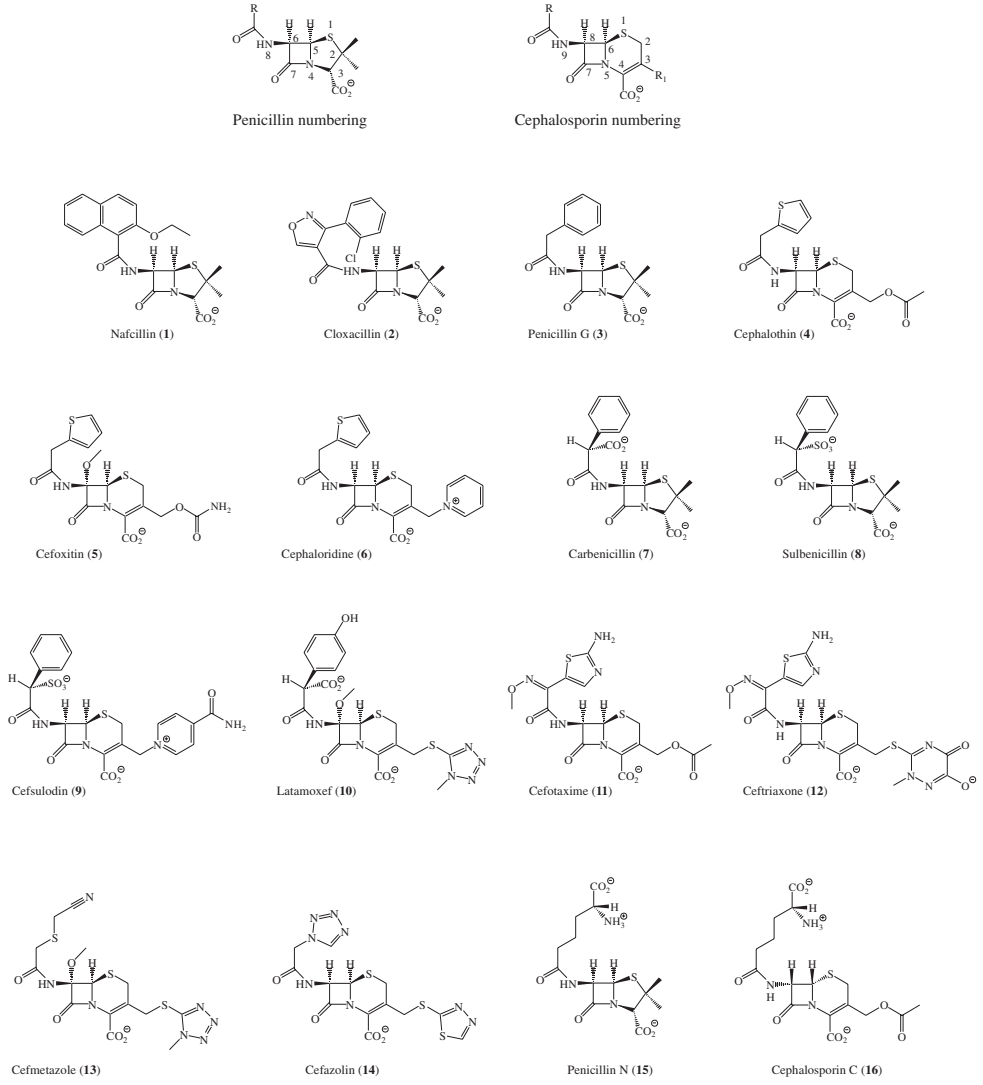


Figure 1. Structures of β-lactams at neutral pH, with atomic numbering for penicillins and cephalosporins.

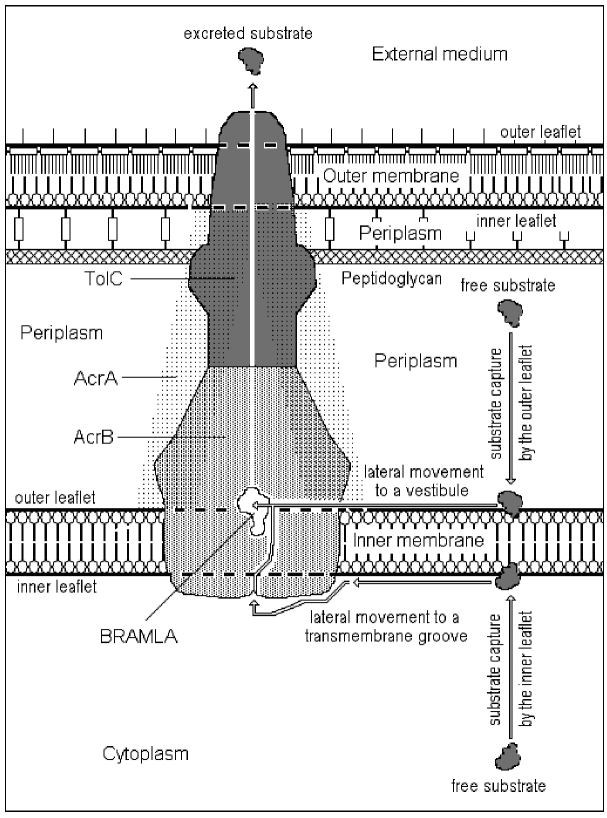


Figure 2. AcrAB-TolC efflux pump. TolC is docked to AcrB. Only one vestibule is visible in this orientation, while the other two are placed in the back side of the AcrB trimer, at the joint lines of the monomers. Arrows show the substrate efflux pathway starting from the periplasm and cytoplasm.

lic character of these compounds and the majority of antibiotics, due to which all of them behave as classical detergents or surfactants. This includes drug concentration at the hydrophobic-hydrophilic interface, self-association, interaction with biomembranes and proteins,[42] penetration through cellular membranes and intestinal absorption. Most drugs, wheather natural or synthetic, are amphiphilic,[42] as are amino-acids and proteins,[48,49] because of which protein folding and protein interactions with diverse kinds of molecules occur. AcrAB-TolC belongs to the Resistance Nodulation Division (RND)[22–25] family of transporters whose substrates are rather amphiphilic. However, protein-drug interactions are rarely described in QSAR by drug amphiphilic descriptors.[49] In this work, the nature of β-lactam descriptors including amphiphilicity, is highlighted.

METHODS

MICs were determined as mass concentrations for molecular and ion/zwitterion species (Figure 1) by Nikaido *et al.*[18] at neutral pH, as excreted by strains SH5014 (parent strain) and HN891 (AcrAB overproducer) of pathogen bacteria *Salmonella typhimurium*. pMICs are defined as pMIC = –log [MIC/(mol dm^{-3})]. Molecular descriptors were generated employing chemical compositions and structural formulae of β-lactams (Figure 1) and some atomic constants. A few descriptors were from the previous study.[14]

Descriptors and biological activities were autoscaled, *i.e.*, reduced by the respective mean values and then divided by the respective variances, prior to every chemometric analysis due to differences in their orders of magnitude. Three types of chemometric methods[1] were employed in this work: Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA) and Partial Least Squares (PLS) regression. PCA is a data compression method, in which original molecular descriptors with usually substantial intercorrelations are linearly combined into principal components (PCs) which are mutually orthogonal. The PCs are arranged in decreasing order of the total variance contents, so that only a few PCs are sufficient to visualize and explain the original data. A preprocessed data matrix $\mathbf{X}(I\times J)$ for n molecules and m descriptors is decomposed into two matrices **T** and **L** such that $\mathbf{X} = \mathbf{TL}^{\mathrm{T}}$, where **T** is the scores matrix, representing positions of compounds in the new coordinate system. **L** is the loadings matrix, *i.e.*, the transformation matrix connecting original descriptors and PCs. HCA is another, two-dimensional way to present the original data in form of the dendograms. Euclidean distances d_{ij} between compounds i and j are calculated using elements of **X**, resulting in the similarity matrix **S**. The same procedure can be applied to descriptors. Elements of **S** are similarity indices defined as $S_{ij} = 1 - d_{ij}/d_{\max}$, where $d_{\max}$ is the largest distance between any pair of compounds or descriptors.

