JOURNAL OF CHEMOMETRICS
J. Chemometrics 2004; **18**: 242–252
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/cem.867

QSAR study of β-lactam antibiotic efflux by the bacterial multidrug resistance pump AcrB†

Márcia M. C. Ferreira* and R. Kiralj

Laboratório de Quimiometria Teorica e Aplicada, Instituto de Química, Universidade Estadual de Campinas, Campinas, SP 13084-981, Brazil

Received 22 January 2004; Revised 30 April 2004; Accepted 2 May 2004

AcrAB-TolC is the most important efflux pump system of Gram-negative bacteria, responsible for their resistance to lipophilic and amphilic drugs. In this work, HCA–PCA studies were performed to investigate the relationship between efflux activities (negative logarithm of minial inhibitor concentration, pMIC) of three strains of *S. thypimurium* with respect to β-lactams, and to analyze the relationship between lipophilicity parameters calculated by different methods. The analyses demonstrate that pMICs strongly depend on properties of both bacterial strains and drug molecules, and that lipophilicity parameters do not necessarily contain the same information about the drugs. QSAR studies have shown that the calculated lipophilicities, in some cases, are non linearly related to the pMICs originated by active AcrAB-TolC bacterial pumps, due to the existence of β-lactams with nitrogen- and sulfur-rich substituents. Among the most important β-lactam molecular properties quantitatively related to pMICs are lipophilicity and electronic and hydrogen–bonding properties. Parameters describing these properties were included in the QSAR study to obtain parsimonius regression models for MICs. β-Lactams were classified as good, moderately good and poor AcrAB-TolC substrates. Their stereoelectronic molecular properties, especially the Y-component of the molecular dipole moment and hydrogen binding properties, reflected this classification. Copyright © 2004 John Wiley & Sons, Ltd.

KEYWORDS: AcrB multidrug resistance efflux pump; β-lactam antibiotics; QSAR; PLS

1. INTRODUCTION

β-Lactam antibiotics are the most widely used antibacterial agents, primarily inhibiting penicillin-binding proteins (PBPs) responsible for the construction and maintenance of the bacterial cell wall. Resistance of bacteria to β-lactams and other antibiotics is becoming an increasing problem in the treatment of infectous diseases. A major means of this resistance is the overproduction of multidrug resistance (MDR) efflux pumps, which excrete a wide range of compounds [1–4]. Some MDR efflux pumps play a very important role in bacterial resistance to antibiotics [5–7]. Moreover, unknown physiological roles of MDR pumps, their possible use as antibacterial targets and as aids in cell-based screening for novel antibacterial compounds, make these membrane transport systems very attractive in medical, biochemical and chemical studies [4].

*Correspondence to: M. M. C. Ferreira, Laboratório de Quimiometria Teorica e Aplicada, Instituto de Química, Universidade Estadual de Campinas, Campinas, SP 13084-862, Brazil.
E-mail: marcia@iqm.unicamp.br
†Part of this paper was presented at the *8th Scandinavian Symposium on Chemometrics*, Mariehamn, Finland, 14–18 June 2003.
Contract/grant sponsor: State of São Paulo Funding Agency.

The AcrAB-TolC efflux pump and its homologs, identified in *Escherichia coli* [4,8], *Salmonella typhimurium* [9], *Salmonella enterica* [5], *Haemophilus influenza* [10] and other bacteria, possess intrinsic resistance to a wide variety of lipophilic and amphiphilic toxic compounds from the bacteria's natural environment (bile salts, detergents, fatty acids, organic solvents, cationic dyes, etc.) and to drugs (antibiotics, antiseptics, chemotherapeutic agents, etc.) [1–4,11–13]. This pump consists of the inner-membrane transporter AcrB [14–17] which belongs to the resistance-nodulation-cell division superfamily (RND) [3], the outer-membrane channel-tunnel protein [18] of the TolC family [17], and the periplasmic linker lipoprotein AcrA from the membrane fusion family [19]. AcrAB-TolC pump exports toxic compounds from the cytoplasmic or periplasmic space directly into the external medium, bypassing the outer membrane barrier [7,14,18,20].

X-ray and electron diffraction structure determinations of AcrA [21], AcrB [14–17] and TolC [18] enabled a new insight into structure–function relationships and efflux mechanisms for the AcrAB-TolC pump. On the other hand, one of the factors which significantly influences the multidrug efflux rate is the AcrB substrates, considered as structurally unrelated even if attention is payed only to antibiotics [2,6]. The

substrates can be neutral or bear multiple charged groups, ranging from small molecules such as *n*-hexane [11] and nalidixic acid [8], β-lactams such as nafcillin and cephalosporin C [9], to relatively large molecules such as erythromycin [2] and rifampicin [1]. Paulsen *et al.* [3] pointed out that drugs' physical characteristics, such as charge, lipophilicity or amphiphilicity, rather than their structural characteristics, are to be a key determinant in specificity of proton motive MDR pumps. In a simple analysis Nikaido *et al.* [9] noticed that there was some quantitative relationship between the β-lactam side chain lipophilicity and MICs (minimal inhibitor concentration or drug efflux rate).

The goal of the present study was to establish multivariate quantitative lipophilicity–MIC and β-lactam structure–MIC relationships for *S. typhimurium* strains [9]. In spite of the general opinion that the structures of AcrB antibiotic drug substrates have not much in common, this study relied on structure and structural characteristics (molecular descriptors) of the substrates, which definitely determine their physical characteristics and biochemical behavior. This approach is in accord with quantitative structure–activity relationship (QSAR) [22–29] and structure–activity relationship (SAR) [30–32] studies recently performed in our group on various classes of drugs. The QSAR study in this work was performed by means of the partial least squares (PLS) [33–36] method. The relationships in the set of various lipophilicity parameters and in the set of MICs originating from different bacterial strains were investigated by means of principal component analysis (PCA) and hierarchical cluster analysis (HCA) [34–36].

2. METHODS

2.1. MICs for bacterial strains

The MIC of a drug with respect to a bacterial strain is defined as the lowest concentration of the drug at which no growth of the strain is observed in a period of time and under specified experimental conditions. Mass concentration MICs for 16 penicillines and cephalosporins (Figure 1) with respect to bacterial strains *S. typhimurium* SH5014 (parent strain) and its mutants low resistant SH7616 and AcrAB overproducer HN891, were used from the literature [9]. In this work negative logarithms of molar MICs, pMICs, are used (Table I).

2.2. Modeling of drugs

Molecular structures of the 16 β-lactam antibiotics, *n*-hexane, erythromycin and rifampicin were refined or modeled by molecular modeling package PC Spartan Pro 1.0.5 [37] using atomic coordinates from the 3D Pharmaceutical Structure Database [38], the Cambridge Structural Database [39] or two-dimensional formula (Figure 1). A conformational search for the molecules (except for semirigid large erythromycin and rifampicin) was performed by the Monte Carlo molecular dynamics method incorporated in the Spartan package, and the most stable conformers were optimized by semi-empirical molecular orbital method PM3 [40] in the package.

2.3. Lipophilicity parameters

Logarithm of the octanol–water partition coefficient $\log K_{OW}$ was from Nikaido *et al.* [9] (Table II). As many computer programs do not calculate lipophilicity contribution for a number of charged and delocalized functional groups, cationic forms for **6** and **9** (with protonated carboxylates) and neutral forms for other β-lactams were modeled and optimized using the Spartan package as described above, and the gas-phase lipophilicity ($\log P_{GC}$) was calculated. Free online JME Molecular Editor was used to model these species in SMILES format and the octanol–water partition coefficient ($\log P_w$) was calculated by the IA_log*P* Predictor at the Interactive Analysis log*P* and log*W* Predictors website [41]. Submitting SMILES files to a free web program ALOGPS 2.1 [42], lipophilicity parameters based on different computational approaches were calculated: $\log P_s$, $\log P_{IA}$, $\log K_{WIN}$ and $\log P_X$ (originally named ALOGPs, IA_LOGP, KOWWIN and XLOGP, respectively). The number fraction w_C of hydrophobic carbon atoms, defined as the number of hydrophobic carbon atoms (all carbon atoms except those in C═O, C–O$^-$ and CN groups) divided by the number of all non-hydrogen atoms, was calculated from two-dimensional chemical formulae. S_f, the surface fraction of hydrophobic carbon atoms, was calculated analogously to w_C: instead of atom counts, their CPK [43] atomic surface areas from optimized geometries of compounds (in charged forms at neutral pH, see Figure 1) were used. Parameters w_C and S_f were considered as structure-based lipophilicity parameters.

