Computational analysis of structure-based interactions and ligand properties can predict efflux effects on antibiotics

ARTICLE *in* EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY · MARCH 2012

Impact Factor: 3.45 · DOI: 10.1016/j.ejmech.2012.03.008 · Source: PubMed

3 AUTHORS, INCLUDING:

Aurijit Sarkar
Virginia Commonwealth University
17 PUBLICATIONS **127** CITATIONS
SEE PROFILE

Glen Kellogg
Virginia Commonwealth University
145 PUBLICATIONS **3,471** CITATIONS
SEE PROFILE

Published in final edited form as:
Eur J Med Chem. 2012 June ; 52: 98–110. doi:10.1016/j.ejmech.2012.03.008.

Computational Analysis of structure-based interactions and ligand properties can predict efflux effects on antibiotics

Aurijit Sarkar, **Kelcey C. Anderson**, and **Glen E. Kellogg***
Department of Medicinal Chemistry & Institute for Structural Biology and Drug Discovery, Virginia Commonwealth University, Richmond VA 23298-0540, USA

Abstract

AcrA-AcrB-TolC efflux pumps extrude drugs of multiple classes from bacterial cells and are a leading cause for antimicrobial resistance. Thus, they are of paramount interest to those engaged in antibiotic discovery. Accurate prediction of antibiotic efflux has been elusive, despite several studies aimed at this purpose. Minimum inhibitory concentration (MIC) ratios of 32 β-lactam antibiotics were collected from literature. 3-Dimensional Quantitative Structure Activity Relationship on the β-lactam antibiotic structures revealed seemingly predictive models (q^2 = 0.53), but the lack of a general superposition rule does not allow its use on antibiotics that lack the β-lactam moiety. Since MIC ratios must depend on interactions of antibiotics with lipid membranes and transport proteins during influx, capture and extrusion of antibiotics from the bacterial cell, descriptors representing these factors were calculated and used in building mathematical models that quantitatively classify antibiotics as having high/low efflux (>93% accuracy). Our models provide preliminary evidence that it is possible to predict the effects of antibiotic efflux if the passage of antibiotics into, and out of, bacterial cells is taken into account – something descriptor and field-based QSAR models cannot do. While the paucity of data in the public domain remains the limiting factor in such studies, these models show significant improvements in predictions over simple LogP-based regression models and should pave the path towards further work in this field. This method should also be extensible to other pharmacologically and biologically relevant transport proteins.

Keywords

efflux pumps; β-lactam antibiotics; 3D QSAR, HINT; hydropathic interactions; docking and scoring

1. Introduction

Antimicrobial drugs have been crucial tools of healthcare for decades due to their effectiveness in control of bacterial infections. However, soon after their discovery, it was

*Corresponding Author, glen.kellogg@vcu.edu, phone: +1 804 828-6452, fax: +1 804 827-3664.

Description of Supporting Information: Descriptors for all antibiotics included in this study are provided, along with graphical representations of both linear and quadratic relationships between the various descriptors and logMIC. A preliminary investigation into modeling building with quadratic treatment of the descriptors is also described. Lastly, the surface property map for TolC is shown.

realized that some pathogens rapidly developed resistance to antibiotics [1,2]. Initially, this problem was overcome by discovery of new classes of antibiotics such as aminoglycosides, macrolides and glycopeptides, but it became increasingly clear that bacteria had an impressive array of defensive mechanisms that conferred resistance to many modes of attack [1]. Organisms that cause pneumonia and cutaneous infections, *Streptococcus pneumonia*, *Streptococcus pyogenes* and *Staphylococci*, are now resistant to almost all of the older, first generation, β-lactam antibiotics [2] like penicillin, which were discovered through screening of mold samples [3,4,5]. Members of the *Enterobacteriaceae* and *Pseudomonas* families, which cause diarrhea, urinary infection and sepsis, are also resistant [2]. The development of bacterial resistance to antibiotics has mostly been attributed to their excessive use in the clinic as well as at home [1,2]. The therapeutic usefulness of several drug classes has been prolonged through clever drug design, e.g., semisynthetic modifications of β-lactam antibiotics have given us second- and third-generation agents. However, for the sake of sustainability and clarity of future endeavors in antibiotic design, it is essential to understand the routes of drug resistance.

1.1. Efflux pumps

One primary mechanism of antibiotic resistance is extrusion of the foreign chemical, which is termed efflux. In 1980, it was reported that tetracycline could be actively effluxed from the bacterial cell [6]. Since then, many efflux-related mechanisms have been discovered. Efflux pumps are transporters involved in extrusion of toxic substances from cells, thereby limiting the detrimental effects of these substances [7]. They may be substrate-specific and responsible for transporting biological compounds such as bile salts, or may be promiscuous and transport structurally diverse compounds such as various classes of antibiotic drugs [8]. Overexpression of these structurally complex and versatile proteins may thus lead to antibiotic resistance. While efflux pumps are present in both Gram-positive and Gram-negative bacteria and also in eukaryotes, antibiotic resistance due to efflux is a bigger problem in Gram-negative bacteria than in Gram-positive bacteria [9]. This is due to the presence of an outer membrane in Gram-negative bacteria that demonstrates comparatively lower permeability and complements the efflux activity of these pumps.

Several such pump systems have been described: *Campylobacter jejuni* (CmeABC) [10,11], *Escherichia coli* (AcrAB-TolC, AcrEF-TolC, EmrB, EmrD) [12], *Pseudomonas aeruginosa* (MexAB-OprM, MexCD-OprJ, MexEF-OprN and MexXY-OprM) [12], *Streptococcus pneumonia* (PmrA) [13], *Salmonella typhimurium* (AcrAB) [14] and *Staphylococcus aureus* (NorA) [15]. These pumps basically fall into five major families, including the MF (major facilitator), MATE (multidrug and toxic efflux), SMR (small multi-drug resistance), ABC (ATP-binding cassette) and RND (resistance-nodulation-division) families [16]. It has been shown that co-expression of multiple types of efflux pumps can cause an additive or even multiplicative effect on drug resistance [17].

AcrAB was first described as an efflux system in 1995 [18]. AcrB (Figure 1A) is responsible for efflux of bile salts, thus protecting enteric *E. coli* from the detrimental effects of these powerful detergents [19]. As is typical with other members of the RND-type efflux protein systems, AcrAB is also a proton antiporter. AcrA and AcrB homologues in *Haemophilius influenzae*, HI0894 and HI0895 respectively, are also responsible for drug efflux [20]. TolC is a trimeric 12-stranded α/β barrel (Figure 1B) forming a tunnel and thus providing the exit duct for substrates. The importance of AcrA-AcrB-TolC in multidrug resistance has been described in several publications [21–26], where knock-out, knock-in and mutation studies were used in an attempt to describe the extent to which the AcrA-AcrB-TolC transporter is responsible for expulsion of structurally diverse antibiotics from bacterial cells.

1.2. MIC ratios

Efflux for substrates has generally been measured and reported as MIC (minimum inhibitory concentration) ratios, i.e., ratio of MIC in the presence of the efflux pump to MIC in the absence of efflux pump, e.g., through knockout [21–27]. MIC ratios represent two aspects of antibiotic movement between the bacterial cell and the extracellular medium – influx into the cell, and efflux out of the cell. While the influx of antibiotic into the bacterial cell is poorly understood, it is likely to be partly regulated by Lipinski's rule-of-five [28,29]. An additional factor is the facilitation of antibiotic influx by large Outer Membrane Proteins (OMPs) called porins. It has been shown that the right balance of affinity and repulsion is required for passage of antibiotics through OmpC porins, whose expression is favored *in vivo* [30]. Since the sizes of the antibiotic molecules are comparable to the diameter of porin channels, diffusion of the former through the latter is likely to be affected by not only physicochemical interactions between the two, but also by the size of antibiotic under consideration. Similarly, efflux of antibiotics must be dependent on the recognition of antibiotic by the efflux pump along with its interactions with various regions of its structure.

The balance between penetration of antibiotic into the cell, and the ability of efflux pumps to evacuate it from the same, determines its periplasmic concentration. Since an antibiotic must reach sufficient concentration within the periplasm in order to penetrate further into the bacterium and reach its target, any factors that might alter the ability of an antibiotic to accumulate within this region is of clinical significance. Hence, while MIC ratios are unable to differentiate between rates of influx and efflux, they remain more than a crude metric for efflux efficiency. They represent (with due consideration of numerous other factors that affect bioavailability) the degree of increased dosing required in order to achieve the same clinical effect as would have been achieved in the absence of the efflux pumps.

1.3. Kinetic measurements of efflux

In very recent work by Lim, Nagano and Nikaido [31,32], kinetic parameters, $K_{0.5}$ and V^{max}, for AcrA-AcrB-TolC efflux of a dozen β-lactam antibiotics were reported. In order to negate the effect of influx rate on these kinetic experiments, porins that are large enough to permit facile influx of all substrates into the cell were incorporated into the bacterial cell. The kinetics of efflux were then measured on these carefully manipulated systems. Perhaps the most surprising result was that $K_{0.5}$ and V^{max} for cloxacillin and ampicillin are relatively similar, whereas their MIC ratios are vastly different: 2 to 4 for ampicillin vs. 128 to 256 for cloxacillin [23–27]. This implies that differences in their influx rates (or intracellular enzymatic hydrolysis by β-lactamase) must be largely responsible for their differences in MIC ratios [31,32].

With this new understanding of efflux kinetics, the long-held key conclusion from MIC ratio measurements that efflux pumps prefer to “extrude” hydrophobic ligands [24] can be reexamined. Instead of clarity, however, this leads a pair of conundrums: 1) if the AcrA-AcrB-TolC pump is similarly efficient at extruding all compounds, as the kinetic data seems to indicate [32], then are the observed higher MIC ratios for more hydrophobic substrates (counterintuitively) suggesting that more hydrophobic substrates are *less* able to transport across the membrane into the cell? 2) As neither influx nor β-lactamase activity should be significantly affected by the presence or absence of the pump, why do the MIC ratios range over such a broad spectrum as compared to the pump kinetic parameters? While it is clearly important to understand the thermodynamics and kinetics of cell-antibacterial substrate interactions as a complete package, and the recent results on efflux kinetics by AcrA-AcrB-TolC are certainly a crucial piece of the puzzle, other facets of the phenomenon demand continuing investigation.

In light of the above, along with the clinical significance of MIC ratios (*vide supra*) and the availability of data from literature in the public domain (with all due respect to the emerging quantitation of efflux kinetics), this work will focus on MIC ratios.