In PLS regression, the matrix **X** of molecular descriptors is correlated with the vector **y** of biological activities. The regression equation $\mathbf{y} = \mathbf{Xb}$ is solved by maximizing the variance between **y** and **T**, which results in k principal components (latent variables). The PLS regression model is built by using k essential principal components. The model is internally validated by leave-one-out crossvalidation, in which every compound is predicted from a regression that was built without using the data for this compound. The following statistical parameters are important to access the goodness of a PLS model:

a) standard error of calibration $\mathrm{SEC} = [\Sigma_i\ (y_{ei} - y_{ci})^2/(n - k - 1)]^{1/2}$,
b) standard error of leave-one-out crossvalidation $\mathrm{SEV} = [\Sigma_i\ (y_{ei} - y_{vi})^2/n]^{1/2}$,
c) correlation coefficient of calibration $R^2 = 1 - [\Sigma_i\ (y_{ei} - y_{ci})^2]/[\Sigma_i\ (y_{ei} - \langle y_{ei}\rangle)^2]$,
d) correlation coefficient of leave-one–out crossvalidation $Q^2 = 1 - [\Sigma_i\ (y_{ei} - y_{vi})^2]/[\Sigma_i\ (y_{ei} - \langle y_{ei}\rangle)^2]$,
e) linear correlation coefficient of calibration $R = [\Sigma_i\ (y_{ei} - \langle y_{ei}\rangle)(y_{ci} - \langle y_{ci}\rangle)]\ /\ [\Sigma_i\ (y_{ei} - \langle y_{ei}\rangle)^2]^{1/2}\ [\Sigma_i\ (y_{pi} - \langle y_{pi}\rangle)^2]^{1/2}$,

f) linear correlation coefficient of leave-one-out cross-validation

$$Q = [\Sigma_i (y_{ei} - \langle y_{ei} \rangle)(y_{vi} - \langle y_{vi} \rangle)] / [\Sigma_i (y_{ei} - \langle y_{ei} \rangle)^2]^{1/2} [\Sigma_i (y_{vi} - \langle y_{vi} \rangle)^2]^{1/2}$$

where:

i is the summation index; y_e are experimental values of $\boldsymbol{y}$; y_c and y_v are values of $\boldsymbol{y}$ obtained from calibration and crossvalidation, respectively; and $\langle y_e \rangle$, $\langle y_p \rangle$ and $\langle y_v \rangle$ are mean values of y_e, y_p and y_v, respectively. It is important to note that squares $(R)^2$ and $(Q)^2$ of linear correlation coefficients R and Q, although numerically close, are not equal to correlation coefficients R^2 and Q^2.

Variable selection for the PLS model for strain HN891 was started by an initial cut-off in linear correlation coefficients for descriptor-activity correlations (absolute values < 0.600). The second part of variable selection consisted of systematic manual variable elimination, a procedure that was aided by the following items: a) Descriptors with pronounced dispersion and non-uniform distribution of data (for example, chance correlations and false non-linearity) in descriptor-activity correlograms were eliminated; b) Descriptors from the same clusters in HCA dendograms and PCA loadings plot were eliminated as much as possible; c) Descriptors with small contribution to the PLS regression vector were eliminated; d) Descriptors whose presence resulted in worsened PLS statistics than when excluded from the model were also eliminated; e) Descriptors of the same type (like steric, topological, electronic, hydrogen bonding, compositional and combined) were eliminated as much as possible; f) Descriptors generated in the same way were eliminated as much as possible; g) Descriptors that were complex to obtain and interpret were eliminated as much as possible. pMIC(HN891) and pMIC(SH5014) are highly correlated (linear correlation coefficient is 0.980),[14] which justified the use of the same descriptors in the two final PLS models. External validation was performed by excluding β-lactams **7**, **10** and **14**, which represent broad activity ranges and behave differently in exploratory analyses from previous work.[14]

Exploratory analysis (PCA and HCA with incremental linkage) related to strain HN891 was performed using the same set of descriptors (data set A). Two sets of descriptors, with moderate to high values of absolute correlation coefficients (> 0.600) with the two pMICs were used in an additional PCA analysis (data set B related to HN891 activities, and data set C related to SH5014 activities). All chemometric analyses were performed by using programs Matlab 5.2[50] and Pirouette 3.02.[51]

RESULTS AND DISCUSSION

Molecular Descriptors

Application of the *a priori* approach to the set of 16 β-lactams (Figure 1) resulted in 105 new molecular descriptors, which, together with 21 descriptors from the earlier study,[14] formed the complete data set. Table I contains descriptors that have absolute correlation coefficients with one or both pMICs (for HN891 and SH5014 strains) above 0.600. Compositional (CM), steric (ST), electronic (EL), hydrogen bonding (HB), topological (TP), hydrophobic (HP) and complex descriptors were generated by considering mainly atoms, bonds, electrons and other structural units, and sometimes from the literature (Pauling atomic electronegativities) and chemical knowledge (fragment characteristics based on well-known atomic, group or element properties). Descriptors were generated for β-lactam molecules/ions (global descriptors) and their side chains (local descriptors). Some were made as a function of one or more local or global descriptors: Gaussian transforms, sums, differences, ratios and functions normalized by the number of structural units (atoms, bonds, *etc.*), among others. All topological descriptors were counted as Wiener or Randić indices of the zeroth or first order,[31] their transforms or rational functions.

It is important to comment about the definition of molecular fragments. The central molecular fragment was considered as a fragment with no or little structural variation: the β-lactam ring and small exocyclic groups (carboxy, carbonyl and methoxy in **5**, **10** and **13**). Therefore, side chains (two variable fragments) are as defined in Figure 1: the left fragment always contains substituent R, and the right fragment includes R_1 in cephalosporins and its analogue C2-Me_2 in penicillins. The left fragment, depending on descriptor definitions (Table I), includes only R, R-CO-N8 (penicillins) or R-CO-N9 (cephalosporins), R-CO-N8-C6 (penicillins) or R-CO-N9-C8 (cephalosporins). The right fragment in cephalosporins is defined most frequently as C3-R_1, and sometimes as R_1. For calculation of Wiener and Randić indices of the zeroth order, the variable fragments included chemical bonds with the central fragment: N8-C6 (penicillins) or N9-C8 (cephalosporins) for the left fragment, and C2-S1 and C2-C3 (penicillins) or C2-C3 and C3-C4 (cephalosporins). Figures 3–5 illustrate generation of the most important molecular descriptors for a good (**2**) and bad (**12**) substrate of the MDR pump AcrAB-TolC. N_{ar1} can be obtained as the number of non-H atoms in the right fragment CR_1/CMe_2 as defined in Figure 3a (3 for **2** and 12 for **12**). **2** and **12** are in anionic and dianionic forms at neutral pH, respectively. From their respective formulas $[C_{19}H_{17}ClN_3O_5S]^-$ and $[C_{18}H_{16}N_8O_7S_3]^{2-}$ 29 and 52 atoms, and 152 and 190 valence electrons are counted. This gives 3.30 and 3.65 valence electrons per atom (V_{av}). Figure 3b shows hydrogen bond (HB) donor and acceptor atoms in **2** and **12**. Counting these atoms (N_{DA}) and dividing them by the total number of non-hydrogen (non-H) atoms in the anions, one gets the number fraction of HB donors and acceptors w_{DA}: 7/29 = 0.241 for **2** and 14/36 = 0.389 for **12**. Figure 3c illustrates the definition of the valence elec-

tron contents of non-H atoms in the side chains, relative to carbon (Z). Anionic/cationic charges are ignored and, therefore, one counts these numbers of valence electrons and atoms: a) for **2**: 58 electrons and 13 atoms from the left fragment, and 12 electrons and 3 atoms from the right fragment; b) for **12**: 47 electrons and 10 atoms from the left fragment, and 53 electrons and 11 atoms from the right fragment. Finally, simple calculation gives $Z = v(\mathrm{R}) + v(\mathrm{R_1}') - 8 = 58/13 + 12/3 - 8 = 0.462$ for **2** and $Z = 47/10 + 53/11 - 8 = 1.518$ for **12**. Descriptor L, the number of non-H atoms along the longest chain in a molecule, includes 4 atoms of six- or five-membered rings in the side chains, as defined in Figure 3d. L is equal to 13 for **2**, and to 20 for **12**. The number of non-H atoms, when divided by the number of domains of distinct hydrophobic/hydrophilic character (illustrated in Figures 4a and 4b), gives the average size of domains D_{av}, being equal to 29/6 = 4.8 for **2** and 36/13 = 20 for **12**. W_{sd} is calculated from the number of non-H atoms and the first order Wiener index for the extended left and right chains (W, N_{r} and W_1, N_{r1}, respectively), as defined in Figure 5a. Calculation of the Wiener index for the right fragment of **12** is shown in Figure 5b, with arbitrarily numbered atoms and the topological distance matrix (matrix with diagonal symmetry), partial sums and the total sum W_1. Final calculations give $W_{\mathrm{sd}} = [W/(N_{\mathrm{r}})^3 - W_1/(N_{\mathrm{r1}})^3]^2 = [492/17^3 - 4/3^3]^2 = 0.002304$ for **2** and $W_{\mathrm{sd}} = [306/14^3 - 196/12^3]^2 = 0.000004$ for **12**.