2.4. Other molecular properties

Geometrical, electronic and hydrogen bond (HB) parameters, were calculated for penicillins and cephalosporins. Based on two-dimensional chemical formulae, the following parameters were counted or calculated: the number of charged groups (N_{CH}); the number of nitrogen and sulfur atoms (N_{NS}, sulfonate sulfur excluded); the number fraction of heteroatoms (H_f, all N, S, O, Cl atoms counted with respect to all non-hydrogen atoms); the number of all π- and lone pair electrons divided by CPK [43] molecular surface (σ); the number of hydrogen bond acceptors (A_{HB}, which is equal to the number of lone pairs in N atoms, and in carbonyl, oxide, sulfonate and hydroxyl O atoms); the sum of overal numbers of valence electrons for substituents R and R_1 reduced by 8 (Z, hydrogen atoms excluded; two methyl groups were considered as R_1 in penicillins); and some other properties. Geometrical and electronic properties, like dipole moment (D) and its components, the difference between the energies of the highest occupied and the lowest unoccupied molecular orbitals (Δ), were calculated by the Spartan package. Polarizability properties like the third-order molecular polarizability (γ) were calculated using semi-empirical method PM3 in MOPAC [44]. There were around 50 calculated molecular properties.

2.5. Chemometrics: principal component analysis and hierarchical cluster analysis

Most chemical, biochemical and biological applications of data analysis are by nature multivariate, and some of the most suitable methods for such cases are PCA and HCA, which are used in this work. HCA and PCA analyses were performed for biological activities, pMICs, so the data matrix had dimensions 16×3 (16 = the number of β-lactam antibiotics; 3 = the number of different types of pMIC originated

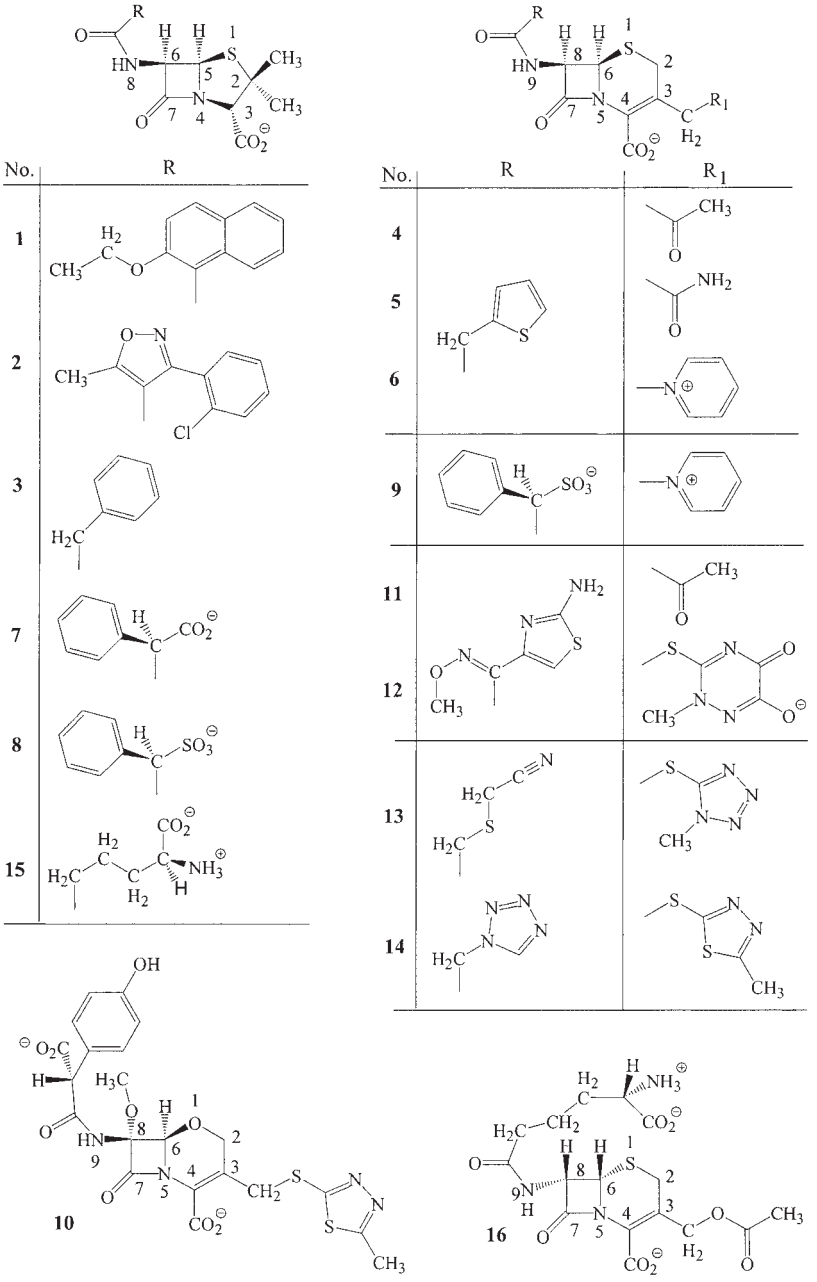

Figure 1. Chemical structures of β-lactams at neutral pH, with the atomic numbering for penicillines and cephalosporins.

from bacterial strains). Furthermore, the lipophilicity parameters were organized into a data matrix 16×9 or 16×7 ($16 =$ the number of β-lactams; 9 or $7 =$ the number of lipophilicity parameters including and excluding the two structure-based lipophilicity parameters). All data were always autoscaled, and incremental linkage method was used in HCA.

2.6. Chemometrics: regression methods

Some lipophilicity variables were non-linearly related to the pMICs, so square x^2 and gaussian $\exp[-(x - x_o)^2]$ terms of some lipophilicity parameters x were introduced. The selected lipophilicity variables (absolute correlation coefficient with pMICs over 0.4) were then used to build the PLS regression models for each pMIC, and the models were validated by leave-one-out cross-validation. The variable selection for other molecular parameters was performed, and PLS models for each pMIC including lipophilicity, electronic and HB parameters were constructed and internally validated. In the final stage of regression analysis, three samples from the training set were excluded, and the PLS models based on all types of parameters were rebuilt and externally validated by these three samples. All statistical and chemometric analyses were carried out by using software Matlab 6 [45] and Pirouette 3.01 [46].

3. RESULTS AND DISCUSSION

3.1. PCA and HCA of biological activities

PCA and HCA on experimental biological activities (Table I) were carried out in order to extract structural information on both the antibiotics and the bacterial strains. One should bear

Table I. Biological activities[a] for penicillines and cephalosporins

No.	Compound	Form[b]	HN891	SH5014	SH7616
1	Nafcillin	Anion	2.310	2.607	4.714
2	Cloxacillin	Anion	2.629	2.930	5.338
3	Penicillin G	Anion	4.019	4.621	6.126
4	Cephalothin	Anion	4.695	4.996	5.598
5	Cefoxitin	Anion	4.427	5.029	5.631
6	Cephaloridine	Zwitterion	4.715	4.715	5.017
7	Carbenicillin	Dianion	4.073	4.675	5.277
8	Sulbenicillin	Dianion	4.112	4.714	5.316
9	Cefsulodin	Anion-zwitterion	3.919	3.919	3.919
10	Latamoxef	Dianion	6.318	6.637	6.637
11	Cefotaxime	Anion	5.959	6.579	7.181
12	Ceftriaxone	Dianion	6.364	6.665	6.966
13	Cefmetazole	Anion	5.674	5.975	5.975
14	Cefazolin	Anion	5.055	5.357	5.357
15	Penicillin N	Anion-zwitterion	4.652	4.652	4.652
16	Cephalosporin C	Anion-zwitterion	4.414	4.414	4.414

[a]Biological activities: $pMIC(X) = -\log(X)$, where X is minimal inhibitor concentration (MIC) with respect to strain HN891, SH5014 and SH7616.
[b]Species at neutral pH, according to Nikaido *et al.* [9].