1.4. Computational modeling of efflux

Interestingly, despite the tremendous importance of efflux as an ancillary consideration in drug discovery, very few computational studies designed to predict the effect on proteins other than P-glycoprotein [33–35] have been reported. However, two-dimensional quantitative structure-activity relationship (QSAR) studies performed previously [36,37] yielded what appeared on the surface to be remarkable regression equations for efflux of β-lactam substrates by the AcrA-AcrB-TolC pump from *Salmonella typhimurium* (16 compounds, target: minimal inhibitory concentration for three strains, reported r^2 from cross validation (q^2) > 0.9). No surprising conclusions were made in this study, i.e., molecules showing more hydrophilic character, including hydrogen bonding capability, were likely to be poor efflux substrates, and that efflux correlated with properties like LogP (for partitioning the drug between 1-octanol and water), the y-axis component of electrostatic dipole moment, the surface area of hydrophobic carbons, the number fractions of carbons and heteroatoms and the number of charged groups and the number of nitrogen and sulfur atoms, all of which would supposedly influence interactions between pump and substrate [36]. However, the number of descriptors in the equations (up to 9) suggests serious overfitting of such a small data set and the inclusion of multiple methods of LogP prediction within the same QSAR equation is also a concern: while LogPs calculated by different methods do not necessarily encode exactly the same information about a molecule, they *must* be largely correlated, and some of the other descriptors, e.g., surface area of hydrophobic carbons, also likely correlate with LogP. Most importantly, it does not appear that this type of model has been embraced by potential users of efflux prediction in drug design, possibly because of its poor chemical and physical interpretability.

1.5. Aims of this work

Factors that affect influx and efflux of an antibiotic into the bacterial cell govern its effectiveness. Since the clinically relevant MIC ratios encode the sum total effects of antibiotic retention within the cell, it would be beneficial if we could predict MIC ratios based on descriptors governing influx/efflux of these molecules. If we ignore the diffusion of antibiotic within the periplasmic matrix and probability-based recognition of the same by efflux pumps, the value of MIC ratio will depend on the mode of entry of antibiotic into the bacterial cell, along with its interactions with the efflux pump structure.

We are proposing here that characteristics of the antibiotic such as size and hydrophobicity, along with its scored interactions with efflux pumps, can be used to predict its MIC ratio. In this manuscript, we show preliminary evidence that this can be achieved in practice. We believed that emerging structural information for pump molecules [38,39], combined with modeling tools that effectively characterize hydrophobic interactions and related effects, could illuminate the process of efflux. Our key technology is the HINT model [40–42], which is an empirical computational tool based on the free energy of solvent transfer between two phases, 1-octanol and water, representing hydrophobic and polar biological environments, respectively. HINT has been used to evaluate ligand binding [43,44], protein-protein associations [45–47] and other phenomena involving biological molecules [48,49], and has been generally successful in quantitating the free energies of these interactions. In addition, qualitative representations of biological processes involving molecular associations have been developed within the HINT paradigm; e.g., we recently developed molecular models validating previously proposed dual site mechanisms for inhibition of paramyxovirus hemagglutinin-neuraminidase [50,51].

2. Results and Discussion

The intrinsic function of efflux pumps is to expel extraneously acquired molecules that could harm the cell. These are promiscuous proteins that are largely responsible for antibiotic resistance in Gram-negative bacteria. In contrast to normal receptors and enzymes that bind or else catalyze reactions involving small molecules at a specific site, the function of these proteins to *transport* these small molecules – usually over is fairly large distances. Perhaps due to this difference in function, and the multiple steps involved in what is clearly a dynamic rather than static process, their activity has been resistant to computational chemistry/biology attempts at prediction. In fact, for a long time, the only useful known trait of the AcrA-AcrB-TolC efflux pump was that it transports hydrophobic molecules more easily [24]. However, recently available crystallographic data for the AcrB and TolC components of this efflux pump has enabled a more systematic and structural evaluation of the efflux mechanism [38].

In obtaining and curating a high quality and relevant data set for our analysis, we restricted this study to 32 β-lactam compounds (see Scheme 1) and 12 other compounds from several (non-β-lactam) families (see Scheme 2), whose efflux data (MIC ratios) were available in published reports [17,23–25,27]. Although similar data are available for the aminoglycoside and macrolide classes of antibiotics, they are not included in the curated set; the former because they are hydrophilic and therefore not effluxed as readily by the AcrAB pump – the AcrD pump is apparently more responsible for their efflux than the AcrB pump [52,53] – and the latter because of their large size relative to that of the AcrB entrance. The macrolides have molecular widths of about 13.4 Å, while the range for compounds in this study is 6.0 to 11.0 Å (average 8.4 Å). However, we cannot discount the possibility that these antibiotics might enter TolC via a different route [54], thus still involving parts of the AcrA-AcrB-TolC pump (and concomitantly being affected by the pump knockout mutants). We also recognize that a much larger, but proprietary, set of data is very likely available within pharmaceutical companies, but wanted the entire data set to be available from the primary literature for this initial work.

The experimental measure of efflux used in this study, Log_2(MIC ratio) termed logMIC, is the logarithm base 2 of the ratio of MIC (minimum inhibitory concentration) for a cell with an intact AcrA-AcrB-TolC pump normalized by the MIC for that cell with the pump knocked out. Using this ratio in lieu of MIC itself has benefits as it more clearly represents the change in effectiveness of an antibiotic. In contrast, non-normalized MIC is a poorer target metric as it is dependent on a number of factors such as concentration of cells per unit volume of culture, and thus more laboratory and procedure-dependent than its normalized analog. Although MICs and MIC ratios are not strongly correlated with Nikaido's efflux rates, MIC ratios represent a collection of effects – not limited to, but including, efflux by the pump – that do reflect how these effects alter antibiotic potency *in toto*. The prediction of MIC/MIC ratio is thus of major importance as it provides insight into the overall effectiveness of an antibiotic.

2.1. 3D-QSAR

It seemed possible that binding in the AcrB pocket would affect the extrusion of substrates more than binding elsewhere in the pump because it has been shown that this protein must undergo a conformational change in order to pass substrates into TolC [38]. One way to test this is to perform 3D-QSAR analyses, where the interactions between substrates and a hypothetical, but undefined, receptor are simulated by the molecular fields of the substrates. This approach was applied to the β-lactam data set after conformationally aligning these molecules to simulate their putative binding modes within a binding site presumed to be AcrB. If such an analysis provided predictive models, then we at least partially address our

goal of predicting high or low MIC ratios, and by inference efflux. We performed this 3D-QSAR study of the β-lactam data set with molecules in both their charged and neutral forms using the Comparative Molecular Field Analysis (CoMFA) method of Cramer [55]. The results seem promising for neutral compounds, with a cross-validated r^2 (q^2) = 0.53 (4 components) on the training set of 24 compounds and a (predictive) r^2 = 0.80 for the test set of 8 compounds, as illustrated for a typical run in Figure 2. This model used the CoMFA Standard Steric field, H-bond Steric field and H-bond Electrostatics field, with relative contributions of 27%, 38% and 35%, respectively. CoMFA models for β-lactam substrates in their charged state could not be validated and showed inadequate predictive nature (data not shown).

However, 3D-QSAR experiments are dependent upon the alignment of a *common substructure*, in this case the β-lactam moiety. Thus, this model formalism would not be readily utilizable for the extended dataset including the non-β-lactams. Furthermore, 3D QSAR experiments are based on the crucial assumption that substrates bind to a single, largely static, site in the pump, whether AcrB, as we hypothesized, or elsewhere. In fact, AcrB changes conformation between between its binding and extrusion states [38], which suggests that there is a dynamic change in its interactions with ligands as they are processed. Also of interest is that the binding pocket of AcrB is lined with a number of hydrophobic phenylalanine residues [38], indicating a preference for hydrophobic or, at a minimum, less polar substrates. The conclusion here is that ligands would more favorably bind in their uncharged forms or that the phenylalanine rings could be acting as hydrogen bond (π-cation) acceptors. The 3D-QSAR results above support the assertion of "neutral state" binding to AcrB and may be interpreted as evidence of at least transient binding at this site being a key step in the efflux process.

In contrast, the TolC lumen is exposed to the extracellular environment due to its position on the bacterial outer membrane, and this lumen is very likely solvated – suggesting that ligands bound here would be favored in their charged form. The translocation of substrates from AcrB to TolC must expose them to water, providing a mechanism for transforming them from their uncharged to charged forms. Substrates, by necessity, interact with various parts of AcrB and TolC during their efflux extrusion and this process is certainly affected by a multitude of different interactions, and thus cannot be completely addressed by simple methods such as 2D or 3D-QSAR that are based on molecules in a single state, bound within a single well-behaved binding site/mode with the concomitant assumption of a pharmacophore recognition-driven process. While our 3D-QSAR model is unfortunately not extensible to other substrate families because of the need that the molecules have a common alignment, the results for the β-lactam data set is an indicator of key principles behind the pump mechanism.

2.2. What factors might affect efflux-mediated MIC ratios?

Since MIC ratios describe a balance between the rates of influx through the bacterial outer membrane and efflux through the AcrA-AcrB-TolC pump, any mathematical model for predicting MIC ratios must capture elements of these processes. Influx of antibiotics is likely to be a result of two different routes - a diffusion-controlled membrane permeability process similar to others that have been well modeled with LogP (for 1-octanol/water) of the ligands [55], and a porin facilitated influx into the cell [30], which is affected by size of the antibiotic in question (*vide supra*).

It is also known that AcrB has a very hydrophobic binding pocket [38]; it is not unlikely that LogP affects the rate of efflux as well. Previous studies have reported correlation between LogP and MIC ratios [35]. Although it would be preferable to use experimental LogP values, they are not uniformly available for the compounds of interest, and in fact often are

only relevant for the conditions, e.g., pH, temperature, etc., under which they were measured. LogP values were estimated using several methods; the best correlation between experimental and predicted LogP is shown by ALogPs (Table 1). LogP alone (see Figure 3) is clearly insufficient to predict logMIC effectively ($r^2 = 0.48$).

$$\text{logMIC} = 1.46 + (1.29) * \text{LogP} \quad \text{(eq. 1)}$$

The size of the substrate molecules would seem to be one of their most critical features, since it has been reported that influx of antibiotics through porins is affected by the size of antibiotic [30]. Larger molecules pass through porins very slowly, if at all. Since ingress of large molecules would be low, it can be expected that fast efflux will cause an increase in MIC ratios for these compounds. Larger molecules should be expected to show a positive correlation with MIC ratio. In order to account for conformational changes that a compound may undergo in *squeezing* through an opening, as a crude assessment of size, molecular widths were calculated by MD simulation. These values for each molecule are listed in Table 1.

