Chemometric and QSAR prediction of the multidrug resistance of VmrA efflux pump from *Vibrio parahaemolyticus*

Rudolf Kiralj and Márcia Miguel Castro 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 OBJECTIVES OF THIS WORK

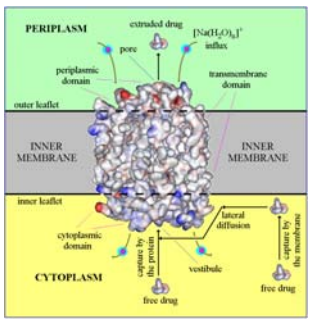
1) ***V. parahaemolyticus***, like other non-cholera *Vibrio* species, contaminates most marine animals in coastal waters, and causes frequently food poisoning (associated gastroenteritis), wound and soft tissue infections, septicemia, and other infections. Among several thousand of infected persons worlwide per year, there are over 10% cases with severe diseases that may end in death when immunocompromized persons are infected. To get more insight into the multidrug resistance (MDR) mechanism of this microbe, particularly VmrA efflux pump and its function, is one of the objectives;

2) To perform QSAR (Quantitative Structure-Activity Relationship) & chemometric study of structurally unrelated substrates of the VmrA, as extruded by ***E. coli*** strains: KAM32 and KAM32/pVCJ6 (with VmrA);

3) To develop QSAR and SAR means to predict to which drugs and xenobiotics ***V. parahaemolyticus*** will be sensitive or resistant due to its efflux ability by means of the VmrA pump.

THE VmrA PUMP, ITS MULTIDRUG RESISTANCE (MDR) EFFLUX MECHANISM, AND ITS SUBSTRATES

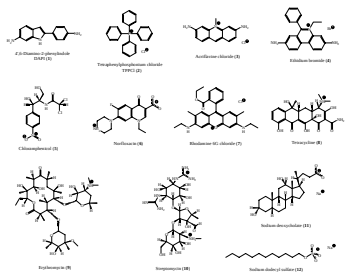
1



Structure of the Na+/multidrug transporter VmrA and drug efflux a Gram-negative cell of *V. parahaemolyticus* AQ334 [M. M. C. Ferreira, R. Kiralj, unpublished], as modeled from primary structure [J. Chen *et al., J. Bacteriol.*, 184 (2002) 572-576.]. Influx of Na+ in salty medium causes complex formation of Na+ with Asp and Glu in transmembrane, cytoplasmic, and periplasmic domains of VrmA. This provokes allosteric changes in the VmrA, and opening the periplasmic window, pore. Drug molecules may be are captured by the inner leaflet of the inner membrane and by lateral diffusion are brought to the vestibule. They may be captured directly by the cytoplasmic domain and brought to the vestibule through which they enter into the central cavity where they accumulate before being extruded from the pump. VmrA is a secondary active transporter that uses electrochemical potential of Na+ across the membrane as its energy source. It is a defense mechanism of *V. parahaemolyticus* against several structurally unrelated drugs and xenobiotics.

Morphology of *Vibrio parahaemolyticus* cells according to electron microscopy experiments (obtained from WWW resources).

2



The only VmrA substrates with known efflux activities [J. Chen *et al., J. Bacteriol.*, 184 (2002) 572-576.] (Table 1). The active organic parts of these twelve substrates (neutral pH) belong to twelve distinct classes of organic compounds.

3

Table 1. Experimental data for drugs 1 - 12

No.	CSD source[a]	Formula[b]	pMIC(pVCJ6)[c]	pMIC(KAM)[d]	pMICΔ[e]
1	ASUMEG	$C_{24}H_{29}N_2$	3.242	5.961	2.719
2	DURDAV01	$[C_{24}H_{20}P]^+$	3.468	4.672	1.204
3[f]	-	$[C_{14}H_{14}N_2]^+$	3.909	5.114	1.205
4	ETHDIB	$[C_{21}H_{20}N_2]^+$	4.392	4.994	0.602
5	CLMPCL02	$[C_{17}H_{17}ClN_2O_3S]^+$	5.810	5.810	0
6	XAYGEJ	$C_{18}H_{17}FN_2O_3$	7.027	7.027	0
7	QIMMEE	$[C_{20}H_{21}N_2O_2]^+$	4.777	4.777	0
8	TETCYH10	$[C_{22}H_{25}N_2O_4]^+$	5.949	5.949	0
9	NAVTEJ	$[C_{27}H_{34}O_{11}N]^+$	5.264	5.264	0
10	STOSEH10	$[C_{21}H_{41}O_2N_2]^+$	5.464	5.464	0
11	GOLWIN	$[C_{24}H_{52}O_4]^-$	5.318	5.318	0
12	SATLUU	$[C_{12}H_{25}O_4S]^-$	6.461	6.461	0