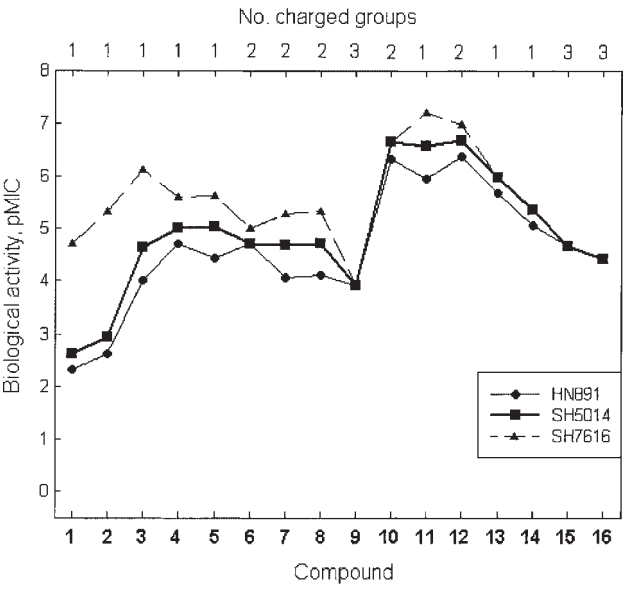

Figure 2. Comparison of pMICs for β-lactams. pMIC for *S. typhimurium* SH5014 strain is placed in between pMICs for mutants SH7616 and HN891. The number of charged groups in antibiotic molecules at neutral pH shows that different bacterial strains are not distinguished well when excreting highly charged antibiotics.

in mind that MIC of an antibiotics means the minimal molar concentration of this drug necessary to stop bacterial development. Consequently, drugs with high MIC (or low pMIC) are good AcrAB-TolC substrates. Figure 2 demonstrates correlations among pMICs. There is a high correlation between pMICs for strain SH5014 and its mutant HN891 (correlation coefficient $r = 0.98$), and lower between pMICs for SH5014 and other mutant SH7616 ($r = 0.76$). This finding is in agreement with the fact that the strain HN891 is more similar to the parent SH5014 than SH7616, due to the strain preparations [9].

The number of charged groups, N_{CH}, in β-lactam molecules seems to be important for the three pMICs originating from different bacterial strains (Figure 2). There are three β-lactams with three charged groups, i.e. zwitter-anions **9**, **15** and **16**. They are not distinguished by the three bacterial strains (see Figure 2 where the three curves coincide). Nikaido *et al.* [9] pointed out that such β-lactams are hydrophilic, and in contrast to lipophilic ones, are equally poorly distributed in the membrane bilayer of any bacterial strain. β-Lactams with two charged groups have moderate differences in pMICs corresponding to different bacterial strains: zwitter-ion **6** and dianions **7**, **8**, **10** and **12**. Anionic β-lactams **1**–**5**, **11**, **13** and **14** bear only one charge, and are exctreted quite differently by the three AcrAB-TolC efflux pumps (see Figure 2). One can notice in the same figure that there is almost parallel increase/decrease of the three pMICs for most of the antibiotics. Some of them (**2**–**5** and **11**) disrupt this parallelism.

The first two PCs are sufficient to describe over 99.6% of the the original information. Molecules in the PCA scores plot (Figure 3, left) are labeled by N_{CH} and the number

Table II. Lipophilicity descriptors for penicillins and cephalosporins

No.	$\log K_{OW}$[a]	$\log P_w$	$\log P_s$	$\log P_{IA}$	$\log K_{WIN}$	$\log P_X$	$\log P_{GC}$	w_C	S_f	$\log P_{exptl}$[b]
1	3.62	3.14	3.21	3.14	3.79	3.28	2.105	0.621	0.666	
2	2.95	2.57	2.61	2.57	3.22	3.04	0.905	0.552	0.555	2.48
3	2.43	1.10	1.92	0.80	1.85	1.49	−0.201	0.565	0.614	1.83
4	2.11	−0.01	0.62	−0.01	0.16	−0.20	−1.771	0.462	0.498	0.00
5	2.11	−1.16	−0.23	−1.16	−0.72	−1.08	−1.338	0.429	0.452	
6	2.11	−1.25	−1.30	1.09	−3.53	0.45	−0.618	0.571	0.609	
7	1.90	1.21	0.96	2.19	1.19	1.84	0.477	0.500	0.538	1.13
8	1.90	0.42	−0.20	0.42	0.81	0.39	−0.013	0.481	0.505	0.59
9	1.90	−1.07	−1.48	1.55	5.64	−0.83	−0.807	0.500	0.511	
10	1.30	−2.46	−0.57	−2.46	−2.20	−0.53	1.033	0.444	0.470	
11	0.78	−0.77	0.13	−0.77	0.64	−0.55	−1.590	0.400	0.428	
12	0.78	−1.22	0.52	−1.66	−1.05	−0.32	−0.541	0.361	0.378	
13	0.30	0.08	−0.62	0.08	−1.87	−1.80	1.139	0.367	0.403	−0.62
14	−0.30	0.80	0.01	0.80	−1.87	−1.04	0.310	0.379	0.406	
15	−2.40	−2.85	−1.43	−2.85	−3.29	−2.95	−1.672	0.417	0.460	
16	−2.40	−2.14	−2.20	−2.14	−4.80	−3.70	−2.976	0.393	0.494	

[a]K_{OW} was reported as ~0 for samples **15** and **16**. Original K_{OW} from Nikaido *et al.* [9] was transformed into $\log(K_{OW}+1)$ in this work.
[b]Experimental values for log*P* for octanol–water partition, obtained from ALOGPS software.

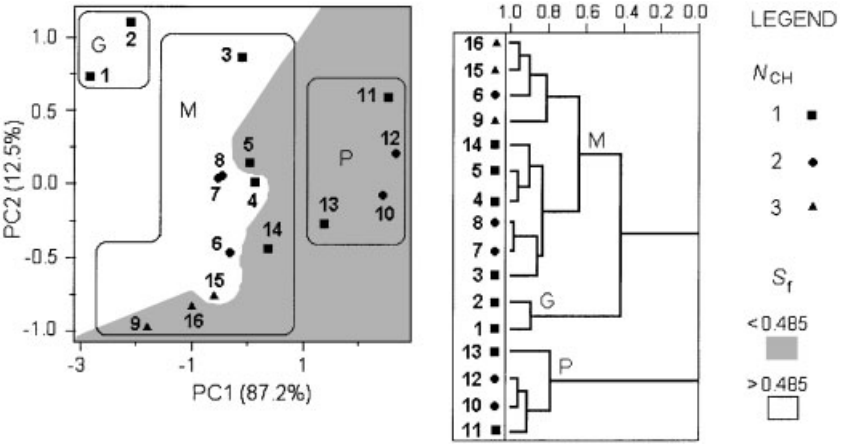

Figure 3. Left: PCA scores plot shows the position of β-lactams in the space defined by principal components PC1 and PC2. Right: HCA dendogram of samples (β-lactams). Both plots exhibit clustering of good (G), moderately good (M) and poor (P) AcrB substrates. Also, other clustering of the antibiotics with respect to their molecular properties (N_{CH}, S_f) is visible.

fraction S_f of hydrophobic carbon atoms. These labels can help in explaining the relationships between antibiotic molecular properties and the pMICs. Roughly, PC1 discriminates β-lactams with low pMICs (good AcrAB-TolC substrates, cluster G consisting of **1** and **2** on the left-hand side of the plot) and moderate pMICs (moderate pump substrates, cluster M) from those with high pMICs (poor AcrAB-TolC substrates, **10**–**13**, cluster P). The second PC (PC2) discriminates molecules with respect to their variability in pMICs for different bacterial strains (as observed in Figure 2). Note that the molecules **2**–**5** and **11**, which have already shown unusual behavior in Figure 2, are in the upper part of the PCA plot. Furthermore, PC2 is related to the number of charged groups N_{CH}. The hydrophobicity label S_f also shows variations, by decreasing along PC1 and slowly increasing along PC2. This result is in accord with the observation of Nikaido *et al.* [9] that MICs are proportional to the antibiotic side-chain hydrophobicity. Clustering of the β-lactams in HCA (Figure 3, right) also includes clusters G, M and P from PCA. Going from G to M and further to P cluster in PCA and HCA, one can notice that in average, molecules have more charged groups and less hydrophobic carbon atoms, although the cluster M contains a large variety of β-lactams. The samples dendogram reveals two-membered sub-clusters of structurally similar molecules (**4**, **5**), (**7**, **8**), (**15**, **16**) and (**10**, **12**), with a similarity index over 0.90, which is an indication for existence of structure–activity relationships for β-lactam antibiotics.