In addition, to account for the dynamic nature of efflux, intermolecular interactions between efflux substrates and various zones or compartments of the pump must be coordinated in order to transport the substrates through the pump. Thus, the substrate molecules were docked into various locations within AcrB and TolC in their charged and uncharged forms (Figure 1). These docked positions are direct representations of interactions between substrate and the efflux pump subunits. Using the HINT model to score these interactions gives us an additional advantage of taking desolvation energy and entropy into account at these loci [40].

Important sites were thus initially surveyed with β-lactams within the binding and extrusion states of AcrB. Since the TolC lumen is open to the extracellular space, ligands present here should be solvated and exist mostly in their charged forms. β-lactams were used to also survey the potential sites within TolC.

2.3. Regression models of properties and interaction scores

A partial least squares (PLS) analysis was conducted to explore the interplay between the molecular properties of the substrates and the interactions between the charged/uncharged β-lactam antibiotics and the AcrB and TolC proteins. Multiple equations of correlation were obtained through exhaustive exploration of the descriptor space. Only those descriptors that showed both interpretable trends and significant contributions to a model fitting the set of 32 β-lactam molecules were retained.

With the important descriptors and binding sites thus identified, the 12 non-β-lactam substrates were docked and scored at those sites. These descriptors were also subjected to rigorous statistical screening in order to ascertain the probability of non-normal distribution. The best combination of descriptors yielded the multilinear model:

$$\begin{aligned}\text{logMIC} = &-1.31 - \left(1.7 \times 10^{-4}\right) * \text{HINT}_{\text{nB}} \\ &- \left(5.3 \times 10^{-4}\right) * \text{HINT}_{\text{cE}} \\ &+ \left(6.9 \times 10^{-4}\right) * \text{HINT}_{\text{AcrB(hole)}} \\ &- \left(1.0 \times 10^{-3}\right) * \text{HINT}_{\text{Z3}} \\ &+ 1.10 * \text{LogP} + 0.43 * \text{MolWidth}\end{aligned} \quad \text{(eq.2)}$$

Here, $HINT_{nB}$ is the HINT score of the neutral substrate docked to the AcrB binding state. Similarly, $HINT_{cE}$ is the HINT score of the charged substrate in the AcrB extrusion state. $HINT_{AcrB(hole)}$ and $HINT_{Z3}$ represent the HINT scores of the charged substrate at the intermonomeric space ("the hole") of AcrB and zone 3 of TolC, respectively. LogP values are as predicted by the ALogPs algorithm and MolWidth is the molecular width calculated as described in the methods below.

Cross-validation with leave-one-out on the data set yielded a q^2 of 0.56 and an r^2 of 0.66 with 2 components for equation 2 (P value < 10^{-6}). This equation produced results superior to LogP based prediction of antibiotic efflux (eq. 1) as confirmed by an F-test (P value < 0.002). In twenty trials, the same descriptor set was also found to be uncorrelated with randomly permuted MIC ratios, virtually eliminating any possibility of equation 2 being a chance correlation. Figure 4A displays the predictive model of equation 2 and Table 2 sets out the predicted MIC ratios and deviations calculated. To further evaluate the predictive ability of the model, the data set of 44 compounds was randomly divided into training sets of 33 compounds and test sets of 11 compounds. New models were built with leave-one-out cross-validation over the training sets and used to predict the logMIC of their corresponding test sets. The predicted test set logMIC for a typical run of this nature is illustrated in Figure 5 (A and B).

It is probably more important for drug discovery and design purposes to determine whether or not the MIC ratio of an antibiotic will be altered in the presence of efflux pumps than it is to predict its precise numerical MIC ratio. We will call those antibiotics that show much higher MIC ratios "high efflux" and those with relatively low MIC ratios "low efflux". These terms are not to be confused with the *rate* of efflux and are only meant to categorize the *effectiveness* of these compounds for inhibition of bacterial growth in the presence vs. absence of efflux pumps. Using a definition of "high" efflux as logMIC ≥ 4, the equation 2 model was able to identify low/high efflux molecules with a 93.18% (41/44) success rate. Other results, i.e., with different high/low cutoffs, are summarized in Table 3. Note that even the predictions in error with this binary classification scheme are often fairly close to the experimental efflux (see Table 2). A similar analysis conducted with the LogP-only model of equation 1 shows that, while both models perform well with a cutoff of 4, equation 2 is more predictive, especially at lower cutoff values (Table 3). This clearly suggests that the inclusion of interaction-based descriptors contributes to the predictive power of models under more stringent conditions. In summary, this method allows reliable predictions for whether higher concentrations of a given antibiotic will be required to achieve the same therapeutic effect in the presence of the AcrA-AcrB-TolC pump as when the pump is absent.

2.4. Model and descriptor interpretation

A key requirement for a useful predictive model is that the physicochemical implications of the model's descriptors are interpretable and intuitive. The descriptors must yield not only statistical information but also chemical information that can be applied to fruitful drug design. However, when considered together, as in a regression equation, the model should have more value than the sum of its parts, i.e., the individual descriptors. Table 4 shows the fractional contribution of each descriptor for the model. All descriptors in equation 2 were found to have a meaningful contribution in the prediction of logMIC values, with LogP having the largest, to the model. In this section the descriptors and their qualitative and quantitative contributions to the overall model are described.

First, we examine the role played by molecular size (width) in the biological process. Molecular width (MolWidth) possesses a positive correlation coefficient; i.e., larger substrates increase MIC ratio. This is reassuring because of two reasons: (1) it is already well known that influx of molecules is affected by their size because they must permeate the

outer bacterial membrane by passing through porins [30]. Since the size of the porin lumen is comparable to the size of antibiotics, larger molecules simply will not enter the bacterial cell easily. Coupled with efflux, this would translate into a much larger effect on MIC ratios; (2) additionally, we propose that larger sized antibiotics might also cause an `induced fit' effect on AcrB, thus forcing AcrB to transform from the unbound and flaccid access state to the much larger binding state. Concomitantly, substrate bulk could force a change in tertiary structure that causes AcrB to assume the extrusion state. There is some experimental evidence to support this hypothesis: the crystal structure (PDB ID 2drd) of AcrB clearly shows that the binding pocket is shrunken in the access state, while in the binding state it is wide open with an entrenched ligand [38]. It has previously been suggested that the proton pump mechanism provides energy for conformational change [57]. We propose that binding of larger ligands might supplement this effect by providing steric “encouragement.”

LogP represents more than solubility and related phenomena. This is especially true in the study of MIC ratios because the phenomenon is composed of two independent events: influx of the antibiotic through the outer membrane, followed by extrusion of the same by an efflux pump. LogP plays an important role during permeation of the antibiotic through the outer membrane into the periplasmic region, governed by principles similar to Lipinski's rules [28,29]. Also, since the AcrB binding pocket, which captures antibiotics and other substrates to commence efflux, is hydrophobic, the importance of this descriptor in the process of efflux should be obvious. Indeed, LogP contributes almost 41% to equation 2 (Table 4), but we cannot “partition” this contribution into the multiple effects that LogP represents (*vide infra*).

$HINT_{nB}$ appears in the equation because these substrates appear to bind to the AcrB protein's binding state in their neutral form (as suggested by the 3D QSAR above). It has previously been proposed that efflux pumps capture antibiotics from the periplasmic region, close to the inner membrane [38]. Since this region is lined on either side by lipid bilayers, this environment should be less polar than the cytosolic medium. We propose that this could partially shift the acid-base equilibria for ligands in the periplasmic space somewhat towards their non-ionic (more hydrophobic) forms. As there are several phenylalanine residues in the AcrB binding pocket, it is thus quite hydrophobic, and this, the entrance to the pump, should preferentially bind less polar ligands, or those that are in a non-ionic form at the time of capture. This suggests that a number of substrate molecules may linger near the entrance and only pass within when their equilibrium-mediated ionization state matches the requirements of the AcrB binding pocket. Confusingly, however, this term in equation 2 has a negative correlation coefficient that suggests strong binding here *disfavors* efflux. This, in a sense, would appear to be true, as very strong binding to this site should cause the substrate to be “stuck” and not effluxed. However, we believe that this observation could be an artifact resulting from the small quantity and nature (value range) of training data available. A larger, more comprehensive, dataset may invert this sign, or even possibly indicate a non-linear behavior (see more in Supporting Information).

Once the ligand is bound to the binding state of AcrB, the protein undergoes a conformational change to assume the extrusion state. The ligand must leave the AcrB binding pocket at this stage. $HINT_{cE}$ represents the interaction of ligands at this site. It also has a negative correlation coefficient in the models, which is intuitively true – strong affinity for the extrusion state should be detrimental for efflux. On the other hand, the molecule must now enter the intermonomeric region of AcrB, the interactions at which are described by the $HINT_{AcrBHole}$ descriptor. The positive sign for this descriptor in our model suggests that an attraction towards this region might help extrude molecules from the AcrB binding pocket.

On traversing through the TolC lumen, the substrate would successively interact with multiple positions on the protein. In accord with our hypothesis that stronger interactions would slow down extrusion of ligands, the term $HINT_{Z3}$ correlates negatively with logMIC in our regression equations. This is easily explained by looking at the inner surface map of TolC (see Supporting Information, Figure S4). Zone 3, which happens to be a deep pocket, is found in the center of the TolC lumen surface.

In order to gain a complete understanding of efflux with respect to the parameter set we have applied, or indeed any set of parameters, the data set should probe a wide range of both experimental and descriptor values. That is, unfortunately, not the case for the currently available data. Even the most significant parameter in equation 2, LogP, is being probed over a narrow range; clearly, although not indicated by equations 1 or 2, a very high LogP is likely to inhibit influx and/or efflux and also compromise the MIC ratio. Previous QSAR analyses have invoked quadratic representations of LogP to account for this effect [58, 59]. Figure 3 illustrates this principle with the present data. Similarly, while strong affinities for any particular location of the efflux pump might be counterproductive for the passage of efflux substrates (as in equation 2), the substrates must also bind at least moderately. Hence, quadratic representations of these descriptors may be necessary to characterize their effect on efflux. However, this analysis would could as much as double the number of fitted parameters and is thus described in Supporting Information.