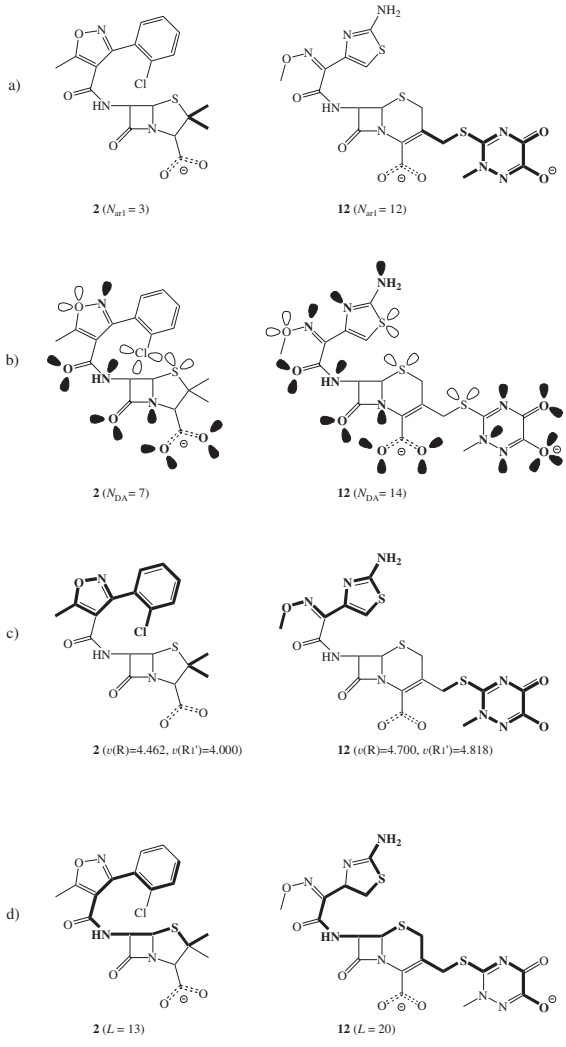


Figure 3. Definition and calculation of some molecular descriptors for **2** and **12**: a) number of non-hydrogen atoms in the right fragment CR_1/CMe_2 (N_{ar1}); b) number of hydrogen bond donors and acceptors (N_{DA}); c) electron contents of the left (v(R)) and right ($v(R_1)$) fragments without hydrogen atoms; d) number of non-hydrogen atoms along the path from the R_1 end to the R end (L). Hydrogen bond donors and acceptors are marked bold and with lone pairs colored black.

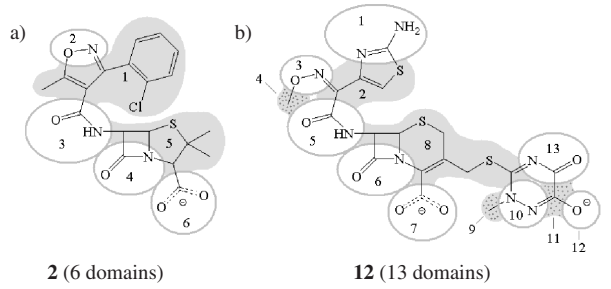
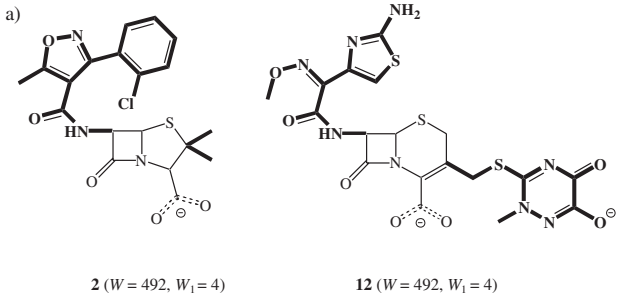
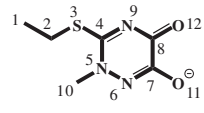


Figure 4. The number and distribution of hydrophobic (gray tones) and polar (white ellipses) domains in a) **2** and b) **12**. Domain numbering starts from the beginning of R and terminates at the end of R_1.





	1	2	3	4	5	6	7	8	9	10	11	12
1	0	1	2	3	4	5	6	5	4	5	7	6
2		0	1	2	3	4	5	4	3	4	6	5
3			0	1	2	3	4	3	2	3	5	4
4				0	1	2	3	2	1	2	4	3
5					0	1	2	3	2	1	3	4
6						0	1	2	3	2	2	3
7							0	1	2	3	1	2
8								0	1	4	2	1
9									0	3	3	2
10										0	4	5
11											0	3
12												0
Sum:	0	3	6	10	15	21	20	18	27	37	38	

Total sum: W_1 = 196
No. atoms: N_{r1} = 12
W_{n1} = 0.1134

Figure 5. Definition and calculation of the topological descriptor W_{sd}. a) Definition of the left and right fragments for the first order Wiener index. b) Calculation of W_1 and normalized W_{1n} for the right fragment of **12**, with arbitrary numbering of non-hydrogen atoms.

PLS Models

Seven molecular descriptors (bold in Table I) were selected for PLS models to predict pMIC(HN891) and pMIC(SH5014). Selected data are in Table II. The two *a priori* models (Table III) have the same number of β-lactams and descriptors and similar statistics as analogous computed models from the earlier work.[14] The computed models used only three *a priori* descriptors, a quantum-chemical descriptor and three lipophilic descriptors.

External validation results for the *a priori* PLS models are in Table III. The two PLS models are similar due to high correlation between pMIC(HN891) and pMIC (SH5014). According to their basic statistics ($Q^2 > 0.81$, $R^2 > 0.95$, SEV < 0.50, 4 PCs), acceptable relative errors

TABLE I. Molecular descriptors used in exploratory analyses and PLS models

No.[a]	Symbol[b]	Definition	Nature[c]	HN891[d]	SH5014[d]
6	N_{ar1}	**number of non-H atoms in CR_1/CMe_2 fragment**	CM/ST/HP	0.647	0.558
8	N_{br1}	number of bonds in CR_1/CMe_2 (H-atoms excluded)	ST	0.614	0.523
9	N_{srr1}	function $N_{ar}+N_{ar1}$; N_{ar} is the number of non-H atoms in RCON	ST	–0.693	–0.600
10	N_{qrr1}	function N_{ar}/N_{ar1}	ST	–0.744	–0.655
11	B_{srr1}	function $N_{br}+N_{br1}$; N_{br} and N_{br1} are numbers of non-H atoms in CR_1/CMe_2 and RCON fragments, respectively	ST	–0.678	–0.596
12	B_{qrr1}	function N_{br}/N_{br1}	ST	–0.755	–0.675
15	w_C[(*)]	number fraction of hydrophobic carbon atoms (all C atoms except those in C=O, C-O- and CN groups)	CM/HP	–0.764	–0.739
21	P_{av}	average Pauling atomic electronegativity	EL	0.745	0.752
23	P_r	average Pauling atomic electronegativity of RCON	EL	0.703	0.703
24	P_{rav}	function $(P_r+P_{r1})/2$; P_{r1} is average Pauling electronegativity of CR_1/CMe_2	EL	0.743	0.744
26	P_{rp}	function P_r*P_{r1}	EL	0.747	0.748
29	N_{pol}	number of polar non-H atoms (C from C=O, C-O^-, CN; S from SO_3; Cl; N)	CM/EL/HB	0.723	0.674
30	K	ratio of numbers of hydrophobic and polar atoms	CM/HB/HP	–0.693	–0.676
31	w_{hyd}	number fraction of non-H hydrophobic atoms	CM/HP	–0.696	–0.667
32	w_{h2}	number fraction of hydrophobic atoms: amphiphilic atoms are O, Cl and S in C-Cl, C-O-C and C-S-C bonds: each such atom is added to hydrophilic C atoms counted for w_C	CM/HP	–0.737	–0.704
33	N_{pol2}	number of polar atoms expressed as $(1–w_{h2})N_{nh}$; N_{nh} is the number of non-H atoms	HB/HP/EL	0.742	0.690
34	D	number of hydrophobic and polar domains	HB/HP/EL	0.727	0.680
35	D_{av}	**average size of domains counted for D, NnH/D**	ST/HP	–0.783	–0.738
36	$N_{rb\text{-}tr}$	function $(N_{rb}–5)^2$; N_{rb} is the number of rigid bonds in R, including double, aromatic and delocalized bonds	ST/HP/EL	–0.556	–0.609
39	V_{av}	**number of valence electrons per atom**	EL	0.656	0.676
41	N_{HB}[(*)]	number of hydrogen bonds, (number of HB donors+acceptors)	HB	0.619	0.561
42	N_{hba}[(*)]	number of hydrogen bond acceptor bonds (number of H bonds originated from HB acceptors = number of lone pairs in N atoms + in carbonyl, oxide, sulfonate and hydroxyl O atoms)	HB	0.638	0.596
44	N_{DA}[(*)]	number of heteroatoms which are HB donors/acceptors	HB/EL	0.727	0.682
45	w_{DA}[(*)]	**number fraction of HB donors and acceptors**	HB	0.686	0.657
46	Z[(*)]	**function v(R)+v(R_1')–8; v(R), v(R_1') are average numbers of valence electrons in R and R_1', respectively; R_1' = R_1/CMe_2; H atoms and +/– charges excluded**	EL	0.703	0.664
47	N_{ns}[(*)]	number of non-σ valence electrons (π-electrons and lone pair electrons; sulfur in β-lactam ring is excluded)	EL/HP/HB	0.628	0.596
48	w_{het}[(*)]	number fraction of heteroatoms	CM/EL/HB	0.772	0.751
49	N_{het}[(*)]	number of heteroatoms (all N, O, S, Cl)	CM/EL	0.749	0.710