The presented PCA and HCA study on pMICs leads to the conclusion that the presence of charged groups and hydrophobic moieties determines the antibiotic behavior with respect to AcrAB-TolC pumps. Highly charged β-lactams with small hydrophobic fraction w_C or S_f might be poor AcrB substrates due to energetically unfavorable pump–drug interaction in molecular recognition.

3.2. PCA and HCA of lipophilicity parameters

In ideal case, the parameters log$P$ and log$K$ for β-lactams (Table II) should be highly correlated, form a compact cluster and be described by only one PC. The purpose of PCA–HCA study of these parameters is to reveal how much they deviate from this ideal situation and, consequently, how reliable lipophilicity–MIC relationships can be established and used for the study of pump–drug interactions.

Nikaido *et al.* [9] calculated the lipophilicity contribution of 6-substituents in penicillins and 7-substitutents in cephalosporins. This approach does not take into account the difference in the β-lactam ring of penicillins and cephalosproins, nor the presence of other substituents in cephalosporins (7-methoxy in **5**, **10**, **13**, and various R_1 substituents; oxazolidine ring in **10**).

The PCA–HCA study of β-lactam lipophilicity in this work included lipophilicity logK_{OW} of Nikaido *et al.* [9] and calculated parameters logP_w, logP_s, logP_{IA}, logK_{WIN}, logP_X, logP_{GC}, w_C and S_f. The correlation coefficients among them range from 0.29 to 0.96 (0.56 to 0.87 if excluding w_C and S_f). These parameters make a heterogeneous data set and, when compared with experimental log$P$ of some drugs (Table II), sometimes give quite different results. The first three PCs describe only 89.3% of the original data, which is far from the ideal case. There is no simple explanation why the lipophilicity parameters form three-membered (logK_{OW}, S_f, w_C) and four-membered (logP_w, logP_s, logP_X, logP_{IA}) clusters in PCA and HCA (Figure 4) while two lipophilicity parameters (logP_{GC} and logK_{WIN}) are isolated. Removal of w_C and S_f causes some changes in both PCA and HCA. The percentage of total variance explained by the first three PCs (89.8%) is slightly increased, logP_{GC} and logK_{WIN} are farther from others which are organized in one cluster (results not shown) and the scores are also changed. In Figure 5 (based on seven lipophilicity parameters, i.e. excluding w_C and S_f), four groups can be seen based on the structure of substituent R: aromatic two-ring substituent (group I), phenyl group in substituent (group II), an other π-system in substituent (five-membered ring or CN, group III), and an aliphatic substituent (group IV). This arrangement of samples along PC1 reveals the increase in lipophilicity with chemical structure and aromatic character of substituent R. Two more facts can

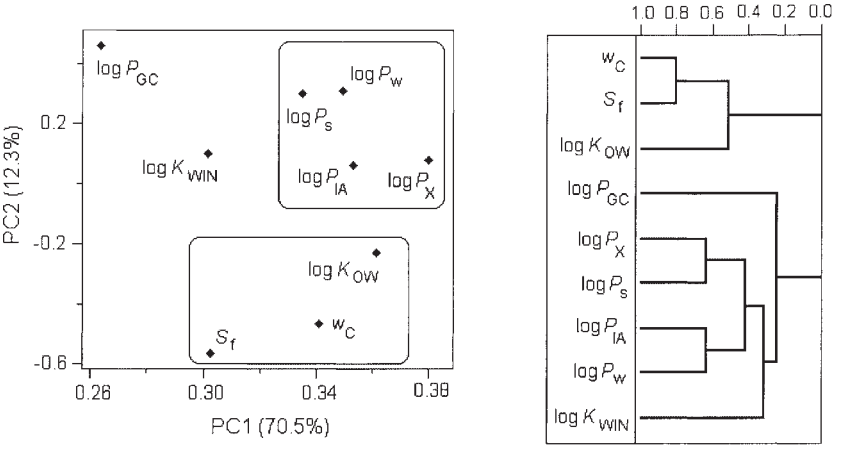

Figure 4. PCA loadings plot (left) and HCA dendogram on variables (right) for nine lipophilicity parameters.

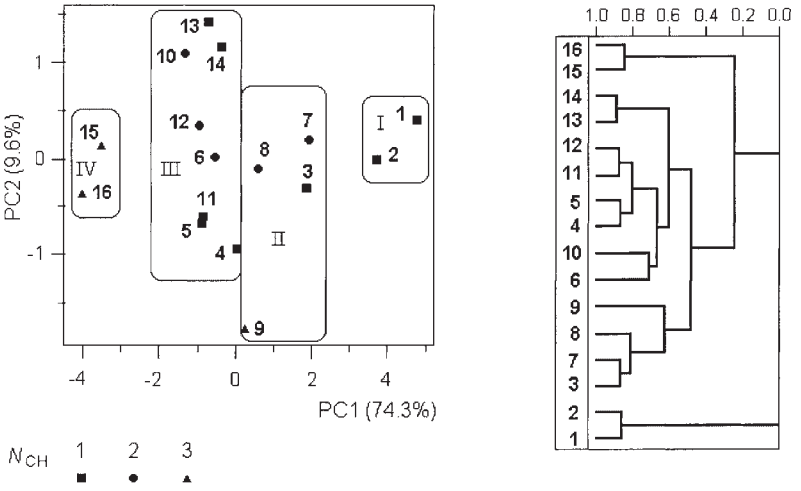

Figure 5. PCA scores plot (left) and HCA dendogram on samples (right) for seven lipophilicity parameters.

be stated. First, samples with N_{CH} equal to 1, 2 or 3 occupy definite areas in the PC1–PC2 space. Second, samples with a ring in R_1 (**1–5, 7, 8, 11, 15, 16**) are placed mainly in the central part with respect to PC2. Similar trends in the scores plot may be observed if S_f and w_C are included in the data set (results not shown).

The present results on lipophilicity parameters provide a general picture of the β-lactams under study, but there still remains the question of which lipophilicity parameters: the most reliable for these antibiotics, especially in QSAR study. Is lipophilicity, besides the number of charged groups N_{CH}, the main or only antibiotic property which determines drug efflux by bacterial AcrAB-TolC pump? Answers may be obtained from regression models employed in QSAR study.

3.3. Quantitative lipophilicity–pMIC and structure–pMIC relationships for β-lactams

In QSAR studies lipophilicity parameters are important variables [47,48]. Usually they are linearly related to pharmacological activity (MICs), but in the more general case this relationship is not linear [48–50]. Nikaido *et al.* [9] discussed the linear relationship between MIC and the logK_{OW}. In this work, the relationship between pMICs and lipophilicity parameters (Table II) is studied by chemometric methods in order to gain more insight into the mechanism of bacterial efflux pump–drug interactions. Visual inspection of data indicates that the three biological activities pMIC(HN891), pMIC(SH5014) and pMIC(SH7616) are probably in non-linear relationship to logK_{OW}, logP_s, logK_{WIN} and logP_X, and are linear with respect to w_C and S_f. Relationship pMIC(HN891)–logK_{OW} is a good illustrative example for such non-linearity (Figure 6), where the points lie along a

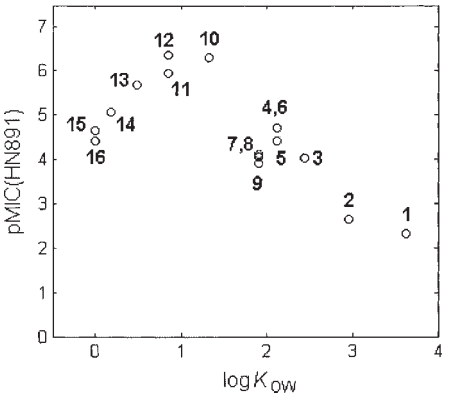

Figure 6. An example of nonlinear pMIC–lipophilicity relationship.