2.5. Proposed mechanism of efflux

Figure 6 summarizes our results with a proposed mechanism for the AcrA-AcrB-TolC efflux process. The dielectric environment within the periplasmic region is unknown, but it is likely that there is less water present between the two lipid bilayers than in either the extracellular or cytoplasmic regions (see Figure 6A that indicates the color scales used in the remaining panels of Figure 6). Furthermore, both peptidoglycan chains and a gel containing a wide variety of enzymes, which should tend to reduce the polarity, occupy the periplasmic space. Efflux substrates in the periplasm would exist in a reversible equilibrium between their charged and uncharged forms that slightly favors the uncharged forms (Figure 6B, inset) in this (slightly) more hydrophobic region. The uncharged forms will be more likely captured by the AcrB hydrophobic pocket (Figure 6B), upon which AcrB will assume the extrusion state (Figure 6C) partly due to the bulk of the substrate, as indicated by the positive correlation between logMIC and molecular width. It should be pointed out that if a substrate molecule binds too tightly to the binding state form of AcrB, or if it cannot sterically trigger the extrusion state of AcrB, that substrate would appear to be immune from extrusion and may block the pump's function. The absence of favorable substrate binding at this state would also preclude efflux. Once the extrusion state is formed, the AcrB entrance is closed and thus isolated from the periplasmic space but now open towards TolC, exposing the substrate to water present in the TolC lumen (Figure 6C). Similarly, the ligand's ionization equilibrium is concomitantly shifting towards the charged form (Figure 6D, inset), leading to release of the substrate into the intermonomeric space of AcrB (Figure 6D). Again, if the substrate ligand binds too weakly or too tightly to the AcrB extrusion state, it may not be effluxed and, in fact, may in the latter case block the pump. However, both these “clogging” events involve equilibria that may reverse to unclog the pump. The charged efflux substrate can now diffuse through water present in the TolC lumen (Figure 6E). Ultimately the substrate will be extruded into the comparatively polar extracellular medium (Figure 6F).

3. Conclusions

Despite the engineering of Nature to facilitate the extrusion of undesired molecules within a cell, there are likely to be multiple reasons why a particular antibiotic is more (or less)

amenable to extrusion by AcrA-AcrB-TolC efflux pumps. In addition to the obvious descriptors of hydrophobicity (LogP) and size, the ligand's ability to bind *and* release from various pockets within the pump machinery is at least as critical as the aforementioned descriptors that are not cognizant of its interactions with the pump. However, despite the relatively successful predictions of logMIC by the model described above, there are a number of considerations inherent in the approach that should be discussed. For example, the primary is the dataset itself. Unfortunately, the available data is both relatively small in quantity and skewed towards the lower range of logMIC values, which corresponds to the range of molecules of more clinical interest. Predictions for ligands in this upper range are thus less reliable – most of the outliers are of this nature. Our development of regression model equations with six descriptors on 44 substrates using PLS is not ideal. Although we have some comfort from the fact that these models were subjected to cross-validation, which yielded good statistical parameters, Wold and Dunn [60] state that, even when using PLS, regression studies are only valid when the number of independent variables is far less than the number of dependent target values. Nevertheless, the absence of any detectable correlation between our descriptor set and scrambled observations (logMIC values) clearly shows that our model (eq. 2) is not a result of overfitting or a statistical artifact. Clearly, we would like to have a larger data set, but restricted this analysis to the β-lactam class of antibiotics and some non-β-lactam antibiotics because of their loose chemical similarities that, in turn, suggest extrusion by the same pump and mechanism. Data from a wider class of antibiotics are potentially available in databases, but their use may be premature in testing a new computational method.

A second consideration is that, while MIC ratios have historically been used to study efflux, there are inherent flaws in this measure. Nikaido's recent work [30,31] shows that MIC ratios do not accurately describe the *kinetics* of the efflux mechanism as MIC ratios encode both the rate of influx for the antibiotic as well as its rate of efflux by the pump. In their view, MIC ratios are misleading as indicators of efflux. However, MIC ratios do represent the sum of the various effects that lead to a cell's demise when the cell's pump is present. In our models, the key LogP descriptor primarily represents the antibiotic influx, and possibly partitioning of ligands into the binding pocket of AcrB. The remaining descriptors in our models probably do describe the rate of efflux. When a larger and more diverse set of data (beyond β-lactam antibiotics) is available, it would be interesting to build similar predictive models for the $K_{0.5}$ and V^{max} kinetic parameters for the AcrA-AcrB-TolC efflux pump.

In conclusion, by probing and scoring substrate-protein interactions at various locations in a dynamic process, we have proposed a new computational approach that was used here to predict efflux values as represented by MIC ratios of the AcrA-AcrB-TolC efflux pump. The model suggested some interesting mechanistic details about the efflux process that seem intuitively true, such as the possibility that strong binding at a particular region of the efflux pump is detrimental for efflux. A larger, more diverse, data set may reveal additional insight or perhaps major modifications to our statistical models, but we believe principles behind our modeling methods are universally valid. On further development, this approach could be expanded to more non-β-lactam antibiotics, other efflux systems affecting antibiotic efflux, and potentially mammalian efflux systems that have been shown to extrude, among other molecules, anticancer chemotherapeutic agents. The key puzzle piece is obtaining structural data for the protein components of additional pump molecules.

4. Experimental

4.1. Crystal structures of AcrB and TolC

The crysta structures of AcrB and TolC (PDB codes 2drd and 1ek9 respectively) are available in the RSCB database [38,39]. The protein structures were modeled using Sybyl

version 8.1 (Tripos International) [61]. Hydrogens were added, followed by 1000 steps of Powell minimization with Gasteiger-Hückel charges while keeping the heavy atoms still. Then, the entire structure was minimized to a gradient of 0.005 kcal mol^{-1} $Å^{-1}$. Visual inspection revealed no unrealistic steric clashes between residues. Where required, the ligand was removed from the protein binding state (from PDB code 2drd) to void the docking region.

4.2. Efflux data and substrate molecules

A review of literature produced a curated set of 32 β-lactam antibiotics (Scheme 1) and 12 non-β-lactam antibiotics (Scheme 2), for which efflux data has been reported [17,23–25,27] in the form of minimum inhibitory concentration (MIC) ratios. All reported MIC ratios for these compounds were taken and an average MIC ratio was obtained. Because the experimental data are reported as powers of 2, i.e., 1, 2, 4, 8, 16, etc., logarithms (base 2) were calculated for these average MIC ratios and these logarithmic values were used as the dependent variable in our analyses (Table 1). This dependent variable is referred to as logMIC throughout this manuscript. The range of MIC values for wild type (WT) pumps across all sources is also reported here. The narrow range of WT MIC values for each antibiotic clearly indicates that similar if not identical pump strains were used during experimental procedures reported in these referenced works such that these MIC ratios are directly comparable.

As all of the substrate antibiotic molecules have acid and/or base functionalities, the structures were modeled in Sybyl in both their neutral (non-ionic) and charged (usually zwitterionic) forms, and minimized to a gradient of 0.005 kcal mol^{-1} A^{-1}. An attempt was made to initially place each compound in its lowest energy conformation by manually selecting from available rotamers.

4.3. Docking and scoring

The initial methods development for this study was performed with only the β-lactam ligand set (Scheme 1) and then applied to the full, extended dataset. All compounds were docked in both their neutral and charged forms using GOLD version 3.0 [62]. When docked to different parts of the TolC cavity (see Figure 1 and Results and Discussion for further description), 100 positions for each compound in Scheme 1 were obtained at each of eleven different overlapping areas of the protein, giving us a total of 1100 solutions per compound. However, a total of 2000 positions were obtained per compound when docked into the binding state or extrusion state of AcrB. The antibiotics in Scheme 2 were docked at four different positions as selected by the model found for the β-lactam compounds (see Results and Discussion), using identical procedures and parameters. All docked positions were scored using the Hydropathic INTeractions (HINT) forcefield [40–43] as reported in previous work [44,45]. The HINT forcefield has been previously shown to not only estimate enthalpic contributions towards binding but also entropic and solvation contributions [40–44]. The pose corresponding to the highest HINT score for each ligand at each position was selected for further analysis. These, the best-docked positions, were combined into protein-ligand structures that were minimized with 500 iterations. The interactions were re-scored at the minimized positions and these HINT scores were used as descriptors. Further explanation of the various HINT scores utilized in this manuscript was given above.

4.4. LogP calculations

LogP was calculated by using the ALogPs online server [63], which gave values of multiple LogP prediction methods including ALogPs, ALogP, MLogP, XLogP2 and XLogP3 [63–65]. Our regressions, however, stipulated that only one set of LogP values could be used in any resulting equation. The best correlation between predicted LogP and logMIC was shown

by ALogPs values. For comparison, and to evaluate the significance of differences between LogP calculated for charged and uncharged forms of the molecules, LogP values were also calculated for the β-lactams using HINT [40] ("calculate method", "essential" hydrogens only).

4.5. Prediction of molecular width by molecular dynamics simulations

Considering an efflux pump to be a tubular passage, one descriptor that we expected to correlate with efflux was molecular width, i.e., it should have a negative correlation with the rate of efflux (and hence with MIC ratios), akin to a very large ball not being able to pass through a tubular pipe. For this purpose, we used molecular dynamics simulations to calculate the effective width of the efflux substrates. Figure 7 describes the physical concept behind our calculations: we aligned each structure with the z-axis, and calculated the projections of its atoms on the x- and y-axes. The largest projection for each structure was taken as its width. For completeness, we ran a 1 ns simulation for each molecule using the Sybyl molecular dynamics module at 300K, recording snapshots every fs and performed width calculations every 100 fs of simulation. The resulting average of widths, calculated using 10,000 conformers for each ligand, was used as the molecular width descriptor.

4.6. 3D-QSAR

The 3D-QSAR analyses were performed using the Comparative Molecular Fields Analysis fields (CoMFA) module of Sybyl. The β-lactam molecules were manually aligned based on the common β-lactam substructure and re-minimized to ensure no structures with high internal strain energy were present during analysis. The set of 32 β-lactams was divided into two sets, a training set of 24 compounds and a test set of 8 compounds using a random number generator to avoid any bias in selecting training and test sets. 3D-QSAR studies were performed for the charged and neutral forms separately. All compounds thus aligned were placed in a grid (spacing 1 Å, margin 4 Å). Both standard CoMFA fields (hydrogen bonding, electrostatic and steric parameters) and HINT fields (hydropathic) were calculated at each grid point. Partial Least Squares (PLS) and Sample-distance Partial Least Squares (SAMPLS) [66] analyses were performed to correlate positions/properties of various chemical substituents with logMIC. Various descriptor field combinations were comprehensively explored, creating several models to explore the predictability of each. The selected best models, i.e., those showing highest cross-validated r^2 (q^2) at a minimum (optimum) number of components, identified the descriptor field sets to be used for further analysis.

4.7. Partial least squares statistical modeling

Descriptors, as described above, were calculated for each antibiotic molecule in the study (see Table S1 in Supporting Information). For validation, each descriptor was normalized against its highest value, followed by binning into regularly spaced intervals. These frequencies were plotted with SIMCA-P+ [67] to ensure that the expected Gaussian patterns were observed in normal probability plots.