50	R_1	average number of valence electrons in substituent CR_1/CMe_2	EL	0.666	0.651
52	$N_{NS}^{(*)}$	number of N and S atoms (except S in sulfonate groups)	CM/EL/HB	0.713	0.680
53	N_N*	number of nitrogen atoms	CM/EL/HB	0.710	0.670
54	w_N	number fraction of nitrogen atoms	CM/EL/HB	0.677	0.650
55	w_{NS}	number fraction of N and S atoms	CM/EL/HB	0.663	0.629
58	L	**number of non-H atoms in the shortest path from the R end to the R_1 end**	CM/ST/HP	0.695	0.601
59	w_L	number fraction of non-H atoms counted for L	CM/ST/HP	0.621	0.532
61	$L_{av\text{-}tr}$	function $\exp[-(L_{av}-2.10)^2]$; L_{av} is average number of domains counted for D along chain defined by L	HB/HP/EL	–0.620	–0.630
62	D_2	number of domains along chain defined by L; small side chains along the chain are excluded	HB/HP/EL	0.681	0.585
78	N_{r1}	number of non-H atoms in CR_1/CMe_2	CM/ST	0.647	0.558
79	D_{rr1}	function N_r-N_{r1}; N_r is the number of non-H atoms in RCONC	CM/ST	–0.675	–0.598
80	Q_{rr1}	function N_r/N_{r1}	CM/ST	–0.734	–0.656
83	W_{sd}	**function $(W_n-W_{n1})^2$; $W_n=W/(N_r)^3$, $W_{n1}=W_1/(N_{r1})^3$; W, W_1 is the first order Wiener index for RCONC and C_3R_1/C_3Me_2 extended fragment, respectively**	TP	–0.774	–0.713
84	Q_{wrr1}	function W/W_1	TP	–0.768	–0.705
88	Q_{2wrr1}	function W_{no}/W_{no1}; $W_{no}= W/(N_r)^2$; $W_{no1}= W_1/(N_{r1})^2$	TP	–0.655	–0.584
93	V_{r1}	zeroth-order extended Wiener index for CR_1/CMe_2	TP	0.639	0.548
94	K_{r1}	zeroth-order Randić index for CR_1/CMe_2	TP	0.644	0.559
96	K_{r1av}	K_{r1} averaged per atom in the fragment	TP	–0.642	–0.548
97	V_d	function V_r–V_{r1}; V_r is zeroth-order Wiener index for RCON	TP	–0.697	–0.613
98	V_q	function V_r/V_{r1}	TP	–0.753	–0.672
99	K_d	function K_r–K_{r1}	TP	–0.642	–0.576
100	K_q	function K_r/K_{r1}	TP	–0.710	–0.638
101	K_{avd}	function K_{rav}–K_{r1av}; K_{rav}, K_{r1av} is K_r and K_{r1} averaged per atom in the fragment, respectively	TP	0.710	0.594
102	K_{avq}	function K_{rav}/K_{r1av}	TP	0.697	0.580
110	C_{r1}	first order Randić index for CR_1/CMe_2-	TP	0.649	0.557
111	C_{r1av}	C_{r1} per number of bonds in the fragment	TP	0.640	0.573
112	E_{r1av}	E_{r1} per number of bonds; E_{r1} is the 126th descriptor in this table	TP	–0.643	–0.574
113	E_d	function E_r–E_{r1}	TP	–0.714	–0.633
114	E_q	function E_r/E_{r1}	TP	–0.751	–0.668
115	E_{avd}	function E_{rav}–E_{r1av}; E_{rav} is E_r per number of bonds	TP	–0.654	–0.621
116	E_{avq}	function E_{rav}/E_{r1av}	TP	–0.657	–0.620
117	C_d	function C_r–C_{r1}; C_r is the first order Randić index for RCON	TP	–0.726	–0.645
118	C_q	function C_r/C_{r1}	TP	–0.760	–0.681
119	C_{avd}	function C_{rav}–C_{r1av}; C_{rav} is C_r per number of bonds	TP	–0.654	–0.621
120	C_{avq}	function C_{rav}/C_{r1av}	TP	–0.657	–0.620
124	E_{rr1av}	E_{rr1} per number of bonds; E_{rr1} is the first order extended Wiener index for the whole molecule	TP	–0.754	–0.650
125	$w_{NS\text{-}tr}$	function $(w_{NS}-0.400)^2$	CO/EL/HB	–0.654	–0.592
126	E_{r1}	first order Wiener index for CR_1/CMe_2	TP	0.609	0.519

[(a)]Ordinal number as in the complete list with 126 molecular descriptors. [(b)]Molecular descriptors that have absolute correlation coefficients greater than 0.600 with one or two pMICs. Molecular descriptors generated previously[14] are marked with asterisk (*). Descriptors used in the final PLS models are typed bold. [(c)]Simple or composite nature of molecular descriptors: compositional (CM), steric (ST), electronic (EL), topological (TP), hydrogen bonding (HB) and hydrophobic (HP) character. [(d)]Correlation coefficients with pMIC(HN891) and pMIC(SH5014).

TABLE II. Biological activities and molecular descriptors used in PLS models

No.	pMIC(HN891)	pMIC(SH5014)	N_{ar1}	D_{av}	V_{av}	w_{DA}	Z	L	10^6 W_{sd}
1	2.310	2.607	3	4.1	3.04	0.207	0.154	12	2621
2	2.629	2.930	3	4.8	3.30	0.241	0.462	13	2304
3	4.019	4.621	3	4.6	3.05	0.261	0.000	12	480
4	4.695	4.996	6	3.7	3.37	0.269	1.333	15	681
5	4.427	5.029	6	3.5	3.41	0.286	1.333	15	681
6	4.717	4.717	8	4.0	3.20	0.214	0.476	16	9
7	4.073	4.675	3	4.3	3.29	0.308	0.400	12	1616
8	4.112	4.714	3	4.5	3.40	0.333	0.727	12	2025
9	3.921	3.921	11	4.0	3.42	0.306	1.127	19	404
10	6.318	6.637	9	2.9	3.52	0.361	1.295	18	202
11	5.959	6.579	6	3.0	3.48	0.333	1.500	17	92
12	6.364	6.665	12	2.8	3.65	0.389	1.518	20	4
13	5.674	5.975	9	3.0	3.48	0.367	1.350	17	751
14	5.055	5.357	9	4.1	3.62	0.414	1.417	17	0
15	4.652	4.652	3	4.0	3.05	0.375	0.625	13	85
16	4.414	4.414	6	3.1	3.21	0.357	1.475	17	1369

and external validation, the models can be considered comparable to the analogous computed models[14] and therefore, are applicable in QSAR studies.

Descriptors with positive contribution to the increase in biological activity (meaning increase of drug efflux rate, *i.e.*, decrease in pMIC values) are those having negative regression coefficients. The major contributors to the activity are D_{av}, V_{av} and W_{sd}. N_{ar1}, D_{av}, Z, L and W_{sd} positively affect the activity increase. Concluding, larger hydrophobic or polar domains as well as low average electron content (characteristic for hydrophobic and even amphiphilic species) and pronounced differences in size/branching of fragments R and R_1 (this is in favor of penicillins where R_1 is very small) are the main determinants of properties of good pump substrates (bad drugs).