Table III. Some electronic and hydrogen bond descriptors for penicillins and cephalosporins

No.	N_{CH}	N_{NS}	H_f	Z	A_{HB}	σ (Å^{-2})	Δ (eV)	D (Debye)	D_y (Debye)	γ (a.u.)[a]
1	1	3	0.276	0.154	10	0.098	−5.74	25.95	−24.264	4.234
2	1	4	0.345	0.462	11	0.116	−5.96	21.00	−14.409	3.852
3	1	3	0.304	0.000	10	0.095	−6.57	23.21	−22.727	2.595
4	1	4	0.385	1.333	12	0.123	−6.67	15.34	−9.352	3.353
5	1	5	0.429	1.333	13	0.132	−7.63	10.10	7.860	2.690
6	2	5	0.321	0.476	10	0.106	−6.45	14.59	−12.063	6.439
7	2	3	0.346	0.400	14	0.120	−7.64	15.11	−2.689	3.216
8	2	3	0.407	0.727	16	0.129	−7.78	15.72	9.820	3.289
9	3	5	0.389	1.127	19	0.137	−6.75	8.09	−2.429	5.074
10	2	7	0.444	1.295	20	0.145	−7.55	5.28	1.186	7.153
11	1	7	0.467	1.500	15	0.137	−7.48	12.90	−4.780	6.350
12	2	11	0.500	1.518	21	0.149	−5.42	25.55	6.850	12.252
13	1	10	0.500	1.350	15	0.134	−8.04	11.60	6.849	6.698
14	1	11	0.517	1.417	16	0.139	−7.36	4.40	0.840	5.501
15	3	4	0.417	0.625	14	0.108	−8.51	7.97	1.838	2.108
16	3	4	0.429	1.475	16	0.126	−8.24	10.17	−5.502	3.225

[a]Atomic units.

guassian-like curve. The curve maximum is approximately at $\log K_{OW}$ values for the poorest substrates, **10–13** and **14**. The terminal $\log K_{OW}$ values belong to the best substrates **1–3**, and moderately good to poor **15** and **16** (see Figure 3 for the classification of the substrates). Surprinsingly, **10–14** lie in the middle of the $\log K_{OW}$ range. What is a chemical rationale for this behavior? Table III contains selected electronic and hydrogen bonding (HB) parameters which are moderately to highly correlated to the pMICs. Three of them, the heteroatomic contribution H_f, the surface electron density of π and free pair electrons σ_π, the number of nitrogen+sulfur atoms N_{NS}, can aid qualitatively to the response. β-Lactams **10–14**, in contrast to **1–3** and **15**, **16**, are characterized by high N_{NS}, which have as a consequence high σ_π and H_f. This is because substituents R and R_1 in **10–14** are rich in heteroaromatic π-systems, including N and S atoms. In terms of lipophilicity, polarity, electron density and other molecular properties, these substituents are between hydrocarbon and oxygen-rich side chains. Good AcrAB-TolC substrates **1–3** possess highly lipophilic substituents (mainly hydrocarbon moiety), large R and small R_1. R and R_1 in **10–14** must disable good bacterial pump–drug interaction, which, according to molecular properties of **1–3**, should occur between hydrophobic R, R_1 and hydrophobic residues of AcrB. On the other hand, R and R_1 in **10–14** have many hydrogen bond acceptors (nitrogen atoms) in relatively small substituents, and can hardly find a compatible hydrogen donor network in AcrB. Besides, these R and R_1 are defficient in hydrophobic hydrogen atoms, which are necessary to establish many weak drug–receptor interactions [51–54].

Correlation coefficients between biological activity (Table I) and lipophilicity parameters (Table II) varied in a range of 0.07–0.76. Introducing new square and guassian terms for some lipophilicity parameters, the correlation coefficients reached the values 0.76–0.90. Selected lipophilicity parameters which produced the best PLS models are presented in Table IV for each bacterial strain (the top models). It is obvious that the PLS models related to strains HN891 and SH5014 are similar, even having the same molecular descriptors. This just confirms the previous results in this work that these two strains act similarly, and the third is significantly different from each of them. On the other hand, having nine lipophilicity parameters for each activity leads to diversity of lipophilicity parameters and not only one that will produce a reasonably good model. As each method for calculating lipophilicity parameters might introduce new information which may be contained in some electronic descriptors, the presence of all lipophilicity parameters in a regression model is reasonable. Table IV also contains PLS models based on parameters of all types (the bottom models). One can notice that the models based only on lipophilicity were further improved by including electronic and hydrogen bonding parameters, even in the case of strain SH7616. However, what still remains is that this strain is different from the other two. As this strain contains AcrAB-TolC pumps that are ineffective, the efflux of

Table IV. PLS regression models for pMICs

pMIC	Parameters[a]	SEP[b]	Q[c]	R[d]	PCs (%)
HN891	w_C, S_f, GlogK_{OW}, logP_s, SlogP_s, logK_{WIN}, SlogK_{WIN}, logP_{IA}, logP_X	0.467	0.912	0.967	3 (82)
	logP_{IA}, GlogK_{OW}, SlogP_s, H_f, A_{HB}, D_y, N_{NS}	0.209	0.982	0.993	3 (85)
SH5014	w_C, S_f, GlogK_{OW}, logP_s, SlogP_s, logK_{WIN}, SlogK_{WIN}, logP_{IA}, logP_X	0.391	0.942	0.975	2 (82)
	logP_{IA}, GlogK_{OW}, SlogP_s, H_f, A_{HB}, D_y, N_{NS}	0.316	0.962	0.982	3 (85)
SH7616	w_c, S_f, GlogK_{OW}, logP_s, SlogP_s, logK_{WIN}, SlogK_{WIN}, logP_{IA}, logP_X	0.792	0.645	0.886	2 (76)
	logP_{IA}, GlogK_{OW}, SlogK_{WIN}, N_{CH}, H_f, N_{NS}	0.461	0.851	0.930	3 (83)

[a]Transformation of lipophilicity descriptors (see Methods): GlogK_{OW} = exp[−(logK_{OW} − 1.1)^2]; SlogP_s = (logP_s)^2; SlogK_{WIN} = (logK_{WIN})^2.
[b]Standard error of validation.
[c]Correlation coefficient from validation.
[d]Correlation coefficient from prediction.

Table V. Validation of the PLS regression models for pMICs

pMIC	No.[a]	Experimental[b]	Predicted[b]	Percentage error[b]
HN891	**7**	4.073	3.841	5.7
	10	6.318	6.303	0.3
	14	5.055	5.323	5.3
SH5014	**7**	4.675	4.366	6.6
	10	6.637	6.642	0.1
	14	5.357	5.573	4.0
SH7616	**7**	5.277	5.693	7.9
	10	6.637	6.636	0.02
	15	4.652	5.085	9.3

pMIC	SEP[c]	Q^2[d]	R^2[e]	PCs(%)
HN891	0.246	0.975	0.994	3 (88)
SH5014	0.379	0.945	0.981	3 (88)
SH7616	0.539	0.797	0.924	3 (86)

[a]Compounds from the external validation set.
[b]Experimental and predicted pMICs and their difference expressed as percentage error.
[c]Standard error of validation.
[d]Correlation coefficient from validation.
[e]Correlation coefficient from prediction.

β-lactams is not more active. This efflux follows some other mechanism which probably excludes pump–drug recognition, as in the case of the other two strains. As there is no more molecular recognition where hydrogen bonding and electronic interactions would be very important, variables A_{HB} and D_y did not appear in the PLS model for strain SH7616. The PLS models with parameters of all types (Table IV) were further validated by excluding three samples from the data set, to be used for external validaton. One of the excluded samples was different for the strain SH7617, in order to cover a wide range of activities (Table V). The activities for the samples from the external validation set were predicted reasonably well, with relative errors less than 10% (Table V). The parameters for these PLS models were not significantly different from those using 16 samples in the training set (Table IV), although there were some visible differences in the SEP and Q values. Both training and the external validation sets can be seen in measured vs predicted plots accounting for the three bacterial strains HN891, SH5014 and SH7616 (Figure 7).