Using solely the 32 β-lactams, a statistical analysis of docking/scoring scores and substrate physicochemical properties as descriptors was performed using PLS and SAMPLS. During this investigation, only one predicted LogP value was used in any given regression equation to avoid using multiple correlated descriptors. An exhaustive search through the various HINT score sets calculated, as described above, for substrate binding to the various pump zones, singly and in combinations, was performed to find the best possible descriptor combination. This model (*vide supra*), i.e., the combination of descriptors that best quantified logMIC, was then extended to the non-β-lactam set. The complete set of 44 compounds was separated into training sets of 33 compounds and test sets of 11 compounds.

Cross-validation was performed on these training sets and test sets to confirm internal stability of the models.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors would like to acknowledge the assistance provided by Dr. Vishal Koparde (VCU) in assisting with calculations and in preparation of this manuscript, by Dr. Marek Gierlinski (University of Dundee, UK) for assistance in reviewing statistical methods, by Drs. Wright Nichols and Ken Mattes (AstraZeneca, Boston, USA) for stimulating our interest in this problem, and by Dr. Martyn Symmons (University of Cambridge, UK) for providing us with the coordinates of his AcrA-AcrB-TolC efflux pump model that facilitated our understanding of its overall structure and function. This work was partially supported by U.S. NIH grant R01GM071894 (GEK). KCA (Ripon College, Ripon, Wisconsin, USA) was supported by the NIH/NSF Bioinformatics and Bioengineering Summer Institute (BBSI) program at VCU.

References

1. Gold HS, Moellering RC Jr. Antimicrobial-drug resistance. N. Engl. J. Med. 1996; 335:1445–1453. [PubMed: 8875923]
2. Neu HC. The crisis in antibiotic resistance. Science. 1992; 257:1064–1073. [PubMed: 1509257]
3. Walsh C. Molecular mechanisms that confer antibacterial drug resistance. Nature. 2000; 406:775–781. [PubMed: 10963607]
4. Spratt BG. Penicillin-binding proteins and the future of β-lactam antibiotics. J. Gen. Microbiol. 1983; 129:1247–1260. [PubMed: 6352855]
5. Broome-Smith J, Spratt BG. An amino acid substitution that blocks the deacylation step in the enzyme mechanism of penicillin-binding protein 5 of *Eschiricia coli*. FEBS Lett. 1984; 165:185–189. [PubMed: 6319180]
6. McMurry L, Petrucci RE Jr. Levy SB. Active efflux of tetracycline encoded by four genetically different tetracycline resistance determinants of *Eschericia coli*. Proc. Natl. Acad. Sci. USA. 1980; 77:3974–3977. [PubMed: 7001450]
7. Webber MA, Piddock LJV. The importance of efflux pumps in bacterial antibiotic resistance. J. Antimicr. Chemother. 2003; 51:9–11.
8. Bambeke VF, Balzi E, Tulkens PM. Antibiotic efflux pumps. Biochem. Pharmacol. 2000; 60:457–470. [PubMed: 10874120]
9. Nikaido H. Multidrug efflux pumps of gram-negative bacteria. J. Bacteriol. 1996; 178:5853–5859. [PubMed: 8830678]
10. Lin J, Michel LO, Zhang Q. CmeABC functions as a multidrug efflux system in *Campylobacter jejuni*. Antimicrob. Agents. Chemother. 2002; 46:2124–2131. [PubMed: 12069964]
11. Pumbwe L, Piddock LJV. Identification and characterisation of CmeB, a *Campylobacter jejuni* multidrug efflux pump. FEMS Microbiol. Lett. 2002; 206:185–189. [PubMed: 11814661]
12. Poole K. Efflux mediated resistance to fluoroquinolones in gram-negative bacteria. Antimicrob. Agents Chemother. 2000; 44:2233–2241. [PubMed: 10952561]
13. Gill MJ, Brenwald MP, Wise R. Identification of an efflux pump gene pmrA, associated with fluoroquinolone resistance in *Streptococcus pneumonia*. Antimicrob. Agents Chemother. 1999; 43:187–189. [PubMed: 9869592]
14. Nikaido H. Preventing drug access to targets: cell surface permeability barriers and active efflux in bacteria. Seminars Cell Developmental Biol. 2000; 12:215–233.
15. Kaatz GW, Seo SM. Inducible NorA-mediated multidrug resistance in *Staphylococcus aureus*. Antimicrob. Agents Chemother. 1995; 39:2650–2655. [PubMed: 8592996]
16. Saier MH Jr. Molecular phylogeny as a basis for the classification of transport proteins from bacteria, archaea and eukarya. Adv. Microbiol. Physiol. 1998; 40:81–136. [PubMed: 9889977]

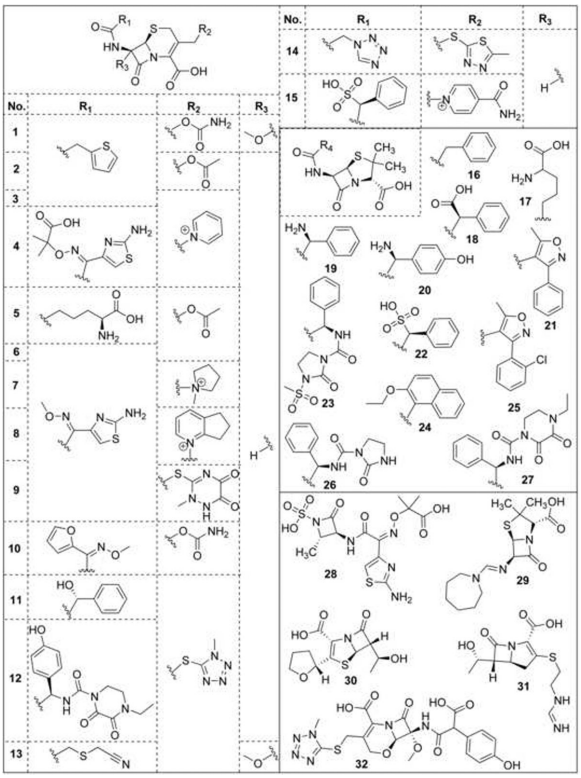
17. Lee A, Mao W, Warren M, Mistry A, Hoshino K, Okumura Y, Ishida H, Lomovskaya O. Interplay between efflux pumps may provide either additive or multiplicative effects on drug resistance. J. Bacteriol. 2000; 182:3142–3150. [PubMed: 10809693]
18. Ma D, Cook DN, Alberti M, Pon NG, Nikaido H, Hearst JE. Genes acrA and acrB encode a stress-induced efflux system of *Escherichia coli*. Mol. Microbiol. 1995; 16:45–55. [PubMed: 7651136]
19. Thanassi DG, Cheng LW, Nikaido H. Active efflux of bile salts by Escherichia coli. J. Bacteriol. 1997; 179:2512–2518. [PubMed: 9098046]
20. Sánchez L, Wubin P, Viñas M, Nikaido H. The acrAB homolog of Haemophilus influenza codes for a functional multidrug efflux pump. J. Bacteriol. 1997; 179:6855–6857. [PubMed: 9352940]
21. Gotoh N, Murata T, Ozaki T, Kimura T, Kondo A, Nishino T. Intrinsic resistance of *Escherichia coli* to mureidomycin A and C due to expression of the multidrug efflux system AcrAB-TolC: comparison with the efflux systems of mureidomycin-susceptible *Pseudomonas aeruginosa*. J. Infect. Chemother. 2003; 9:101–103. [PubMed: 12673418]
22. Yang S, Clayton SR, Zechiedrich EL. Relative contributions of the AcrAB, MdfA and NorE efflux pumps to quinolone resistance in *Escherichia coli*. J. Antimicr. Chemother. 2003; 5:545–556.
23. Mazzariol A, Cornaglia G, Nikaido H. Contributions of the AmpC β-lactamase and the AcrAB multidrug efflux system in intrinsic resistance of *Escherichia coli* K-12 to β-lactams. Antimicr. Agents Chemother. 2000; 44:1387–1390.
24. Nikaido H, Basina M, Nguyen V, Rosenberg E. Multidrug efflux pump AcrAB of *Salmonella typhimurium* excretes only those β-lactam antibiotics containing lipophilic side chains. J. Bacteriol. 1998; 180:4686–4692. [PubMed: 9721312]
25. Nishino K, Yamada J, Hirakawa H, Hirata T, Yamaguchi A. Roles of TolC-dependent multidrug transporters of *Escherichia coli* in resistance to β-lactams. Antimicr. Agents Chemother. 2003; 47:3030–3033.
26. McMurry LM, Oethinger M, Levy SB. Overexpression of marA soxS or acrAB produces resistance to triclosan in laboratory and clinical strains of Escherichia coli. Fed. Euro. Microbiol. Soc. Microbiol. Lett. 1998; 166:305–309.
27. Sulavik MC, Houseweart C, Cramer C, Jiwani N, Murgolo N, Greene J, DiDomenico B, Shaw KJ, Miller GH, Hare R, Shimer G. Antibiotic susceptibility profiles of *Escherichia coli* strains lacking multidrug efflux pump genes. Antimicr. Agents Chemother. 2001; 45:1126–1136.
28. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv. Drug Deliv. Rev. 1997; 46:3–26. [PubMed: 11259830]
29. Lipinski CA. Drug-like properties and the causes of poor solubility and poor permeability. J. Pharmacol. Toxicol. Meth. 2000; 44:235–249.
30. James CE, Mahendran KR, Molitor A, Bolla J-M, Bessonov AN, Winterhalter M, Pagès JM. How β-lactam antibiotics enter bacteria: a dialogue with the porins. PLoS One. 2009; 4:e5453. [PubMed: 19434239]
31. Nagano K, Nikaido H. Kinetic behavior of the major multidrug efflux pump AcrB of *Escherichia coli*. Proc. Natl. Acad. Sci. USA. 2009; 106:5854–5858. [PubMed: 19307562]
32. Lim SP, Nikaido H. Kinetic parameters of efflux of penicillins by the multidrug efflux transporter AcrAB-TolC of *Escherichia coli*. Antimicrob. Agents Chemother. 2010; 54:1800–1806. [PubMed: 20160052]
33. Hou T, Wang J, Zhang W, Wang W, Xu X. Recent advances in computational prediction of drug absorption and permeability in drug discovery. Curr. Med. Chem. 2006; 13:2653–2667. [PubMed: 17017917]
34. Johnson SR, Zheng W. Recent progress in the computational prediction of aqueous solubility and absorption. Amer. Assoc. Pharma. Sci. J. 2006; 8:E27–E40.
35. Clark DE. Computational prediction of ADMET properties: recent developments and future challenges. Ann. Reports. Comput. Chem. 2005; 1:133–151.
36. Ferreira MMC, Kiralj R. QSAR study of β-lactam antibiotic efflux by the bacterial multidrug resistance pump AcrB. J. Chemometrics. 2004; 18:242–252.