HCA Analysis

Descriptors used in the PLS modeling are two extensive properties (N_{ar1}, L) related to molecular/fragment size, and five intensive descriptors (D_{av}, V_{av}, w_{DA}, Z, W_{sd}) accounting for overall shape, electronic features, HB potency and lipophilic character of molecules/fragments. Absolute correlation coefficients for descriptor intercorrelations vary in the range 0.38–0.95. HCA analysis (Figure 6a) exhibits two main clusters defined by similarity indices 0.61 and 0.63. The clusters take into account positive (W_{sd}, D_{av}) and negative (N_{ar1}, L, Z, V_{av}, w_{DA}) correlations with pMICs. The descriptors are grouped in accordance with their nature: predominantly structural (W_{sd}, D_{av}), electronic (Z, V_{av}) and complex (N_{ar1}, L) descriptors, and an isolated HB descriptor (w_{DA}). However, when

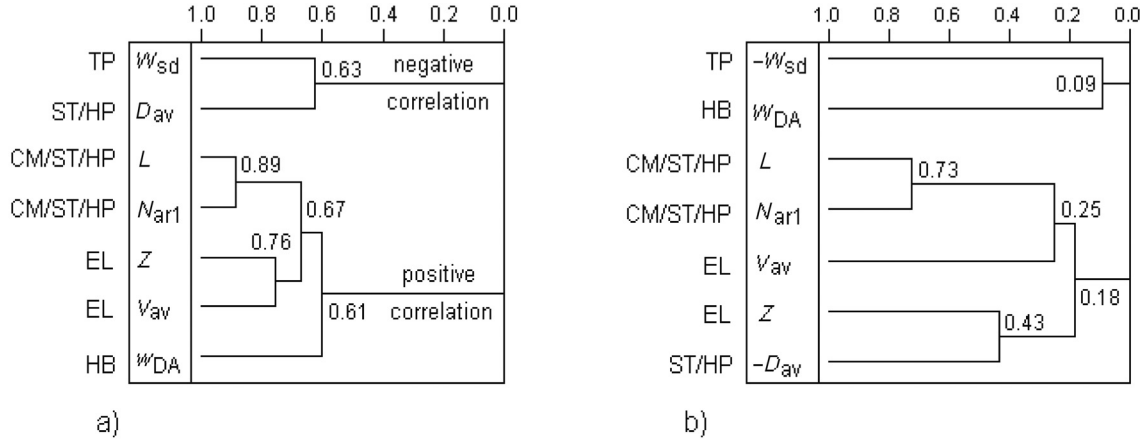


Figure 6. HCA dendograms for seven selected molecular descriptors that characterize MDR efflux activity of strain HN891. Similarity indices and the sign of descriptor-activity linear correlation coefficients are marked in the dendogram. a) Original descriptors. b) Descriptors with always positive correlation coefficients with biological activities.

TABLE III. Comparison of the HN891 and SH5014 PLS models with regression statistics

Parameters	HN891 model	SH5014 model
Regression model		
Training set / External validation set[a]	16 / 0	16 / 0
PCs (%Var)[b]	4 (85.9 %)	4 (84.4 %)
SEV[c]	0.461	0.491
SEC[c]	0.294	0.294
Q, Q^2[d]	0.913, 0.829	0.906, 0.816
R, R^2[d]	0.976, 0.952	0.977, 0.955
R.e. (10.00 %)[e]	3	1
Max. R.e.[e]	14.4 %	11.2 %
Mean R.e.[e]	4.5 %	4.5 %
Regression vector[f]	N_{ar1}: –0.152 (0.647)	N_{ar1}: –0.296 (0.558)
(Correlation coefficients)	D_{av}: –0.687 (–0.783)	D_{av}: –0.763 (–0.738)
	V_{av}: 0.527 (0.656)	V_{av}: 0.809 (0.675)
	w_{DA}: 0.121 (0.686)	w_{DA}: 0.084 (0.657)
	Z: –0.299 (0.704)	Z: –0.347 (0.664)
	L: –0.257 (0.695)	L: –0.433 (0.601)
	W_{sd}: –0.607 (–0.774)	W_{sd}: –0.625 (–0.713)
External validation[g]		
Training set / External validation set[a]	13 / 3	13 / 3
PCs (%Var)[a]	4 (85.4 %)	5 (88.5 %)
SEV_{ev}	0.607	0.673
SEC_{ev}	0.332	0.329
Q_{ev}, Q^2_{ev}	0.855, 0.703	0.834, 0.661
R_{ev}, R^2_{ev}	0.972, 0.945	0.978, 0.956
Predictions (% Errors)[h]	**7**: 3.993 (2.0)	**7**: 4.713 (–0.8)
	10: 6.003 (5.0)	**10**: 6.550 (1.3)
	14: 5.633 (–11.4)	**14**: 5.854 (–9.3)

[a]Number of β-lactams in the training and external validation sets.
[b]Number of used principal components and the corresponding % variance of the ***X*** data matrix.
[c]Standard deviations of the PLS model: SEV – standard error of leave-one-out crossvalidation, SEC – standard error of prediction (calibration).
[d]Correlation coefficients of the PLS model: Q – Linear correlation coefficient of leave-one-out crossvalidation, Q^2 – correlation coefficient of leave-one-out crossvalidation, R – linear correlation coefficient of calibration, R^2 – correlation coefficient of calibration.
[e]Relative errors: R.e. ≥ 10.00 % – number of β-lactams with relative error ≥ 10.00 %, Max. R.e. – maximum relative error, Mean R.e. – mean relative error (calculated from absolute values of relative errors).
[f]Regression vector components for descriptors in autoscaled form and descriptor-activity linear correlation coefficients (in brackets).
[g]Common parameters for the training set in external validation: SEV_{ev} – standard error of leave-one-out crossvalidation, SEC_{ev} – standard error of prediction (calibration), Q_{ev} – linear correlation coefficient of validation, Q^2_{ev} – correlation coefficient of validation, R_{ev} – linear correlation coefficient of calibration, R^2_{ev} – correlation coefficient of calibration.
[h]Predicted biological activities and relative errors (in brackets) for β-lactams from the external validation set.

$-W_{sd}$ and $-D_{av}$ are used, to eliminate the differences in signs of the correlation coefficients, the new clustering pattern is different (Figure 6b). It is based on descriptor intercorrelations and absolute activity-descriptor correlations rather than on the nature of the descriptors.

The dendogram of β-lactams (Figure 7) has two main clusters defined by similarity indices 0.56 and 0.60, which exhibit structural differences between penicillins and cephalosporins (with the exception of **6**). Sub-clusters **I**, **II**, **III** (**A** and **B**) and **IV,** defined by similarity indices 0.69–0.78, are related to the respective biological activity values. The compounds have been classified previously as good (G), moderately good (M) and poor (P) substrates of the AcrAB-TolC pump (as marked in Figure 7), based on exploratory analysis for pMICs data.[14] However, the new exploratory analysis with *a priori* descriptors shows a more complex situation. The sub-cluster **I** contains the best substrates (G and some M), while moderately good substrates (M) are divided among **II**, **IIIA** and **IV**. Poor substrates (P) are in the sub-cluster

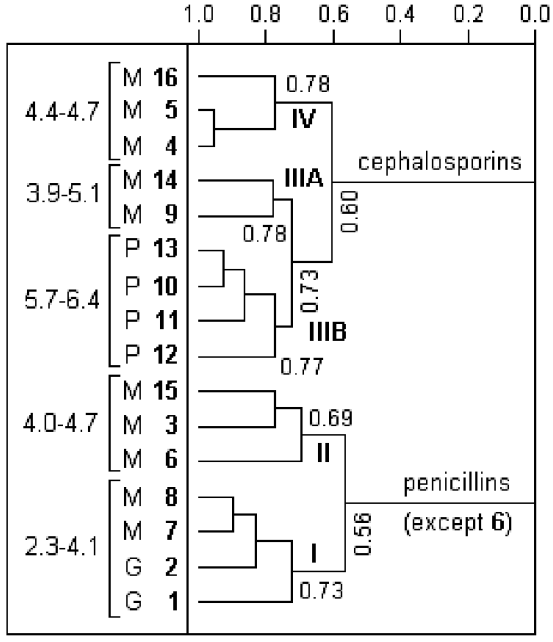


Figure 7. HCA dendogram for seven selected molecular descriptors that characterize MDR efflux activity of strain HN891. β-Lactams are labeled as good (G), moderately good (M) and poor (P) MDR substrates according to the previous chemometric study.[14] Ranges of experimental pMICs for sub-clusters and similarity indices are given also.

IIIB. From the structural point of view, β-lactams with the shortest and very short side-chains (-NH-CO-R, -C2Me$_2$- and –CH2-R$_1$) belong to **I** and **II**, respectively. This is visible from the respective low values of L and N_{ar1}, and the high values of W_{sd} (Table II). The sub-cluster **IIIB** consists of the molecules richest in heteroatoms (N, S, O as noted before[14]), described by high V_{av}, w_{DA}, Z and L, and low D_{av}. The largest molecules and those with the longest side chains belong to this sub-cluster. Molecules in **IIIA** have similar characteristics, with the exception of greater values of D_{av}. β-Lactams in **IV** are of moderate size and are characterized by intermediate values of most descriptors (N_{ar1}, D_{av}, L, W_{sd}, w_{DA}, V_{av}). In general, the liphophilic or amphiphilic character of β-lactams is related to better efflux (lower pMIC), as has been reported.[19] This is also visible in the clustering pattern of the HCA dendogram.