Obviously, the β-lactams efflux rate, expressed via pMICs, is determined by drugs' lipophilicity and electronic and HB properties. This conclusion is in accordance with the previous PCA and HCA results in this work. The reader can observe that AcrAB-TolC substrates **1**–**3** have a highly negative Y-component of the dipole moment D_y (Table III), and substrates **10**–**13** small negative or even positive D_y. The pMIC(HN891)–D_y relationship is illustrated by three examples in Figure 8. β-Lactams with hydrophobic substituents (like **1**) are characterized by well-defined continuous lipophilic and hydrophilic regions, what result in highly negative D_y. Introducing more polar substituents R and R_1, depending on a stable molecular conformation, D_y can vanish (atomic dipoles cancel each other, as in **14**), or even become positive (**8**). Van Bambeke *et al.* [1] discussed the amphiphilic character of antibiotics including β-lactams, pointing out the existence of well-defined lipophilic and hydrophilic regions in molecules, which agrees with our results. A more general qualitative scheme for

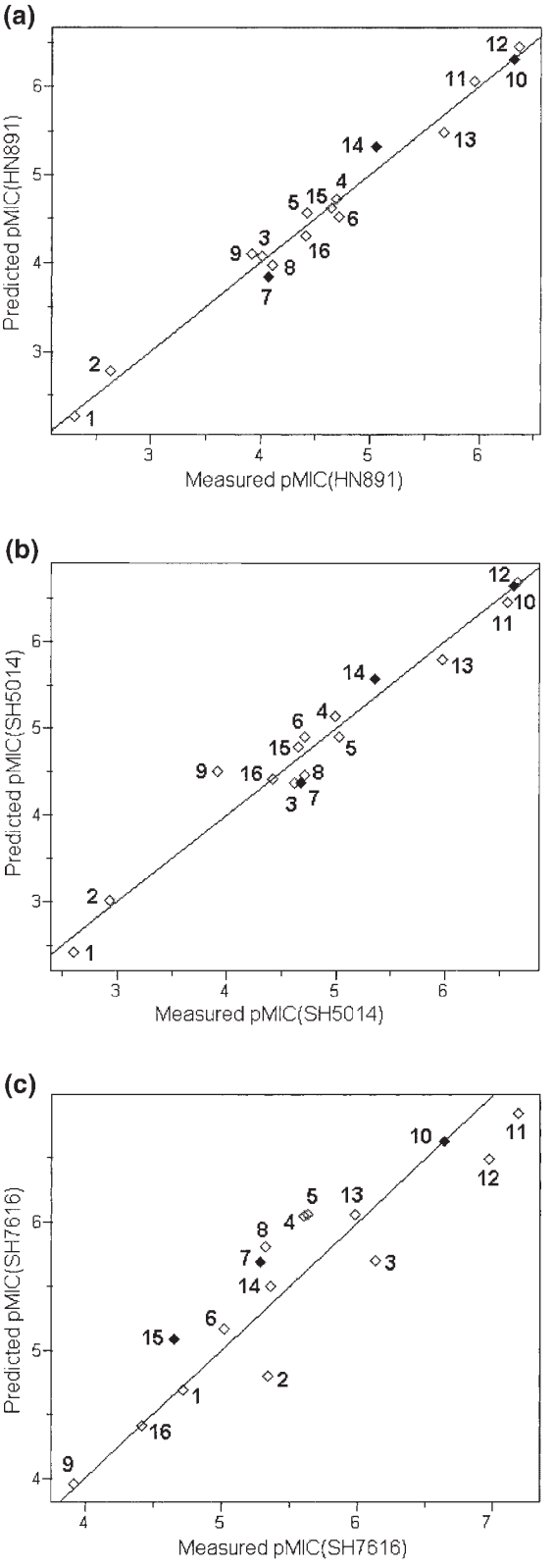

Figure 7. Experimental vs predicted pMICs: (a) HN891, (b) SH5014, (c) SH7616. Solid squares account for the samples from the external validation sets.

AcrAB-TolC–drug recognition in terms of steric and electronic fits can be proposed (Figure 9). Stereoelectronic properties of R and R_1 can enhance or weaken the drug–receptor fit, by influencing molecular conformation and electronic properties such as for example, the most important dipole

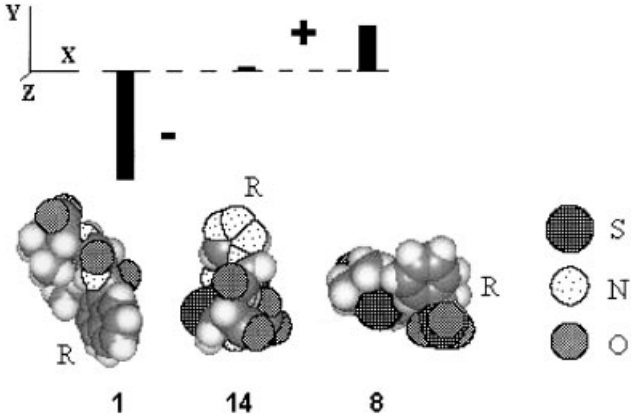

Figure 8. β-Lactam molecules with the most negative (**1**), close to zero (**14**) and the most positive (**8**) dipole moment component along the *y*-axis (D_y). The molecules lie in the *xy*-plane. Heteroatoms are rastered differently from C and H. Substituent R is also labeled.

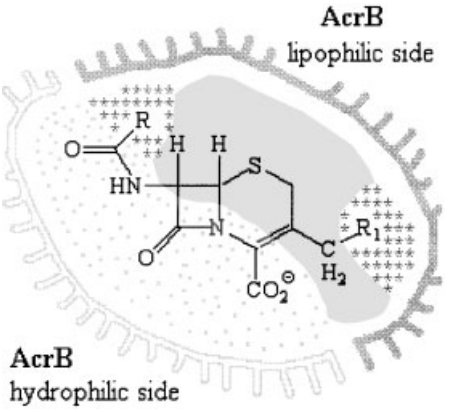

Figure 9. A simplified two-dimensional representation of stereoelectronic AcrAB-TolC bacterial pump–chephalosporin fit. The picture is most appropriate for description of AcrB vestibule–β-lactam fit. The predominantly polar (dotted area) and hydrophobic (gray area) drug regions align along the compatible domains of a pump's recognition site (receptor). The R and R_1 substituents (starred areas) can be polar, amphiphilic or hydrophobic and, consequently, may influence the drug–receptor fit.

moment Y-component D_y (see Figure 8). Summarizing, PLS models from Table IV reveal that basically three factors are important for β-lactam efflux by the bacterial AcrAB-TolC pumps: lipophilicity and electronic and HB drug properties, related mostly to their conformation and charge distribution.

3.4. Biochemical background of the QSAR results

The AcrAB-TolC efflux pump is presented in Figure 10 using atomic coordinates for AcrB [14] and TolC [18] proteins from the Protein Data Bank [55] (PDB codes 1IWG and 1EK9, respectively), and supposing a possible three-dimensional arrangement of AcrA proteins for which a three-dimensional structure is not yet known [21]. All the three components are from *Escherichia coli*, and are trimers. AcrB monomer units are colored differently. TolC protein is docked to the AcrB, and both are additionally bound to the AcrA monomers.

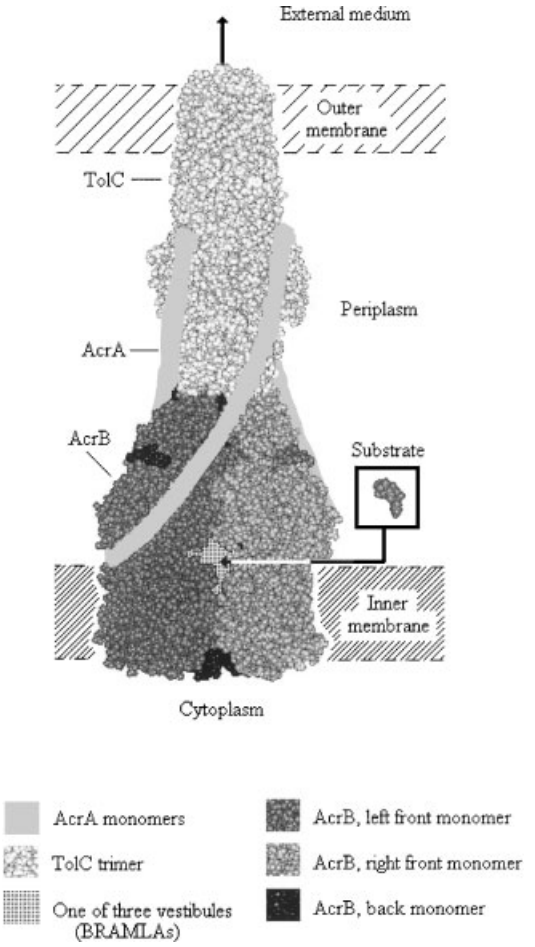

Figure 10. AcrAB-TolC efflux pump. TolC is manually docked to AcrB. Only one vestibule is visible in this orientation, while the other two are placed at the back of the AcrB trimer, at the joint lines of the monomers. Arrows show the substrate efflux pathway starting from the periplasm.