37. Kiralj R, Ferreira MMC. Molecular graphics approach to bacterial AcrB protein-β-lactam antibiotic molecular recognition in drug efflux mechanism. J. Mol. Graph. Model. 2006; 25:126–145. [PubMed: 16406715]
38. Murakami S, Nakashima R, Yamashita E, Matsumoto T, Yamaguchi A. Crystal structures of a multidrug transporter reveal a functionally rotating mechanism. Nature. 2006; 443:173–179. [PubMed: 16915237]
39. Kononakis V, Sharff A, Koronakis E, Luisi B, Hughes C. Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein transport. Nature. 2000; 405:414–419.
40. Kellogg GE, Abraham DJ. Hydrophobicity: is $LogP_{o/w}$ more than the sum of its parts? Eur. J. Med. Chem. 2000; 35:651–661. [PubMed: 10960181]
41. Cozzini P, Fornabaio M, Marabotti A, Abraham DJ, Kellogg GE, Mozzarelli A. Simple intuitive calculations of free energy of binding for protein-ligand complexes. 1. Models without explicit constrained water. J. Med. Chem. 2002; 45:2469–2483. [PubMed: 12036355]
42. Spyrakis F, Amadasi A, Fornabaio M, Abraham DJ, Mozzarelli A, Kellogg GE. The consequences of scoring docked ligand conformations using free energy correlations. Eur. J. Med. Chem. 2007; 42:921–933. [PubMed: 17346861]
43. Tripathi A, Fornabaio M, Kellogg GE, Gupton JT, Gewirtz DA, Yeudall WA, Vega NE, Mooberry S. Docking and hydropathic scoring of polysubstituted pyrrole compounds with antitubulin activity. Bioorg. Med. Chem. 2008; 16:2235–2242. [PubMed: 18083520]
44. Simoni D, Invidata FP, Eleopra M, Marchetti P, Rondanin R, Baruchello R, Grisolia G, Tripathi A, Kellogg GE, Durrant D, Lee RM. Design, synthesis and biological evaluation of novel stilbene-based antitumor agents. Bioorg. Med. Chem. 2009; 17:512–522. [PubMed: 19117761]
45. Abraham DJ, Kellogg GE, Holt JM, Ackers GK. Hydropathic analysis of the non-covalent interactions between molecular subunits of structurally characterized hemoglobins. J. Mol. Biol. 1997; 272:613–632. [PubMed: 9325116]
46. Burnett JC, Kellogg GE, Abraham DJ. Computational methodology for estimating changes in free energies of biomolecular association upon mutation. The importance of bound water in dimer-tetramer assembly for β37 mutant hemoglobins. Biochemistry. 2000; 39:1622–1633. [PubMed: 10677211]
47. Burnett JC, Botti P, Abraham DJ, Kellogg GE. Computationally accessible method for estimating free energy changes resulting from site-specific mutations of biomolecules: systematic model building and structural/hydropathic analysis of deoxy and oxy hemoglobins. Prot. Struct. Funct. Genet. 2001; 42:355–377.
48. Cashman DJ, Kellogg GE. A computational model for anthracycline binding to DNA: tuning groove-binding intercalators for specific sequences. J. Med. Chem. 2004; 47:1360–1374. [PubMed: 14998326]
49. Cashman DJ, Rife JP, Kellogg GE. Docking and hydropathic analysis of Hoechst 33258 with double-stranded RNA. Med. Chem. Res. 2003; 12:445–455.
50. Porotto M, Fornabaio M, Greengard O, Murrell MT, Kellogg GE, Moscona A. Paramyxovirus receptor-binding molecules: engagement of one site on the hemagglutinin-neuraminidase protein modulates activity at the second site. J. Virol. 2006; 80:1204–1213. [PubMed: 16414997]
51. Porotto M, Fornabaio M, Kellogg GE, Moscona A. A second receptor binding site on human parainfluenza virus type 3 hemagglutinin-neuraminidase contributes to activation of the fusion mechanism. J. Virol. 2007; 81:3216–3228. [PubMed: 17229690]
52. Rosenberg EY, Ma D, Nikaido H. AcrD of *Escherichia coli* is an aminoglycoside efflux pump. J. Bacteriol. 2000; 182:1754–1756. [PubMed: 10692383]
53. Elkins CA, Nikaido H. Substrate specificity of the RND-type multidrug efflux pumps AcrB and AcrD of *Escherichia coli* is determined predominantly by two large periplasmic loops. J. Bacteriol. 2002; 184:6490–6498. [PubMed: 12426336]
54. Yu EW, Aires JR, McDermott G, Nikaido H. A periplasmic drug-binding site of the AcrB multidrug efflux pump: a crystallographic and site-directed mutagenesis study. J. Bacteriol. 2005; 187:6804–6815. [PubMed: 16166543]

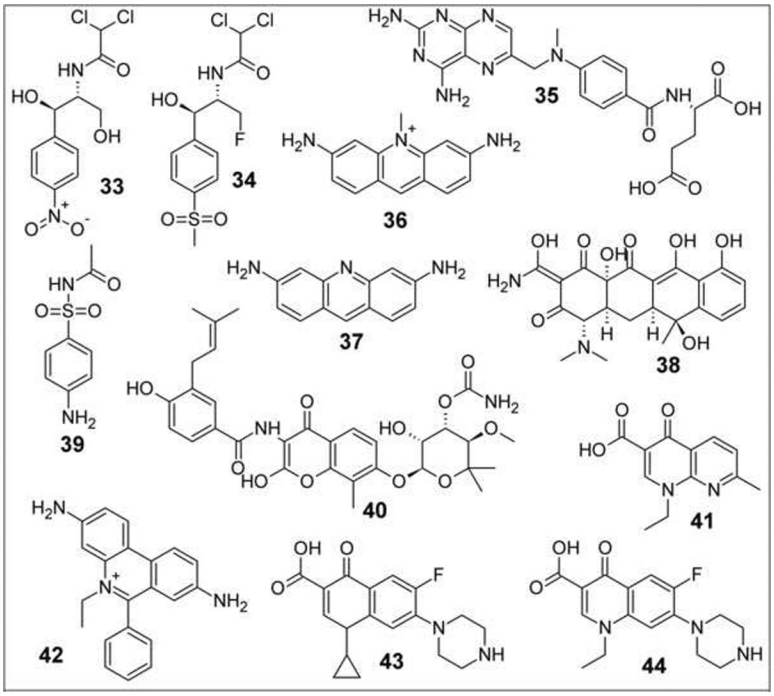
55. Cramer RD III, Patterson DE, Bunce JD. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. J. Am. Chem. Soc. 1988; 110:5959–5967. [PubMed: 22148765]
56. Tayar NE, Tsai RS, Testa B, Carrupt P-A, Hansch C, Leo A. Percutaneous penetration of drugs: a quantitative structure-permeability relationship study. 1991; 80:744–749.
57. Su C-C, Li M, Gu R, Takatsuka Y, McDermott G, Nikaido H, Yu EW. Conformation of the AcrB multidrug efflux pump in mutants of the putative proton relay pathway. J. Bacteriol. 2006; 188:7290–7296. [PubMed: 17015668]
58. Hansch C, Fujita T. ρ-σ-π Analysis. A Method for the Correlation of Biological Activity and Chemical Structure. J. Am. Chem. Soc. 1964; 86:1616–1626.
59. Romanelli GP, Cafferata LFR, Castro EA. An improved QSAR study of toxicity of saturated alcohols. J. Mol. Struct. (Theochem.). 2000; 504:261–265.
60. Wold S, Dunn WJ III. Multivariate quantitative-structure activity relationships (QSAR) – conditions for their applicability. J. Chem. Inf. Comput. Sci. 1983; 23:6–13.
61. Tripos Sybyl molecular modeling suite. version 8.1www.tripos.com/sybyl
62. Jones G, Willett P, Glen RC, Leach AR, Taylor R. Development and validation of a genetic algorithm for flexible docking. J. Mol. Biol. 1997; 267:727–748. [PubMed: 9126849]
63. Tetko IV, Gasteiger J, Todeschini R, Mauri A, Livingstone D, Ertl P, Palyulin VA, Radchenko EV, Zefirov NS, Makarenko AS, Tanchuk VY, Prokopenko VV. Virtual computational chemistry laboratory - design and description. J. Comput.-Aided Mol. Des. 2005; 19:453–463. [PubMed: 16231203]
64. VCCLAB. Virtual Computational Chemistry Laboratory. 2005. http://www.vcclab.org
65. Tetko IV. Computing chemistry on the web. Drug Discov. Today. 2005; 10:1497–1500. [PubMed: 16257371]
66. Bush BL, Nachbar RB Jr. Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. J. Comput.-Aided Mol. Des. 1993; 7:587–619. [PubMed: 8294948]
67. Umetrics, AB. Umeå; Sweden: http://www.umetrics.com

Highlights

- Prediction of antibiotic efflux by the AcrA-AcrB-TolC protein pump is investigated.
- A new QSAR method using docking scores at multiple sites in the pump is described.
- One model presented is > 93% accurate in classifying high and low efflux compounds.
- The approach is extensible to other antibiotic families and other pump proteins.



Scheme 1.
The chemical structures for the β-lactam antibiotic compounds used in the study.



Scheme 2.
The chemical structures for the other antibiotic compounds used in the study.