[a]CSD codes for the structures retrieved from the CSD database. Complete or partial structures were used in molecular modeling. [b]Formula for the drug's organic component in neutral, cationic/protonated (+), anionic (–) or zwitterionic (±) state as applied in molecular modeling. [c]Efflux activity pMIC of *E. coli* strain KAM32/pVCJ6. [d]Efflux activity pMIC of *E. coli* strain KAM32. [e]Difference between the two efflux activities. [f]The simplest drug that was not found in the CSD, and was modeled by Titan program.

Appropriate structures from the Cambridge Structural Database (CSD) were retrieved and modified, then optimized at PM3 semi-empirical level. The activities are pMIC = -log(MIC/mol$\cdot$l^{-1}), where MIC is Minimal Inhibitory Concentration of the drugs as extruded by two strains of *E. coli*: KAM32 strain without VmrA, and KAM32/pVCJ6 with VmrA (VmrA from *V. parahaemolyticus* AQ3334). The pMICΔ is a measure of the MDR effect of VmrA, defined as pMICΔ = abs[pMIC(pVCJ6) – pMIC(KAM32)].

MOLECULAR DESCRIPTORS

4

IMPORTANT MOLECULAR DESCRIPTORS

Table 2. Molecular descriptors above the cut-off (0.500) in correlation with the activities