This complex represents the functional AcrAB-TolC bacterial efflux pump, placed in the bacterial membrane. The AcrB pump is in contact with cytoplasm, but is placed in the inner membrane and periplasm. Most TolC lies in periplasm and partially in the outer membrane, while AcrA is completely in periplasm. Since AcrB is a trimer, there are three holes along the lines of contact of the monomeric units called vesitubles (only one is presented in Figure 10).

It is generally accepted that the mechanism of substrate efflux from periplasm includes the following phases [14,16–18]: (1) the substrate enters the closest vestibule; (2) passing the vestibule's channel, the substrate comes to the central hole of the AcrB trimer; (3) after that, the substrate leaves AcrB and enters the TolC channel, through its narrow channel called pore; and (4) the substrate travels through the TolC channel until it leaves the bacteria cell and enters the external medium. The vestibules, the central cavity, the pore and other parts of the TolC channel represent recognition sites of the AcrAB-TolC pump, and can differentiate hydrophobic/amphiphilic and hydrophilic substrates [14–18]. In detailed analysis of special properties of vestibules and the pore, the authors of this paper have shown recently (Kiralj *et al.*, submitted) the following interesting facts on AcrB–substrate molecular recognition: (1) the vestibule entrance

possesses a characteristic shape that can be recognized even in small pictures (see Figure 10) and can be called BRAMLA (BRAzil Map-Like Area); (2) BRAMLA and the pore recognition site interact with substrate molecules and affect their efflux rate (biological activity of the pump); (3) this interaction includes molecular size, shape, electronic, hydrogen bonding and hydrophobicity complementarity between a substrate molecule and the receptor (a recognition site in AcrB); and (4) hydrophilic drugs tend to establish several hydrogen bonds or polar–polar interactions with receptors, and these interactions are much stronger than other individual interactions between hydrophobic groups. Figure 9 describes drug–BRAMLA interaction.

PCA and HCA results in this work showed that lipophilicity of β-lactams is very important for their efflux. PLS models confirmed this, but also pointed out that electronic and hydrogen bonding properties of drugs should not be ignored. Furthermore, it has been shown recently by QSAR studies [56,57] that lipophilicity parameters can be treated as a linear combination of suitable steric and electronic molecular properties, which may aid in mechanistic interpretation of pump–drug interactions. On the other hand, β-lactam antibiotic molecules interact with polar solvents, so charge transfer [58] and changes of β-lactam molecular dipole moment and atomic charges [59] occur. These findings as well as the above results, explain why electronic and HB parameters are important in PLS models.

4. CONCLUSIONS

AcrAB-TolC is the most important efflux pump system of Gram-negative bacteria, mainly responsible for bacterial resistance to lipophilic and amphiphilic drugs, including β-lactams. This was the reason for performing PCA–HCA study on biological activities (efflux rates) of three different strains of *S. thypimurium* with respect to β-lactams, and on lipophilicity parameters calculated by different methods. In the final stage, a QSAR study was performed based on lipophilicity and electronic and HB molecular descriptors. The analyses demonstrate that: (1) biological activities (pMICs) strongly depend both on properties of bacterial strains and drug molecules. β-Lactams were classified as good, moderately good to poor AcrAB-TolC substrates; (2) among the most important β-lactam molecular properties quantitatively related to pMICs are lipophilicity and electronic and hydrogen bonding properties; and (3) lipophilicity parameters calculated in different ways do not necessarily present the same information about drugs, and cannot produce parsimonius regression models for MICs originated by active AcrAB-TolC pumps. Some lipophilicity parameters were non-linearly related to the pMICs, mainly due to the existence of β-lactams with nitrogen- and sulfur-rich substituents. Penicillins and cephalosporins stereoelectronic molecular properties, especially the Y-component of the molecular dipole moment and hydrogen binding properties, reflect the β-lactam classification obtained from PCA and HCA. From this, it is clear that penicillins and cephalosporins which can be characterized as hydrophilic, with good hydrogen bonding properties and able to establish polar-polar interactions with bacterial pump receptors, are bad pump substrates, and so potentially good drugs. In searching for better β-lactam antibiotics, drug design should start with such drugs.

Acknowledgments

This work was supported by the State of São Paulo Funding Agency (FAPESP).

REFERENCES

1. van Bambeke F, Balzi F, Tulkens PM. Antibiotics efflux pumps. *Biochem. Pharmac.* 2002; **60**: 457–470.
2. Nikaido H. Multidrug efflux pumps of gram-negative bacteria. *J. Bacteriol.* 1996; **178**: 5853–5859.
3. Paulsen IT, Brown MH, Skurray RA. Proton-dependent multidrug efflux system. *Microbiol. Rev.* 1996; **60**: 575–608.
4. Sulavik M, Houseweart C, Cramer C, Jiwani N, Murgolo N, Greene J, Di Domenico B, Shaw KJ, Millear GH, Hare R, Shimer G. Antibiotic susceptibility of *Escherichia coli* strains lacking multidrug efflux pump genes. *Antimicrob. Agents Chemother.* 2001; **45**: 1126–1136.
5. Giraud E, Cloekaert A, Kerboeuf D, Chaslus-Dancla E. Evidence for active efflux as the primary mechanism of resistance to ciprofloxacin in *Salmonella enterica* serovar typhimurium. *Antimicrob. Agents Chemother.* 2000; **44**: 1223–1228.
6. Okusu H, Ma D, Nikaido H. AcrAB efflux pump plays a major role in the antibiotic resistance phenotype of *Escherichia coli* multiple-antibiotic-resistance (mar) mutants. *J. Bacteriol.* 1996; **178**: 306–308.
7. Zgurskaya HI, Nikaido H. Bypassing the periplasm: reconstitution of the AcrAB multidrug efflux pump of *Escherichia coli*. *Proc. Natl Acad. Sci. USA* 1999; **96**: 7190–7195.
8. Helling RB, Janes BK, Kimball H, Tran T, Budesmann M, Check P, Phelan D, Miller C. Toxic waste disposal in *Escherichia coli*. *J. Bacteriol.* 2002; **184**: 3699–3703.
9. Nikaido H, Basina M, Nguyen V, Rosenberg EY. Multidrug efflux pump AcrB of *Salmonella typhimurium* exctretes only those β-lactam antibiotics containing lipophilic side chains. *J. Bacteriol.* 1998; **180**: 4686–4692.
10. Sanchez L, Pan W, Vinas M, Nikaido H. The acrAB homolog of *Hamemophilus influenzae* codes for a functional multidrug efflux pump. *J. Bacteriol.* 1997; **179**: 6855–6857.
11. Aono R, Tsukagoshi N, Yamamoto M. Involvement of outer membrane protein TolC, a possible member of the mar–sox regulon, in maintenance and improvement of organic solvent tolerance of *Escherichia coli* K12. *J. Bacteriol.* 1998; **180**: 938–944.
12. Thanassi DG, Cheng LW, Nikaido H. Active efflux of bile salts by *Escherichia coli*. *J. Bacteriol.* 1997; **179**: 2512–2518.
13. White DG, Goldman JD, Demple B, Levy SB. Role of the acrAB locus in organic solvent tolerance mediated by expression of marA, soxS, or robA in *Escherichia coli*. *J. Bacteriol.* 1997; **179**: 6122–6126.
14. Murakami S, Nakashima R, Yamashita E, Yamaguchi A. 2002. Crystal structure of bacterial multidrug efflux transporter AcrB. *Nature* 2002; **419**: 587–593.
15. Pos KM, Diederichs K. Purification, crystallization and preliminary diffraction studies of AcrB, an inner-membrane multi-drug efflux protein. *Acta Cryst.* 2002; **D58**: 1865–1867.
16. Yu EW, McDermott G, Zgurskaya HI, Nikaido H, Koshland DE Jr. Structural basis of multiple drug-binding capacity of the AcrB multidrug efflux pump. *Science* 2003; **300**: 976–980.
17. Cristopher AE, Nikaido H. 3D structure of AcrB: the archetypal multidrug efflux transporter of *Escherichia coli*

