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

Rudolf Kiralj and Márcia M. C. Ferreira. *marcia@iqm.unicamp.br, rudolf@iqm.unicamp.br, http://lqta.iqm.unicamp.br*
Laboratório de Quimiometria Teórica e Aplicada (LQTA), Instituto de Química, Universidade Estadual de Campinas, Campinas SP, 13084-971, Brazil

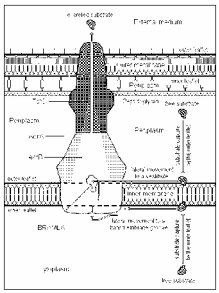
THE PRIMARY OBJECTIVES OF THIS WORK

1) To construct PLS (partial least squares) regression models for 16 β-lactam antibiotics as substrates of the pump AcrB (a component of AcrAB-TolC membrane transporter) that exists in several Gram-negative bacteria), where activities are efflux rates of the compounds excreted from two strains of *S. typhimurium* (HN891 and SH5014);

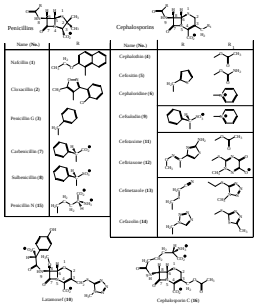
2) To use only *a priori* (not computed) molecular descriptors capable to produce PLS models comparable with those obtained previously from computed molecular descriptors;

3) To explain chemical information observed in chemometric and QSAR analyses.

INTERACTING MOLECULES: AcrAB-TolC PUMP AND β-LACTAMS



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



Chemical structures of β-lactams at neutral pH, with atomic numbering for penicillines and cephalosporins

SELECTED *A PRIORI* MOLECULAR DESCRIPTORS

Molecular descriptors used in exploratory analysis and PLS modeling

No.[a]	Symbol[b]	Definition	Nature[c]	HN891[d]	SH5014[d]
6	N_{nH}	number of non-H atoms in C-R_1 fragment	CM/ST/HP	0.647	0.558
35	Dav	average size of domains counted for D, NnH/D	ST/HP	-0.783	-0.738
39	V_{av}	number of valence electrons per atom	EL	0.656	0.676
45	w_{DA}*	number fraction of HB donors and acceptors	HB	0.686	0.657
46	Z*	function v(R)=v(R1)-S; v(R), v(R1) - average No. valence electrons in R and R1, respectively	EL	0.703	0.664
58	L	number of non-H atoms in the shortest path from the R_1 end to the R end	CM/ST/HP	0.695	0.601
83	W_{sd}	function $(W_w \cdot W_{x1})^2$; $W_w = W/(N_a)^2$, $W_{x1} = W_1/(N_{x1})^2$; W, W_1 - first order Wiener index for R-CO-N-C and C- C-R_2/CMe_2 fragment	TP	-0.774	-0.713

[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[12] are marked with *. 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).

Seven molecular descriptors. Two of them were calculated in previous work (M. M. C. Ferreira, R. Kiralj, J. Chemometr., 18, 2004, 242-252). HN891 and SH5014 are the two strains of *S. typhimurium*.

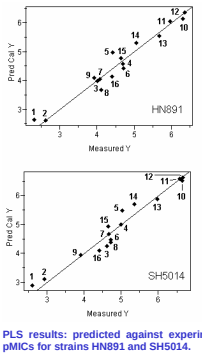
A PRIORI QSAR: PLS MODELING

PLS regression models[a] for pMICs

Model	Molecular descriptors[b]	SEV[c]	Q^2	R^2	PCs (%)[d]
HN891					
this work	N_{nH1}, D_{av}, V_{av}, w_{DA}, Z, L, W_{sd}	0.461	0.913	0.952	4 (86%)
pMIC = − 0.152 N_{nH1} − 0.687 D_{av} + 0.527 V_{av} + 0.121 w_{DA} − 0.299 Z − 0.257 L − 0.607 W_{sd}					
previous work	$GlogK_{OH}$, $SlogP_o$, $logP_{1a}$, w_{tet}, N_{Het}, D_p, N_{H2}	0.209	0.982	0.993	3 (85%)
pMIC = 0.469 $GlogK_{OH}$ − 0.354 $SlogP_o$ + 0.460 $logP_{1a}$ + 0.001 w_{tet} − 0.122 N_{Het} − 0.275 D_p + 0.280 N_{H2}					
SH5014					
this work	N_{nH1}, D_{av}, V_{av}, w_{DA}, Z, L, W_{sd}	0.491	0.906	0.953	4 (84%)
pMIC = − 0.296 N_{nH1} − 0.763 D_{av} + 0.809 V_{av} + 0.084 w_{DA} − 0.347 Z − 0.433 L − 0.625 W_{sd}					
previous work	$GlogK_{OH}$, $SlogP_o$, $logP_{1a}$, w_{tet}, N_{Het}, D_p, N_{H2}	0.316	0.962	0.982	3 (85%)
pMIC = 0.557 $GlogK_{OH}$ − 0.393 $SlogP_o$ − 0.339 $logP_{1a}$ + 0.003 w_{tet} − 0.102 N_{Het} − 0.260 D_p + 0.195 N_{H2}					