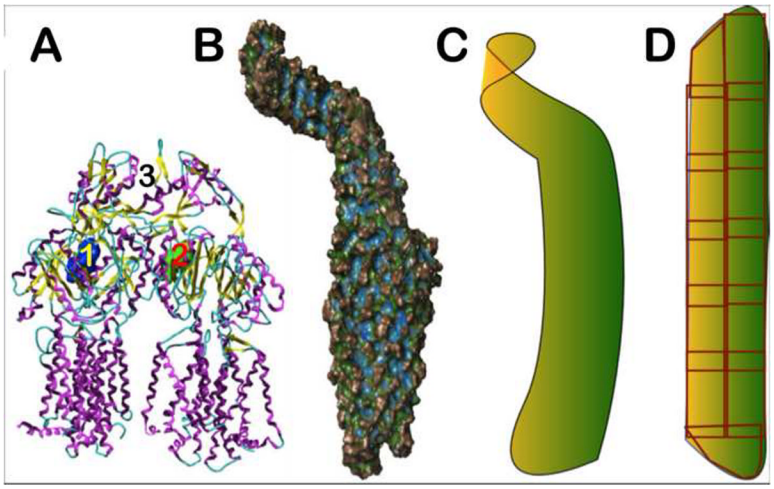


Figure 1. Docking Efflux Substrates Into Different Regions of AcrB and TolC

All 32 β-lactam and 12 non-β-lactam structures were docked into AcrB and several overlapping regions of TolC. **(A)** The hydrophobic pocket of AcrB is shown for both binding state (blue) and extrusion state (green) conformations. The numbers indicate the specific locations of AcrB in which the substrate molecules were docked: 1) binding state, 2) the extrusion state, 3) intermonomer region. **(B)** Molecular surface of the TolC channel colored according to cavity depth. The peaks are shown in copper brown, while troughs are shown in blue. **(C)** Ribbon cartoon of a TolC monomer, depicting its twisted shape. **(D)** The straightened cartoon avatar of one TolC monomer, showing different zones where antibiotic structures were docked. All the zones overlapped sufficiently to ensure a thorough investigation of molecular interactions between the entire TolC lumen surface and each antibiotic.

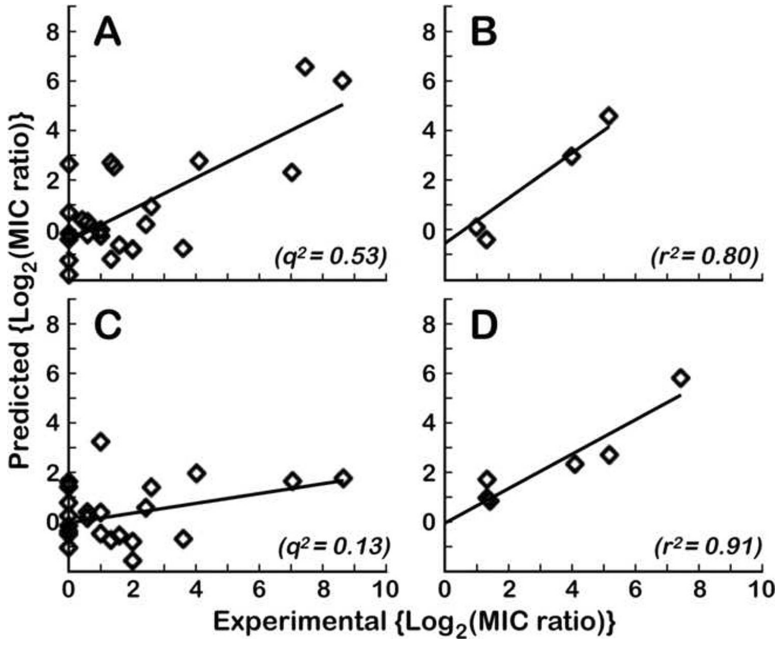


Figure 2. Training set and test set validations for 3D-QSAR models

Two runs show variable predictions of Log_2(MIC ratio) values of substrates. The **(A)** training set and **(B)** test set for the first run are shown along with the **(C)** training set and **(D)** test set for the second run. The cross-validated r^2 (q^2) is indicated for the training sets and the predictive r^2 is indicated for the test sets. While the first run shows a useful model, the second signifies instability of the model.

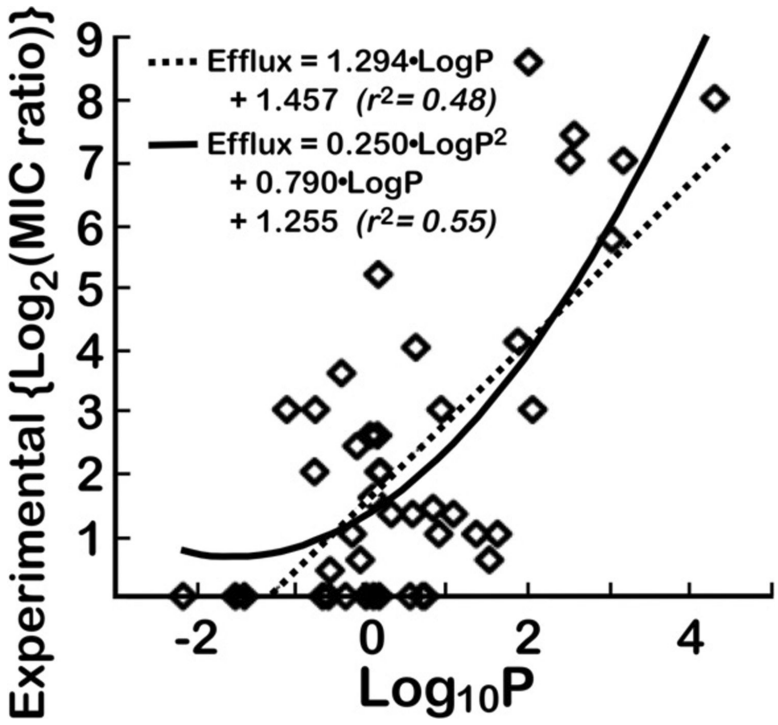


Figure 3. Correlation Between ALogPs Predicted LogP Values and logMIC

The plot of predicted LogP values versus logMIC values shows an r^2 value of 0.48, illustrating the fact that LogP alone does not provide accurate predictions of logMIC values. Also shown is the quadratic fit of logMIC with respect to LogP (logMIC = a + b*LogP + c*Logp^2), which shows an r^2 value of 0.55.

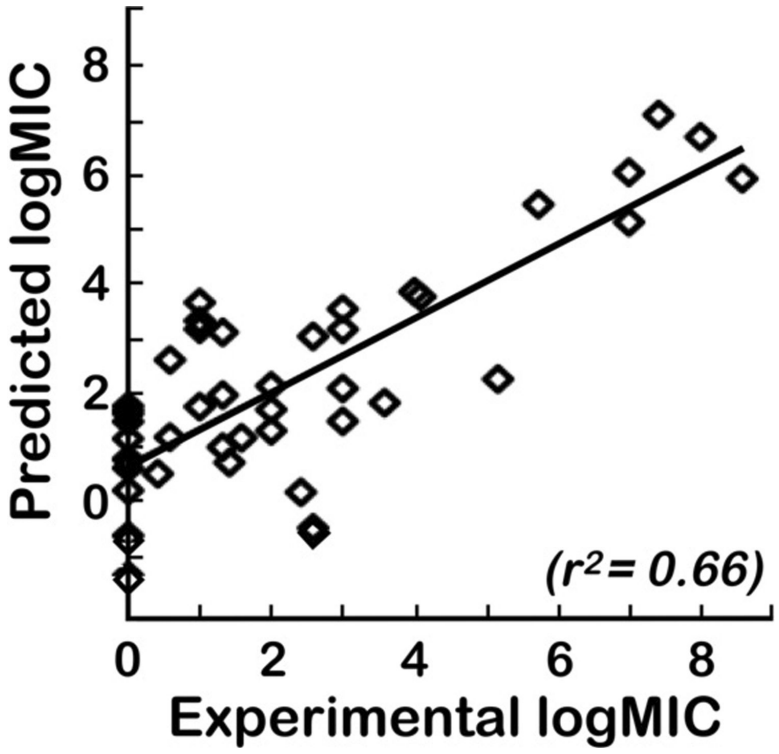


Figure 4. Correlation Plots for Predicted vs. Experimental Efflux as Obtained with PLS Model of equation 2

Predicted versus experimental logMIC values plotted for all 44 ligands based on equation 2.

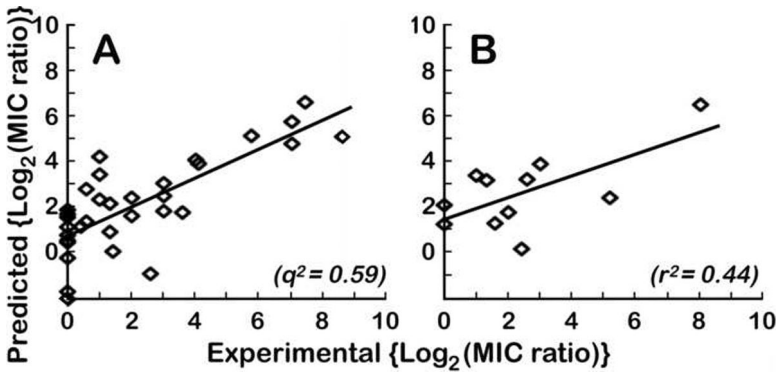


Figure 5. Validation of the PLS Model of equation 2

Predicted Versus Experimental logMIC For Training and Test Set Substrates. **(A)** Correlation between the predicted and experimental logMIC values for training set of 33 compounds using leave-one-out cross-validated model built with descriptors of eq. 2. **(B)** Correlation between the predicted and experimental logMIC values for independent test set of 11 compounds using equation of (A).

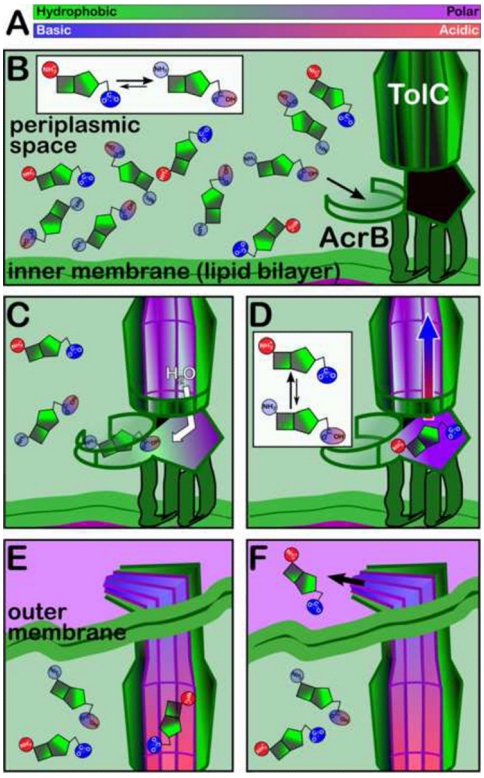


Figure 6. Proposed Efflux Mechanism

(A) Color ramps for hydrophobic-polar (green to purple) and acid-base (blue to red) spectra. **(B)** Antibiotics present in the periplasmic region exist in equilibrium between their charged and uncharged states (inset). The uncharged forms predominate due to the local environment, and are captured by the AcrB hydrophobic pocket. **(C)** AcrB assumes extrusion state partly due to the bulk of the substrate and is isolated from the periplasmic space. The cavity is now open towards TolC, exposing the substrate to water present in the TolC lumen. **(D)** The ionization equilibrium shifts towards the charged form (inset), and the substrate is released into the intermonomeric space of AcrB. **(E)** The charged efflux substrate is now able to diffuse through the water present in the TolC lumen. The electric field present inside the TolC lumen causes orientation of the substrate such that negatively and positively charged groups point towards opposite ends of the protein. **(F)** The efflux substrate is released into the extracellular space.