PCA Analysis

Table IV contains cumulative variances for PC1-PC7 from PCA analyses using data sets A (Table II), B and C (Table I, descriptors with absolute correlation coefficients > 0.600 with pMIC(HN891) and pMIC(SH5014), respectively).

The loadings plots in Figure 8 take into account PC1-PC2 from PCA applied to the data set A (Table II) that was used for the PLS models (Table III) and HCA (Figure 6). PC1 separates descriptors in the same way as in HCA (Figure 6a), *i.e.*, depending on the sign of their correlation coefficients with pMICs. The clustering in the PC1-PC2 plot (Figure 8a) according to the nature of descriptors (structural, electronic, HB and complex) follows practically the same pattern as in HCA. PC2 discriminates descriptors that have a steric and/or hydrophobic nature (D_{av}, L and Na_{r1} at negative PC2) from electronic (V_{av} and Z where PC2 ranges from 0.2 to 0.4) and topological/HB descriptors (W_{sd} and w_{DA} at highly positive PC2).

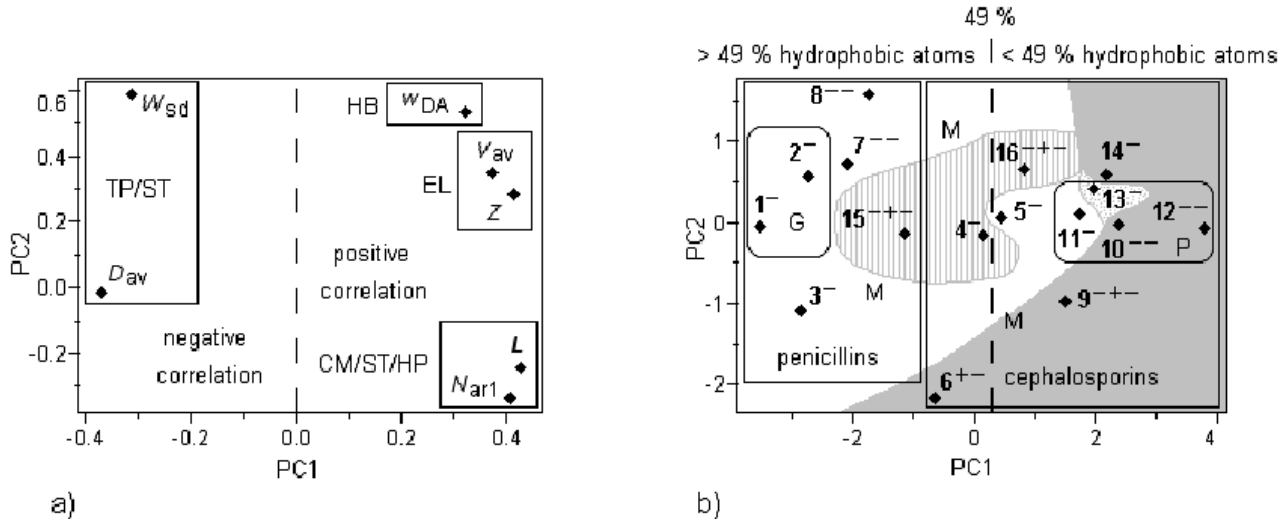


Figure 8. PCA plots with the first two principal components: a) loadings plot and b) scores plot. PCA was based on seven selected molecular descriptors (data set A) that characterize MDR efflux activity of strain HN891. β-Lactams are marked in different ways to distinguish charged species (anion: –, dianion – –, zwitterion + –, anion-zwiterion – + –), β-lactams with different contents of hydrophobic atoms (vertical dashed line at 49 % content), biological activity classes (G, M, P) and species with different ring-containing side chains (gray: R and R$_1$; gray dots: R$_1$; gray vertical lines: no rings; white: R).

TABLE IV. Basic PCA statistics[a] related to the dimensionality[b] of pMIC-related data sets

	Data sets		
	A: 7 descriptors (HN891/SH5014)[c]	B: 66 descriptors (HN891)[d]	C: 44 descriptors (SH5014)[e]
PC	% CVar.	% CVar.	% CVar.
1	**68.36**	**69.25**	**67.67**
2	**78.94**	**79.74**	**78.66**
3	*87.81*	*85.66*	*85.48*
4	*95.07*	*90.35*	*90.38*
5	98.05	93.53	93.46
6	99.68	95.66	95.62
7	100.00	96.80	97.03

[a]Cumulative variance (% CVar.) described by each principal component (PC). [b]Principal components that probably determine the maximum dimensionality of the data are typed italics and those with significant fraction of the total variance are typed bold. [c]Data set A: seven descriptors used in PLS modeling for both strains (Table II). [d]Data set B: descriptors with absolute correlation coefficients > 0.600 with pMIC(HN891). [e]Data set C: descriptors with absolute correlation coefficients > 0.600 with pMIC(SH5014).

The PC1-PC2 scores plot is in Figure 8b. The clustering pattern in this plot is very similar to that in the HCA dendogram in Figure 7, with the exception of groups **IIIA** and **IIIB**, which cannot be identified in the scores plot. PC1 discriminates penicillins from cephalosporins, which is the major determinant of clusters with good (G), moderately good (M) and poor (P) substrates of AcrAB--TolC pump. The limit of 49 % for values of a lipophilic descriptors (hydrophobic surface area fraction S_f), arbitrarily positioned in the plot, discriminates more hydrophobic (at negative PC1) from more polar (at positive PC1) β-lactams. The number and type of charged groups distinguish anions (–), dianions (– –), zwitterion (+ –), and anion-zwitterions (– + –) that tend to cluster. This means that the two properties, hydrophobicity and charge, affect the behavior of β-lactams in interaction with strains of *S. typhimurium*, as discussed previously.[14] β--Lactams in the PC1-PC2 plot are distinguished according to the number and position of rings in fragments R and R_1: species without rings (**4**, **15**, **16** in the central part of the plot), β-lactams with rings only in R (**1**–**3**, **5**, **7**, **8**, **11** in the left and central part), an anion with ring only in R_1 (**13** in the central part), and species with rings in R and R_1 (**6**, **9**, **10**, **12**, **14** in the right part and bottom of the plot). Consequently, the number and position of ring structures in the side chains also affect β-lactam behavior. Descriptors with significant or exclusive structural/hydrophobic character in the PC1-PC2 loadings plot (Figure 8a) determine the behavior of β-lactams in the same part of the PC1-PC2 scores plot (Figure 8b): all penicillins and cephalosporin **6** (a well-defined cluster in Figure 7). Descriptors with predominant electronic and HB nature are the major determinants of the behavior of most cephalosporins, especially of poor substrates (**10**–**13**) and **14** that are the richest β-lactams in nitrogen and sulfur atoms (potential HB donor/acceptor groups). In general, lower values of PC1 are related to more active β-lactams with large R and small R_1 substituents in the side chains (large W_{sd}, small N_{ar1} and L values), molecules that are more hydrophobic and possess well-defined large hydrophobic and polar domains (large D_{av} values). Higher PC1 is related to larger and more polar molecules rich in heteroatoms and HB groups. Hence, PC1 can be considered as a general PC built by all descriptors related to pMIC(HN891) and pMIC(SH5014) with correlation coefficients 0.847 and 0.787, respectively. PC1 describes overall molecular amphiphilicity, which decreases with hydrophobic and increases with polar character of β-lactams. Consequently, the increase of bacterial MDR efflux power is related to the decrease in PC1.

PC1 contains about 2/3 and PC2 significantly smaller percentage (only 11 %) of the total variance. Similar trends are observed in other related PCA analyses (see Table IV), which suggests that the seven selected descriptors (data set A) are good representations of the data sets B and C. Descriptors D_{av} and w_{DA} make the highest contribution to PC2 (Figure 8 top), and thus are best correlated with this PC (correlation coefficient is 0.46 and 0.50, respectively). The scores plot (Figure 8b) shows that PC2 discriminates β-lactams in a rather complex way. The largest molecules as well as smaller ones with pronounced hydrophobic (**1**), amphiphilic (**4**, **5**, **15**) and polar character (**10**–**12**) are placed in the middle of the plot, *i.e.* around PC2 = 0. In general, molecular size and hydrophobicity increase and polar character and HB potency decrease with PC2, placing **3**, **6** and **9** in the bottom part of the plot. The opposite is observed for increasing PC2, due to which **2**, **7**, **8**, **13**, **14** and **16** are positioned in the top part of the plot. Molecular topology also varies along PC2. More branched β-lactams are at high positive PC2, and molecules with more compact R and R_1 are placed at negative PC2. Since both W_{sd} and w_{DA} are positively correlated with PC2, the chemical meaning of this PC is related to the topological distribu-

tion of HB groups, *i.e.* HB potency depends on the size and shape of R and R_1.