No.	Symbol	Definition	R[KAM]*	R[pVCJ6]*
8	Iy	2nd principal moment of inertia	-0.262	**-0.621**
9	Iz	3rd principal moment of inertia	-0.216	**-0.654**
10	FF	log(Dip+1), Dip is molecular dipole moment (Titan)	0.431	**0.721**
14	Nh	No. hydrophobic carbon atoms	**-0.601**	-0.200
15	Na	No. aromatic carbon atoms	-0.469	**-0.645**
17	Np	No. polar (not hydrophobic) atoms (non-H)	0.203	**0.517**
26	Nr	No. ring atoms (non-H)	**-0.570**	-0.410
27	wh	Nh/Nt, number fraction of hydrophobic atoms; Nt is total No. non-H atoms	**-0.566**	**-0.726**
28	wa	Na/Nt, number fraction of aromatic atoms	-0.387	**-0.754**
29	wb	Nb/Nt, number fraction of hydrogen bonding non-H atoms; Nb is No. HB donors/acceptors	**0.562**	**0.579**
30	wp	Np/Nt, number fraction of polar atoms	**0.566**	**0.726**
31	wl	Nl/Nt, number fraction of planes atoms; Nl is No. non-H atoms in all planar fragments	-0.177	**-0.553**
40	wON	NON/Nt, number fraction of O/N atoms; NON is No. oxygen and nitrogen atoms	0.454	**0.558**
45	Bat	B/Nt, No. bonds per atom; B is No. bonds (non-H)	-0.415	**-0.603**
46	wr	Nr/Nt, number fraction of ring atoms	-0.425	**-0.713**
50	DM	molecular dipole moment (E4 method, MOPAC)	**0.694**	**0.788**
54	E4	average polarizability (E4 method, MOPAC)	**-0.611**	**-0.227**
55	BT	β hyperpolarizability along the dipole moment (E4 method, MOPAC)	**-0.663**	-0.456
59	GM	absolute average γ hyperpolarizability (E4 method, MOPAC)	-0.108	**-0.508**
63	Wi3	W(3D)', normalized Wiener index; W is H-depleted Wiener index	**0.510**	0.221
64	HOMO-1	energy of HOMO-1 orbital	**0.584**	0.375
65	HOMO	energy of HOMO orbital	**0.588**	0.376
68	Q-	the most negative ESP atomic charge (non-H)	**-0.672**	**-0.675**
69	Q+	the most positive ESP atomic charge (non-H)	**0.426**	**0.541**
74	QdF	(Q+) – (Q-), the largest ESP charge difference	**-0.539**	**-0.636**
76	%hspk	hspk/S, surface area fraction of hydrophobic C/H atoms; hspk and Sspk are CPK surface areas of hydrophobic C/H atoms and molecule, respectively	**-0.508**	**-0.508**
78	Mrefi	Mref/N, molecular refractivity per atom; Mref is molecular refractivity (ClogP method, CHEMSIDE) & No. all atoms	-0.259	0.606
79	Eshl	eHOMO-1/N, HOMO-1 orbital energy per atom	0.253	0.538
80	Esh	eHOMO/N, HOMO orbital energy per atom	0.252	0.529
83	Enh-Hi	[(HOMO-1)+(HOMO)]/N, frontier orbital energy sum (HOMO-1)+(HOMO) per atom	0.232	0.534
84	Enh-Hl	[(HOMO-1)+(LUMO)]/N, frontier orbital energy sum (HOMO-1)+(LUMO) per atom	0.328	**0.501**
87	HWn	HWN, HM per atom	**0.819**	**0.788**
88	HTn	HTN, hyperpolarizability β per atom	**-0.743**	**-0.881**
89	Bpa	B/Nt; No. bonds per atom; B is No. bonds	-0.165	**0.633**
90	L	No. non-H atoms along the longest bond chain	-0.215	**0.549**
93	Xmin	minimum X coordinate	**-0.398**	**-0.528**
97	Xmax	maximum X coordinate	**-0.638**	**-0.138**
98	Zmax	maximum Z coordinate	**-0.577**	**-0.045**
100	DY	Ymax-Ymin, molecular box width	**-0.543**	**-0.074**
101	DZ	Zmax-Zmin, molecular box height	**-0.538**	**-0.015**
104	sighb	Nb/Snt, HB donors/acceptors surface density; Snt is molecular surface area	0.496	**0.533**
105	sighyd	Nh/Snt, hydrophobic carbons surface density	**0.660**	**0.752**
106	sica	Na/Snt, aromatic carbons surface density	-0.347	**0.686**
107	sigp	Np/Snt, polar atoms surface density	**0.494**	**0.669**
108	sigr	Nr/Snt, ring atoms surface density	**-0.426**	**-0.671**
109	sigon	ON/Snt, ON atoms surface density	**0.575**	**0.536**
110	wb	Nb/Nt, non-H atoms per bond	**0.469**	**0.629**
111	Np2	eNp - 12nt, square function of Np	**-0.527**	**-0.787**
112	ON2	eON - 10nt, square function of ON	**-0.463**	**-0.593**
113	wb2	ewb - 0.6nt, square function of wb	**-0.682**	**-0.784**
114	wa2	ewa - 0.5)', square function of wa	**-0.468**	**0.627**
115	wp2	ewp - 0.4)', square function of wp	**-0.682**	**-0.784**
116	Mrefn2	eMrefn - 0.15)', square function of Mrefn	**-0.485**	**-0.745**
117	sigp2	(sig - 0.022)', sig = Nve/Snt, valence electron surface density; Nve is No. valence electrons	**-0.669**	**-0.674**
118	RD	ratio of actual and standard bond lengths sum	0.232	**0.662**
119	RD2	eRD - 0.98)', square function of RD	**0.583**	**0.787**
120	RRD	1/RD, inverse of RD	-0.243	**-0.673**

*Correlation coefficient with activity pMIC(KAM); [a]Correlation coefficient with activity pMIC(pVCJ6). [b]The bond lengths sums exclude H atoms. Actual bond lengths are from drugs 1 - 12 and standard bond lengths are analogue simple bond lengths are from the simplest molecules: C-C (H-CH$_3$), C-N (CH$_3$-NH$_2$), C-F (CH$_3$-FH$_2$), S-O (HS-OH), C-F (CH$_3$F), C-O (CH$_3$-OH), N-O (H$_2$N-OH), C-Cl (CH$_3$-Cl).