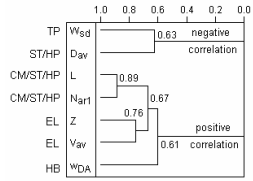
[a]PLS equations are given with all variables in autoscaled form. [b]Molecular descriptors according to the names and definitions given in previous table, except of computed descriptors from previous study: $logP_{1a}$ – a lipophilicity parameter, $GlogK_{OH}$ – Gaussian form of a lipophilicity parameter $logK_{OH}$, $SlogP_o$ – square of a lipophilicity parameter $logP_o$, D_p – Y-component of the molecular dipole moment from PM3 semi-empirical calculations. [c]Standard error of validation. [d]Correlation coefficient from validation. [e]Correlation coefficient from prediction. [f]Number of principal components used and % of the total variance.

PLS modeling of pMICs (MIC = minimal inhibitor concentration) and comparison with the results from the previous work.

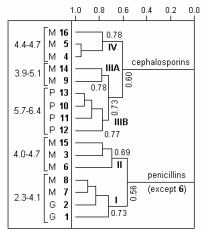


PLS results: predicted against experimental pMICs for strains HN891 and SH5014.

HIERARCHICAL CLUSTER ANALYSIS

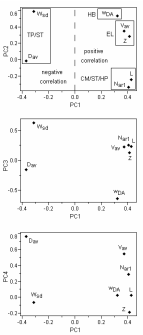


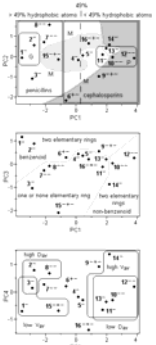
HCA dendogram with variables for seven selected molecular descriptors that characterize MDR efflux activity of strain HN891. Similarity indices and the sign of descriptor-activity correlation coefficients are marked in the dendogram. Clustering of the descriptors with respect to their nature can be observed also.



HCA dendogram with samples for seven selected molecular descriptors that characterize MDR efflux activity of strain HN891. The samples labeled as good (G), moderately good (M) and poor (P) MDR substrates according to the previous chemometric study. The ranges of experimental pMICs for sub-clusters and similarity indices are given also.

PRINCIPAL COMPONENT ANALYSIS

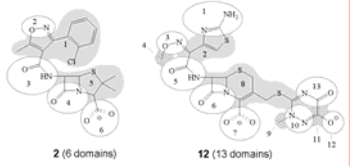




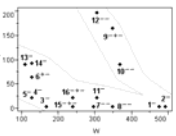
PCA loadings plots with the first four principal component, based on the seven selected molecular descriptors.

PCA scores plots with the first four principal components. PCA was based on the seven selected molecular descriptors that characterize MDR efflux activity of strain HN891. Samples in PC1-PC2 plot are marked with different ways to distinguish charged species (anion: –, dianion – –, zwitterion + –, anion-zwitterion – + –), samples with different content of hydrophobic atoms (vertical dashed line at 49% content), biological activity classes (G, M, P) and samples with different ring-containing side chains (gray: R and R_1; gray dots: R_1; gray vertical lines: no rings; white: R). Various clustering patterns with respect to molecular features may be observed. PCA enables chemical identification of the PCs: 1) PC1 - the general PC -> relationships between molecular hydrophobicity and hydrogen bonding potency, quantitatively related to pMICs; 2) PC2 - hydrogen bonding groups distribution in the side chains; 3) PC3 - hydrophobic groups distribution in the side chains; 4) PC4 - molecular amphiphilicity.

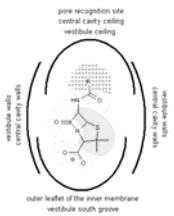
RESULTING CHEMISTRY



Chemical interpretation of an amphiphilicity descriptor D_{av}. Hydrophobic (gray tones) and polar (white ellipses) domains in a good (2) and poor (12) substrates of AcrAB-TolC bacterial pump.



Chemical interpretation of descriptor W_{sd} that indicates interaction between side chains R and R_1. Correlation between the first-order Wiener indices W (for R-CO-N-C) and W_1 (for C- C-R_1/CMe_2) included in the definition of W_{sd}, is visible. Two groups of samples can be noticed.



Schematic representation of stereoelectronic penicillin–pump receptor interactions. The drug consists of large hydrophobic domain (gray) and a few polar domains that form almost continuous polar part of the molecule (gray dots). The substituent R in the side chain (starred), depending on its hydrophobic/polar character, may alter molecular properties and interactions with receptors. Bacterial receptors of the inner membrane or AcrB protein prefer to bind to specific parts of the drug molecule.