PC3 contains 9 % of the original information in this PCA analysis, similar to the other related PCA analyses (Table IV). It seems that this PC maintains some clustering patterns of β-lactams with respect to 5- and 6-membered rings in the side chains R and R1 (scores plots not shown). PC4 contains 7 % total variance, whilst in the other related PCA analyses this contribution is 4–5 % (Table IV). This PC is responsible for certain clustering patterns of β-lactams with respect to molecular topology and hydrophobic character (scores plots not shown).

New Type of Molecular Descriptors: Amphiphilic Descriptors?

PLS models (Table III) and PCA analyses (Table IV) use four principal components. PC3 and PC4 in the regression models and PCA for the data set A have mainly originated from descriptors D_{av} and V_{av}. These descriptors, considered as structural/hydrophobic and electronic, respectively (Figure 6), seem to describe overall molecular amphiphilicity. This is important to note for β-lactams, which are, due to their basic molecular structure, essentially amphiphilic. This fact can be an important determinant for β-lactam interactions with proteins and membranes.

V_{av} is low for more hydrophobic compounds (benzene: 2.50), moderate for amphiphilic (phenol: 2.77), and high for more polar substances (trinitrophenol: 4.42). The V_{av} increase is also related to high contents of unsaturated bonds and low contents of hydrogen atoms, so it is best to compare compounds within the same class. V_{av} values for the β-lactams in this work vary over a relatively narrow range, 3.04–3.65 or 20 %. PLS equations (Table III) clearly demonstrate the decrease in biological activity with the increase in V_{av}.

D_{av} includes hydrophobic, polar, topological and size/shape characteristics of the β-lactam side chains. Figure 4 illustrates extreme values of this descriptor, within the maximum for **2** (4.8) and the minimum for **12** (2.8). **2** is a smaller species with 6 domains along the R-R_1 path of 13 atoms, and **12** is a larger species with 13 domains and 20 atoms along the path. **2** has larger and better defined domains, whilst **12** has pronounced mosaic character. Hence, the polar and hydrophobic groups form almost two continuous domains that facilitate the molecular recognition between β-lactams and AcrB receptors. The average size of domains in **2** (4.8) corresponds to 4- or 5-membered rings, while in **12** this number is rather small (2.8) to form structural units that would interact with amino-acid residues of AcrAB-TolC. D_{av} is elevated for molecules with large domains or with highly pronounced hydrophobic or polar character. The other extreme case is D_{av} = 1 for molecules with many monoatomic domains (the highest possible mosaicity). Amphiphilic molecules thus may have intermediate values of D_{av}, like the β-lactams in this work (2.8–4.8) and some amino-acids (leucine: 2.50; phenyalanine: 4.00; tryptophane: 4.00). The existence of well-defined lipophilic and hydrophilic domains in β-lactams and other antibiotics, as amphiphilic compounds, has been pointed out by van Bambeke *et al*.[24]

To test the amphiphilic character of V_{av} and D_{av}, the respective linear correlation coefficients with 12 lipophilic descriptors from the earlier work[14] and 7 electronic/HB descriptors (5 in PLS modeling and third-order polarizability, also previously calculated,[14] and w_{DA}), were evaluated. Absolute coefficients for V_{av} vary over a large range with lipophilic (0.107–0.607) and electronic/HB descriptors (0.154–0.800). A similar trend is observed for D_{av} with respect to lipophilic (0.099–0.727) and electronic/HB descriptors (0.045–0.782). This indicates the amphiphilic character of V_{av} and D_{av}.

Hansch *et al*. have shown recently[52] that the number of valence electrons, which is directly related to molecular polarizability, is an important molecular descriptor in QSAR. This supports the use of the intensive descriptor V_{av} for QSAR. Besides classical descriptor types, amphiphilic descriptors could be considered as a new class of descriptors. β-Lactam amphiphilicity seems to be important in processes like drug-membrane and drug-AcrB interactions (Figure 2), drug-intestinal transporter[44,45] and general drug-protein interactions (mainly of electrostatic and hydrophobic natures).[43,53] Membranes and proteins involved in β-lactam efflux in Gram-negative bacteria (Figure 2) are generally hydrophilic microinterfaces. The central cavity of AcrB possesses receptor sites able to accommodate a large variety of structurally diverse amphiphilic drugs. β-Lactams establish diverse intermolecular interactions with the receptors, as has been seen during the modeling of AcrB complexes with **1**–**16**[40] and from crystal the structure of AcrB-**1**.[34]

Intramolecular Interactions between β-Lactam Side Chains

Several descriptors in Table II encode information on interactions between the side chains R and R_1: w_{DA} (HB donor/acceptor balance between R and R_1), L and Z (chemical balance between R and R_1), and N_{ar1}, D_{av} and V_{av} (size, shape and electronic features). These descriptors indicate that longer and more flexible side chains with elevated numbers of polar/HB or hydrophobic groups can adopt conformations which enhance intramolecular interactions. The descriptor W_{sd} positively contributes to the activity increase. W and W_1 are included in its definition (Table I) as $W_{sd} = (W_n - W_{n1})^2$, where W_n and W_{n1} are W and W_1 normalized by the cube of the number of corresponding non-H atoms, respectively. β-Lactams with pronounced size/shape differences between the side chains, mainly penicillins, have elevated W_{sd} (large rings: **1**, **2**,

7, **8**; long chain: **16**). β-Lactams with minimum W_{sd} are cephalosporins with similar rings in R and R_1 (**6**, **12**, **14**).

The β-lactam ring and peptide bond form two distinctive regions: a large hydrophobic domain and three polar domains that make a continuous 3D polar/HB region. Cronin *et al.*[47] have noted that intramolecular HB descriptors are important for biological activity of a variety of antibacterials. For example, intramolecular hydrogen bonds participate in stabilization of erythromycin and rifampicin 3D structure.[38] Computed 3D structures of the 16 β-lactams[14] have shown various interactions between hydrophobic, polar, charged and HB groups in the side chains. These interactions stabilize compact, U-shaped or bent conformers. Since linear cylinder-like shapes are preferred for drug interactions with the pump receptors and channels,[38] compact and globular conformers of β-lactams are bad pump substrates. Complex substituents in β-lactams may contain small hydrophobic/polar domains, which makes these molecules unsuitable for interactions with all AcrAB-TolC receptors. Predominantly polar molecules do not accumulate at the inner membrane (Figure 1) and thus, are harder to be excreted.[18] The presence of large R_1 and polar R, and the pronounced overall polar and mosaic character of a β--lactam do not favor drug-receptor interactions. These intermolecular interactions, unlike in most QSAR studies where strong interactions result in stable drug-receptor complex, may be considered sufficiently strong to attract a drug molecule but weak enough to enable further drug movement from one receptor to another along the efflux pathway. Bad β-lactam substrates can be strongly bound to the AcrB receptors, which then disables efficient efflux of such drugs.

The interaction between the two side chains can be direct (contact) or indirect (mediated by the central ring), affecting the efflux rates pMIC. The application of chemometrics in this work has indicated the possibility of such interactions. This may favor the use of *a priori* descriptors to describe complex phenomena such as MDR. Olah *et al*.[54] have recently shown the usefulness of 1D and 2D descriptors as biologically relevant in more than 1600 QSAR-PLS models, which just may confirm that *a priori* descriptors, due to their nature, can be useful in QSAR studies.

CONCLUSION

The chemometric methodologies applied in this work lead to the following conclusions:

1) Advantages of the *a priori* approach are chemical interpretation and understanding of descriptors and drugs (β-lactams), with easy generation of the descriptors and acceptable PLS statistics.
2) Classification of β-lactams as good, moderately good and poor substrates of the AcrAB-TolC pump is confirmed, although the clustering pattern seems to be more complex.
3) Some descriptors are shown to have an amphiphilic nature, which is rare in QSAR studies.
4) Intramolecular interactions between the side chains affect physico-chemical (hydrophobicity, amphiphilicity and hydrogen bonding potency) and biological properties of β-lactams.
5) The *a priori* approach in this work represents an interesting example of how much a relatively complex phenomena such as bacterial multidrug resistance can be treated in a rather simply way.

Acknowledgements. – This work was supported by the State of São Paulo Funding Agency (FAPESP). The authors acknowledge Dr. Carol H. Collins for English revision.