SELECTED MOLECULAR DESCRIPTORS

5

Table 3. Selected molecular descriptors for drugs 1 - 12

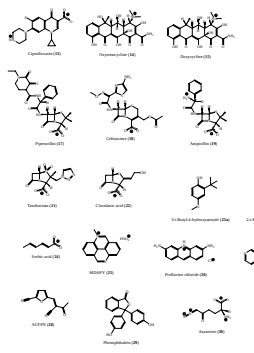
No.	FF	Nh	wa	E1+a.u.	HOMO-1 eV	sighyd/ A	Np2	wh2	Mpolh2/ cm^3 mol^{-1}
1	0.4236	14	0.8824	163.649	-7.857	0.0541	81	0.6500	0.2048901
2	0.6004	24	0.9600	215.097	-13.003	0.0622	121	0.1296	0.3856441
3	0.4807	14	0.8235	203.707	-11.666	0.0527	81	0.6500	0.1632216
4	0.5931	21	0.8333	241.188	-11.251	0.0580	81	0.6756	0.1660001
5	0.8224	10	0.3000	129.924	-10.264	0.0303	4	0.0100	0.1705569
6	1.6830	14	0.6957	173.499	-6.345	0.0396	9	0.0001	0.0155625
7	0.6896	27	0.6061	356.642	-11.168	0.0569	36	0.6176	0.0376536
8	1.1583	19	0.5625	219.420	-11.366	0.0135	1	0	0.002601
9	2.3617	35	0.5098	316.371	-11.867	0.0166	16	0.0071	0.695181
10	1.7979	20	0.1250	249.815	-13.862	0.0329	64	0.0100	0.6768178
11	1.3921	23	0.6071	176.492	-4.333	0.0526	49	0.6400	0.0378961
12	1.6981	12	0	110.602	-0.124	0.0424	49	0.0112	0.0665504

The descriptors were calculated by Titan, MOPAC and Chem3D using optimized geometry, and from molecular formula.

ACKNOWLEDGEMENT: FAPESP

PLS (PARTIAL LEAST SQUARES) AND PCR (PRINCIPAL COMPONENT REGRESSION) PREDICTION OF VmrA RESISTANCE

6



Prediction set of drugs and xenobiotics. *V. parahaemolyticus* is sensitive to clinically used drugs against this bacterium (13-22) and to some agents (23, 24). The bacterium is probably resistant to other xenobiotics from this set (25-30).

7

Table 4. PLS and PCR regression vectors[a] and statistics[b]

Parameter	PLS (pVCJ6)	PCR (pVCJ6)	PLS (KAM)	PCR (KAM)
FCs (%)	1 (70.0)	1 (70.0)	1 (52.8)	1 (53.6)
SEV	0.721	0.683	0.452	0.443
SEP	0.562	0.565	0.368	0.389
Q	0.893	0.818	0.762	0.763
R	0.889	0.888	0.866	0.849
Δ	0.410	0.410	0.288	0.270
FF	0.206	0.209		
Nh			-0.228	-0.260
wa	-0.204	-0.212		
Il4			-0.231	-0.261
HOMO			0.223	0.190
sighyd	-0.215	-0.217	-0.250	-0.258
Np2	+0.225	+0.215		
wh2			-0.258	-0.250
Mrefn2	+0.213	+0.210		

[a]Regression vectors and statistical parameters are given for PLS and PCR modelling of efflux activity of the *E. coli* strains KAM32/pVCJ6 (pVCJ6) and KAM32 (KAM). [b]Parameters: FCs (%) – the number of principal components and % of the total variance that they contain; SEV – standard error of validation; SEP – standard error of prediction; Q – correlation coefficient of validation; R – correlation coefficient of prediction; Δ – average absolute deviation of predicted from experimental activity values.

The PLS & PCR regression statistics for modeling of the both activities shows that there is no significant difference between the PLS and PCR models. The regression coefficients are in favour for pMIC(pVCJ6), whilst the errors are smaller for pMIC(KAM).

8

Table 5. Prediction of the MDR character of *V.parahaemolyticus* to 1-30 by means of PLS and PCR

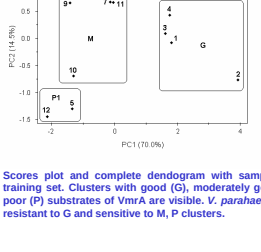
No.	pMICΔ (PLS)	pMICΔ (PCR)	MDR character	Correct prediction	No.	pMICΔ (PLS)	pMICΔ (PCR)	MDR character	Correct prediction
1	1.687	1.720	resistant	+	17	1.137	1.043	resistant	–
2	1.710	1.821	resistant	+	18	0.820	0.783	resistant	–
3	1.259	1.311	resistant	+	19	0.099	0.075	sensitive	+
4	0.903	0.951	resistant	+	20	0.635	0.583	sensitive	+
5	0.454	0.512	?	–	21	0.451	0.482	sensitive	+
6	0.340	0.313	sensitive	+	22	0.614	0.643	resistant	–
7	0.246	0.277	sensitive	+	23a	1.059	1.123	resistant	–
8	0.123	0.138	sensitive	+	23b	1.052	1.117	resistant	–
9	0.872	0.968	resistant	–	24	0.844	0.810	resistant	–
10	0.196	0.230	sensitive	+	25	1.774	1.872	resistant	+
11	0.523	0.499	?	–	26	1.455	1.506	resistant	(–)
12	0.247	0.211	sensitive	+	27	1.604	1.671	resistant	+
13	0.411	0.382	sensitive	+	28	0.625	0.649	resistant	+
14	0.079	0.091	sensitive	+	29	1.281	1.302	resistant	+
15	0.075	0.087	sensitive	+	30	1.016	1.149	resistant	(–)?
16	0.138	0.304	sensitive	+					