likely captures substrates from periplasm. *Drug Resist. Updates* 2003; **6**: 9–13.
18. Koronakis V, Sharff A, Koronakis E, Luisi B, Hughes C. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature* 2000; **405**: 914–919.
19. Thanabalu T, Koronakis E, Hughes C, Koronakis V. Substrates-induced assembly of a contiguous channel from protein export from *E. coli*: reversible bridging for an inner-membrane translocase to an outer membrane exit pore. *EMBO J.* 1998; **17**: 6487–6496.
20. Dinh T, Paulsen IT, Saier MH Jr. A family of extracytoplasmic proteins that allow transport of large molecules across the outer membranes of gram-negative bacteria. *J. Bacteriol.* 1994; **176**: 3825–3831.
21. Avila-Sakar AJ, Misaghi S, Wilson-Kubalek EM, Downing KH, Zgurskaya H, Nikaido H, Nogales E. Lipid-layer crystallization and preliminary three-dimensional structural analysis of AcrA, the periplasmic component of a bacterial multidrug efflux pump. *J. Struct. Biol.* 2001; **136**: 81–88.
22. Ferreira MMC. Multivariate QSAR. *J. Braz. Chem. Soc.* 2002; **13**: 742–753.
23. Alves CN, Pinheiro JC, Camargo AJ, Ferreira MMC, da Silva ABF. A structure activity relationship study of HEPT-analog compounds with anti-HIV activity. *Theochem. J. Mol. Struct.* 2000; **530**: 39–47.
24. Bruni AT, Ferreira MMC. Omeprazole and analogue compounds: a QSAR study of activity against *Helicobacter pylori* using theoretical descriptors. *J. Chemometrics* 2002; **16**: 510–520.
25. Kiralj R, Ferreira MMC. A priori molecular descriptors in QSAR: a case of HIV-1 protease inhibitors. I. The chemometric approach. *J. Mol. Graph. Mod.* 2003; **21**: 435–448.
26. Kiralj R, Ferreira MMC. *A priori* molecular descriptors in QSAR: a case of HIV-1 protease inhibitors. II. Molecular graphics and modeling. *J. Mol. Graph. Mod.* 2003; **21**: 499–515.
27. Kiralj R, Ferreira MMC. QSAR of progestogens: use of *a priori* and computed molecular descriptors and molecular graphics. *QSAR Combin. Sci.* 2003; **22**: 430–448.
28. Kiralj R, Ferreira MMC. Molecular graphics–structural and molecular graphics descriptors in a QSAR study of 17-α-acetoxyprogesterones. *J. Braz. Chem. Soc.* 2003; **14**: 20–26.
29. Pinheiro JC, Kiralj R, Ferreira MMC, Romero OAS. Artemisinin derivatives with antimalarial activity against *Plasmodium falciparum* designed with the aid of quantum chemical and partial least squares methods. *QSAR Combin. Sci.*, 2003; **22**: 830–842.
30. Subramanian S, Ferreira MMC, Trsic M. A Structure–activity relationship study of lapachol and some derivatives of 1,4-naphthoquinone against carcinosarcoma Walker 256. *Struct. Chem.* 1998; **9**: 47–57.
31. Pinheiro JC, Ferreira MMC, Romero OAS. Antimalarial activity of dihydroartemisnin derivatives against *P. falciparum* resistant to mefloquine: a quantum chemical and multivariate study. *Theochem. J. Mol. Struct.* 2001; **572**: 35–44.
32. Vendrame R, Ferreira MMC, Collins CH, Takahata Y. Structure–activity relationships (SAR) of contraceptive progestogens studied with four different methods using calculated physicochemical parameters. *J. Mol. Graph. Mod.* 2002; **20**: 345–358.
33. Geladi P, Kowalski BR. Partial least squares regression–a tutorial. *Anal. Chim. Acta* 1986; **185**: 1–17.
34. Martens H, Naes T. *Multivariate Calibration* (2nd edn). Wiley: New York, 1989.
35. Beebe, KR, Pell R, Seasholtz MB. *Chemometrics: a Practical Guide*. Wiley: New York, 1998.
36. Ferreira MMC, Antunes AM, Melo MS, Volpe PLO. Chemometrics I: multivariate calibration, a tutorial. *Quím. Nova* 1999; **22**: 724–731.
37. *PC Spartan Pro 1.0.5*. Wavefunction: Irvine, CA, 2001.
38. A. Dobashi. *3D Pharmaceutical Structure Database*. Department of Structural Organic Chemistry, Tokyo University of Pharmacy and Life Science: Tokyo; www.ps.toyaku.ac.jp/dobashi/3dpsd/index.htm (accessed 1 September 2003).
39. *The Cambridge Structural Database* (November 2002 Release). Cambridge Crystallographic Data Centre, University of Cambridge: Cambridge.
40. Stewart JJP. Optimization of parameters for semiempirical methods. 1. Method. *J. Comp. Chem.* 1989; **10**: 209–220.
41. *Interactive Analysis logP and logW Predictors Website*. Interactive Analysis: Bedford, MA; www.logp.com (accessed 1 September 2003).
42. *ALOGPS 2.1*. Virtual Computational Chemistry Laboratory; http://146.107.217.178/lab/alogps/ (accessed 1 September 2003).
43. Koltun WL. Precision space-filling atomic models. *Biopolymers* 1965; **3**: 665–679.
44. Lobanov VS. *MOPAC 6.0 for Microsoft Windows*. University of Florida, Miami, FL, 1996.
45. *Matlab 6.1.0.450 release 12.1*. MathWorks: Natick, MA, 2001.
46. *Pirouette 3.01*. Infometrix: Woodinville, WA, 2001.
47. Abraham DJ, Kellog GE. 2000. Hydrophobic fields. In *3D QSAR in Drug Design: Theory, Methods and Applications*, Kubinyi H (ed.). Kluwer/Escom: Dordrecht, 2000; 506–522.
48. Lien EJ. *SAR: Side Effects and Drug Design*. Marcel Dekker: New York, 1987; 41–162.
49. Hansch C, Steward AR, Anderson SM, Bentley D. The parabolic dependence of drug action upon lipophilicity character as revealed by a study of hypnotics. *J. Med. Chem.* 1968; **11**: 1–11.
50. Hyde RM. Relationships between the biological and physiological properties of series of compounds. *J. Med. Chem.* 1975; **18**: 231–233.
51. Cole JC, Lommerse JPM, Rowlands RS, Taylor R, Allen FH. 1998. Use of Cambridge Structural Database to Study Non-covalent interactions: towards a knowledge base of intermolecular interactions. In *Structure-based Drug Design: Experimental and Computational Approaches*, Codding PW (ed.). NATO ASI Series E: Applied Sciences, Vol. 352. Kluwer: Dordrecht, 1998; 113–124.
52. Glusker JP, Lewis M, Rossi M. *Crystal Structure Analysis for Chemists and Biologists*, Chap. 15 and 17. VCH: New York, 1994.
53. Nishio M, Hirota M, Umezawa Y. *The CH/π Interaction*. Wiley-VCH: New York, 1998.
54. Scheiner S, Kar T, Gu Y. Strength of the $C^{\alpha}H\cdots O$ hydrogen bond of amino acid residues. *J. Biol. Chem.* 2001; **276**: 9832–9837.
55. *Protein Data Bank*. Research Collaboratory for Structural Bioinformatics; www.rcsb.org/index.html (accessed 5 September 2003).
56. Kaiser KLE, Niculescu SP. Using probabilistic neural networks to model the toxicity of chemicals to the fathead minnow (*Pimephales promelas*): a study based on 865 compounds. *Chemosphere* 1999; **38**: 3237–3245.
57. van de Waterbeemd H, El Tayar N, Carrupt PA, Testa B. Pattern-recognition study of QSAR substituent descriptors. *J. Comput-Aid. Molec. Des.* 1989; **3**: 111–132.
58. Dutta M, Dutta NN, Bhattacharyya KG. Adsorptive interaction of certain β-lactam antibiotics in aqueous solution: interpretation by frontier-orbital theory. *J. Chem. Eng. Jpn.* 2000; **33**: 303–307.
59. Leon S, Alemán C, García-Alvarez M, Muñoz-Guerra S. Theoretical study of the conformational and electrostatic properties of C4-monosubstituted 2-azetidinones. *Struct. Chem.* 1997; **8**: 39–47.