REFERENCES

1. M. M. C. Ferreira, *J. Braz. Chem. Soc*. **13** (2002) 742–753.
2. J. C. Pinheiro, R. Kiralj, M. M. C. Ferreira, and O. A. S. Romero, *QSAR Comb. Sci*. **22** (2003) 830–842.
3. R. Kiralj, Y. Takahata, and M. M. C. Ferreira, *QSAR Comb. Sci*. **22** (2003) 430–448.
4. M. M. C. Ferreira, *Chemosphere* **44** (2001) 125–146.
5. L. R. Cirino and M. M. C. Ferreira, *Quim. Nova* **26** (2003) 312–318.
6. J. S. Murray, P. Politzer, and G. R. Famini, *J. Mol. Struct.-Theochem* **454** (1998) 299–306.
7. G. R. Famini and L. Y. Wilson, *J. Phys. Org. Chem*. **12** (1999) 645–653.
8. R. Kiralj and M. M. C. Ferreira, *J. Mol. Graphics Modell.* **21** (2003) 435–448.
9. B. Testa and A. I. Bojarski, *Eur. J. Pharm. Sci*. **11(Suppl 2)** (2000) S3–S4.
10. G. Müller, *QSAR Comb. Sci*. **21** (2002) 391–396.
11. R. Kiralj and M. M. C. Ferreira, *J. Mol. Graphics Modell*. **21** (2003) 499–515.
12. C. Hansch and T. Fujita, *J. Am. Chem. Soc*. **86** (1964) 1616–1626.
13. S. M. Free and J. W. Wilson, *J. Med. Chem*. **7** (1964) 395–399.
14. M. M. C. Ferreira and R. Kiralj, *J. Chemom*. **18** (2004) 242–252.
15. R. Kiralj and M. M. C. Ferreira, *J. Braz. Chem. Soc*. **14** (2003) 20–26.
16. R. Kiralj and M. M. C. Ferreira, *J. Chem. Inf. Comput. Sci*. **42** (2002) 508–523.
17. R. Kiralj and M. M. C. Ferreira, *J. Chem. Inf. Comput. Sci*. **43** (2003) 787–809.
18. H. Nikaido, M. Basina, V. Y. Nguyen, and E. Y. Rosenberg, *J. Bacteriol*. **180** (1998) 4686–4692.
19. E. B. Tikhonova, Q. Wang, and H. I. Zgurskaya, *J. Bacteriol*. **184** (2002) 6499–6507.
20. E. Giraud, A. Cloeckaert, D. Kerboeuf, and E. Chaslus-Dancla, *Antimicrob. Agents Chemother.* **44** (2000) 1223–1228.
21. C. Anderesen, *Rev. Physiol. Biochem. Pharmacol*. **147** (2003) 122–165.

22. D. L. Jack, N. M. Yang, and H. M. Saier Jr., *Eur. J. Biochem*. **268** (2001) 3620–3639.
23. M. H. Saier Jr., *Microbiol. Mol. Biol. Rev.* **64** (2000) 354–411.
24. F. Van Bambeke, J. M. Michot, and P. M. Tulkens, *J. Antimicrob. Chemother.* **51** (2003) 1067–1077.
25. F. Van Bambeke, Y. Glupczynski, P. Plésiat, J. C. Pechère, and P. M. Tulkens, *J. Antimicrob. Chemother.* **51** (2003) 1055–1065.
26. S. B. Levy, *J. Antimicrob. Chemother.* **49** (2002) 25–30.
27. R. Wise, *J. Antimicrob. Chemother.* **51**(Suppl 1) (2003) 37–42.
28. H. Akama, T. Matsuua, S. Kashiwagi, H. Yoneyama, S. I. Narita, T. Tsukihara, A. Nakagawa, and T. Nakae, *J. Biol. Chem*. **279** (2004) 25939–25942.
29. M. K. Higgins, E. Bokma, E. Koronakis, C. Hughes, and V. Koronakis, *Proc. Natl. Acad. Sci. U. S. A.* **27** (2004) 9994–9999.
30. S. Murakami, N. Tamura, A. Saito, T. Hirata, and A. Yamaguchi, *J. Biol. Chem*. **279** (2004) 3743–3748.
31. I. T. Paulsen, *Curr. Opin. Microbiol.* **6** (2003) 446–451.
32. S. Murakami, R. Nakashima, E. Yamashita, and A. Yamaguchi, *Nature* **419** (2002) 587–593.
33. E. W. Yu, G. McDermott, H. I. Zgurskaya, H. Nikaido, and D. E. Koshland Jr., *Science* **300** (2003) 976–980.
34. E. W. Yu, J. R. Aires, G. McDermott, and H. Nikaido, *J. Bacteriol*. **187** (2005) 6804–6815.
35. S. Murakami, R. Nakashima, E. Yamashita, T. Matsumoto, and A. Yamaguchi, *Nature* **443** (2006) 173–179.
36. M. A. Seeger, A. Schiefner, T. Eicher, F. Verrey, K. Diederichs, and K. M. Pos, *Science* **313** (2006) 1295–1298.
37. J. Mikolosko, K. Bobyk, H. I. Zgurskaya, and P. Ghosh, *Structure* **14** (2006) 577–587.
38. V. Koronakis, A. Sharff, E. Koronakis, and B. Luisi, *Nature* **405** (2000) 914–919.
39. M. K. Higgins, J. Eswaran, P. Edwards, G. F. X. Schertler, C. Hughes, and V. Koronakis, *J. Biol. Chem.* **342** (2004) 697–702.
40. R. Kiralj and M. M. C. Ferreira, *J. Mol. Graphics. Modell.* **25** (2006) 126–145.
41. B. W. Bycroft (Ed.), *Dictionary of Antibiotics and Related Substances*, Chapman-Hall, London, 1988.
42. P. Taboada, Y. Fernández, and V. Mosquera, *Biomacromolecules* **5** (2004) 2201–2211.
43. F. van Bambeke, E. Balzi, and P. M. Tulkens, *Biochem. Pharmacol*. **60** (2000) 457–470.
44. V. H. Lee, *Eur. J. Pharm. Sci.* **11**(Suppl2) (2000) S41–S50.
45. S. Gebauer, I. Knütter, B. Hartrodt, M. Brandsch, K. Neubert, and I. Thondorf, *J. Med. Chem*. **46** (2003) 5725–5734.
46. D. T. Stanton, P. J. Madhav, L. J. Wilson, T. W. Morris, P. M. Hershberger, and C. N. Parker, *J. Chem. Inf. Comput. Sci.* **44** (2004) 221–229.
47. M. T. D. Cronin, A. O. Aptula, J. C. Dearden, J. C. Duffy, T. I. Netzeva, H. Patel, R. H. Rowe, T. W. Schultz, A. P. Worth, K. Voutzoulidis, and G. Schüürmann, *J. Chem. Inf. Comput. Sci.* **42** (2002) 869–878.
48. D. J. Gordon, J. J. Balbach, R. Tycko, and S. C. Meredith, *Biophys. J*. **86** (2004) 428–434.
49. P. Crivori, G. Cruciani, P. A. Carrupt, and B. Testa, *J. Med. Chem*. **43** (2000) 2204–2216.
50. *Matlab 6.1*, MathWorks, Inc., Natick, MA, 2001.
51. *Pirouette 3.02*, Infometrix, Inc., Woodinville, WA, 2001
52. C. Hansch, W. E. Steinmetz, A. J. Leo, S. B. Mekapati, A. Kurup, and D. Hoekman, *J. Chem. Inf. Comput. Sci.* **43** (2003) 120–125.
53. J. M. Ruso, P. Taboada, L. M. Varela, D. Attwood, and V. Mosquera, *Biophys. Chem*. **92** (2001) 141–153.
54. M. Olah, C. Bologa, and T. I. Oprea, *J. Comput.-Aided Mol. Des*. **18** (2004) 437–449.

SAŽETAK

A priori deskriptori u QSAR: slučaj višestruke otpornosti Gram-negativnih bakterija prema β-laktamima

Rudolf Kiralj i Márcia M. C. Ferreira

Izračunato je više od stotine globalnih i lokalnih molekulnih opisivača *a priori* za sterička, topološka, elektronska i hidrofobna svojstva, kemijski sastav i svojstva vodikovih veza 16 β-laktama. Dva su regresijska modela izgrađena metodom parcijalnih najmanjih kvadrata i ispitana unakrsnom provjerom. *A priori* modeli (Q^2 > 0,80, R^2 > 0,95, SEV < 0,50) su usporedivi s prethodno dobivenim računskim modelima. Raspravljano je i o unutarmolekulnim djelovanjima u β-laktamima i međumolekulnim djelovanjima β-laktam-receptor u smislu molekulskih opisivača.