The PLS & PCR regression models in predicting the MDR character of *V. parahaemolyticus* to 1-30. The cut-off criterion was 0.5 in pMICΔ units, what seems not suitable for 9 agents. Other cut-off value or another methodology should be applied to eliminate wrong and doubtfull predictions. Error accumulation in both predicted pMIC(pVCJ6) and pMIC(KAM32) may contribute to wrong predictions. However, 20 of 31 (65%) agents were predicted correctly as agents to which the bacterium would be resistant or sensitive.

PRINCIPAL COMPONENT ANALYSIS (PCA) AND HIERARCHICAL CLUSTER ANALYSIS (HCA) IN PREDICTION OF VmrA RESISTANCE

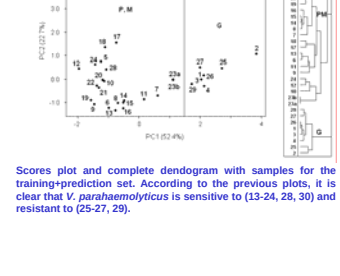
EXPLORATORY ANALYSIS FOR pVCJ6 DATA

9



Scores plot and complete dendogram with samples for the training set. Clusters with good (G), moderately good (M) and poor (P) substrates of VmrA are visible. *V. parahaemolyticus* is resistant to G and sensitive to M, P clusters.

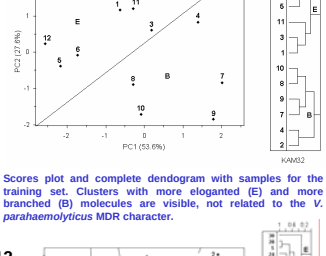
10



Scores plot and complete dendogram with samples for the training+prediction set. According to the previous plots, it is clear that *V. parahaemolyticus* is sensitive to (13-24, 28, 30) and resistant to (25-27, 29).

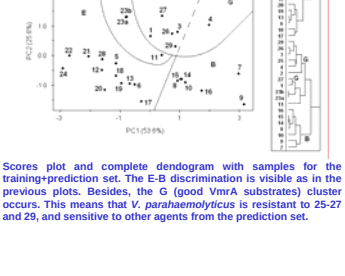
EXPLORATORY ANALYSIS FOR KAM32 DATA

11



Scores plot and complete dendogram with samples for the training set. Clusters with more eloganted (E) and more branched (B) molecules are visible, not related to the *V. parahaemolyticus* MDR character.

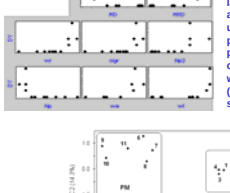
12



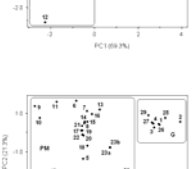
Scores plot and complete dendogram with samples for the training+prediction set. The E-B discrimination is visible as in the previous plots. Besides, the G (good VmrA substrates) cluster occurs. This means that *V. parahaemolyticus* is resistant to 25-27 and 29, and sensitive to other agents from the prediction set.

EXPLORATORY ANALYSIS WITH EIGHT DISCRIMINATORY DESCRIPTORS

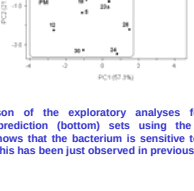
12



The pMICΔ-descriptor correlation plots. These descriptors account for the content of unsaturated bonds, ring and planar fragments, aromatic and polar groups. The descriptors clearly distinguish the agents to which the bacterium is resistant (1-4) from those to which it is sensitive (5-12).



13



Comparison of the exploratory analyses for the training (top) and training+prediction (bottom) sets using the discriminatory descriptors clearly shows that the bacterium is sensitive to the PM and resistant to G cluster. This has been just observed in previous analyses.