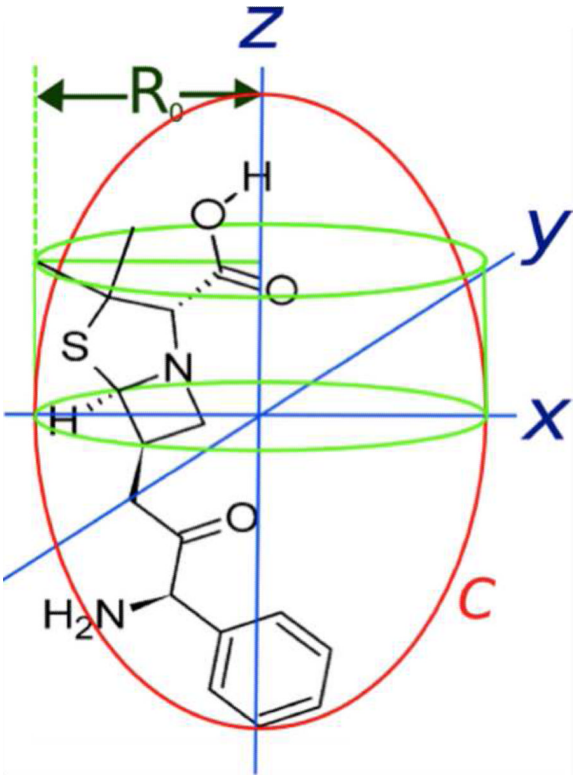


Figure 7. Calculation of Molecular Width

The figure shows ampicillin depicted in a 3D coordinate system, with the longest pair of atoms aligned with the Z axis. Projection of an atom on the XY plane is shown, and its length is depicted by R_0. The largest X or Y component of all such projections was taken as the radius of a cylinder *C* that can hold the entire structure, which is an approximate measure of the width of the structure.

Eur J Med Chem. Author manuscript; available in PMC 2013 June 01.

Table 1

Efflux and molecular parameters for data set molecules (see text for further description).

No.	Antibiotic	Reported MIC ratios[a]					MIC range for WT pump (μg/ml)	Avg. MIC ratio	Log_2 (avg. MIC ratio)	LogP (AlogPs)[c]	Molecular Width (Å)[d]
		Ref 17[b]	Ref 23[b]	Ref 24[b]	Ref 25[b]	Ref 27[b]					
1	Cefoxitin		4	4			4	4	2	0.22	8.82
2	Cefalothin		1	4			2–8	2.5	1.322	0.63	7.87
3	Cephaloridine		2	2			2–8	2	1	1.67	8.47
4	Ceftazidime		1				0.12	1	0	0.78	8.52
5	Cephalosporin C			1			16	1	0	−2.18	8.16
6	Cefotaxime			4	2		0.12	3	1.585	0.14	7.62
7	Cefepime		1	1			0.0075	1	0	−1.54	8.73
8	Cefpirome		1,2				0.015	1.5	0.585	1.57	8.99
9	Ceftriaxone		1	2			0.015–0.12	1.5	0.585	−0.01	8.66
10	Cefuroxime		16		8		1.56–2	12	3.585	−0.24	8.82
11	Cefamandole		4, 8		4		4	5.33	2.415	−0.05	8.28
12	Cefoperazone		2				0.03	2	1	−0.11	11.03
13	Cefmetazole		1	1	2		0.05	1.33	0.415	−0.38	8.44
14	Cefazolin		1	1			0.39–0.5	1	0	−0.4	8.32
15	Cefsulodin		1	1			16–64	1	0	0.6	9.54
16	Penicillin G		2	32			8–16	17	4.087	1.92	7.00
17	Penicillin N			1			8	1	0	−1.43	7.82
18	Carbenicillin		1, 4	4	1		1.56–8	2.5	1.322	1.13	7.07
19	Ampicillin		2		2	4	0.78–12.5	2.67	1.415	0.88	7.07
20	Amoxicillin		1				4	1	0	0.75	7.79
21	Oxacillin		512		256		1024	384	8.585	2.05	8.77
22	Sulbenicillin			4	1		8	2.5	1.322	0.37	7.33
23	Mezlocillin		32				1	36	5.170	0.21	10.16
24	Nafcillin			128	128		200	128	7	3.21	8.85
25	Cloxacillin		128	256	128		256–512	171	7.415	2.61	8.56

No.	Antibiotic	Reported MIC ratios[a]					MIC range for WT pump (μg/ml)	Avg. MIC ratio	Log_2 (avg. MIC ratio)	LogP (AlogPs)[c]	Molecular Width (Å)[d]
		Ref 17[b]	Ref 23[b]	Ref 24[b]	Ref 25[b]	Ref 27[b]					
26	Azlocillin		4, 8				16	6	2.585	0.2	9.92
27	Piperacillin		16				4	16	4	0.67	10.92
28	Aztreonam				1		0.05	1	0	0.06	7.87
29	Mecillinam		2				0.5	2	1	1.41	6.95
30	Faropenem				4		0.39	4	2	0.24	6.98
31	Imipenem		1				0.12	1	0	−0.19	7.36
32	Latamoxef			1			0.12	1	0	0.22	8.44
33	Chloramphenicol	4				8	4–6.25	6	2.585	0.11	7.27
34	Florfenicol					8	6.25	8	3	0.98	8.46
35	Methotrexate					8	640	8	3	−0.91	9.48
36	Acriflavine					128	400	128	7	2.56	9.39
37	Proflavine					8	100	8	3	2.10	5.99
38	Tetracycline					8	1.25	8	3	−0.56	8.88
39	Sulfacetamide					1	2000	1	0	0.15	6.27
40	Novobiocin		32		64	64	32–100	53.3	5.736	3.07	8.68
41	Nalidixic acid					2	3.125	2	1	0.95	9.62
42	Ethidium bromide					256	800	256	8	4.33	9.01
43	Ciprofloxacin					4	0.01	4	2	−0.57	9.07
44	Norfloxacin					1	0.004	1	0	−0.47	9.46

[a] The MIC ratio is the ratio of minimum inhibitory concentration (MIC) in the presence of efflux pump to the MIC in the absence of the pump. All reported MIC ratios are for the *E. coli* AcrAB-TolC pump, or the closely related *S. typhimurium* homolog.

[b] All WT strains used include JC7623, SH5014, TG1, HS414, and ECM1194 while all pump knockout strains include JZM120, SH7616, KAM3/pHSG398 or pHSG299, HS832, and ECM1694.

[c] LogP was predicted by the ALogPs method [63–65].

[d] The molecular width of each efflux substrate was calculated by performing a molecular dynamics simulation of 1 ns duration, followed by aligning the farthest atoms along the Z axis and measuring the projection of all other atoms on the XY plane. Assuming that the molecule spins rapidly, the largest such projection gives a rough measure of its molecular width.

Table 2

Efflux (MIC ratio) predictions for data set molecules.

Antibiotic	Average MIC ratio[a]	MIC ratio prediction (eq. 1)		MIC ratio prediction (eq. 2)	
		MIC ratio	Error	MIC ratio	Error
1	4	3.34	−0.66	3.34	−0.66
2	2.5	4.83	2.33	4.00	1.50
3	2	12.27	10.27	10.13	8.13
4	1	5.52	4.52	3.31	2.31
5	1	0.39	−0.61	0.42	−0.58
6	3	3.11	0.11	2.35	−0.65
7	1	0.69	−0.31	0.69	−0.31
8	1.5	11.22	9.72	6.21	4.71
9	1.5	2.72	1.22	2.37	0.87
10	12	2.21	−9.79	3.64	−8.37
11	5.33	2.62	−2.71	1.19	−4.15
12	2	2.49	0.49	3.46	1.46
13	1.33	1.95	0.62	1.50	0.16
14	1	1.92	0.92	1.61	0.61
15	1	4.70	3.7	2.32	1.32
16	17	15.35	−1.65	13.70	−3.30
17	1	0.76	−0.24	1.21	0.21
18	2.5	7.56	5.06	8.77	6.27
19	2.67	6.04	3.37	1.72	−0.94
20	1	5.38	4.38	2.88	1.88
21	384	17.25	−366.75	59.84	−324.17
22	2.5	3.82	1.32	2.09	−0.41
23	36	3.31	−32.69	4.88	−31.12
24	128	48.82	−79.18	64.80	−63.20
25	171	28.51	−142.49	133.07	−37.60
26	6	3.28	−2.72	8.33	2.33

Eur J Med Chem. Author manuscript; available in PMC 2013 June 01.

Antibiotic	Average MIC ratio[a]	MIC ratio prediction (eq. 1)		MIC ratio prediction (eq. 2)	
		MIC ratio	Error	MIC ratio	Error
27	16	5.00	−11	14.55	−1.45
28	1	2.90	1.9	1.80	0.80
29	2	9.72	7.72	12.73	10.73
30	4	3.40	−0.6	4.49	0.49
31	1	2.31	1.31	0.41	−0.59
32	1	3.34	2.34	3.49	2.49
33	6	3.03	−2.97	0.76	−5.24
34	8	6.61	−1.39	11.73	3.73
35	8	1.21	−6.79	2.89	−5.11
36	128	27.26	−100.74	34.78	−93.22
37	8	18.04	10.04	9.09	1.09
38	8	1.66	−6.34	4.35	−3.65
39	1	3.14	2.14	1.78	0.78
40	53.3	43.06	−10.24	43.32	−9.98
41	2	6.43	4.43	9.12	7.12
42	256	133.27	−122.73	100.92	−155.08
43	4	1.64	−2.36	2.56	−1.44
44	1	1.80	0.8	3.18	2.18

[a]The MIC ratio is the ratio of minimum inhibitory concentration (MIC) in the presence of efflux pump to the MIC in the absence of the pump.

Table 3

Classification accuracy of efflux predictive model.

Definition of "High" Efflux	% Correct Predictions by		
	(eq. 1)[a]	(eq. 2)	(eq. 3)[b]
≥4	90.91 (40/44)	93.18 (41/44)	93.18 (41/44)
≥3	79.54 (35/44)	79.55 (35/44)	84.09 (37/44)
≥2	54.54 (24/44)	72.73 (32/44)	70.45 (31/44)
≥1	72.73 (32/44)	77.27 (34/44)	72.73 (32/44)

[a] LogP only.

[b] See supporting information for information regarding equation 3.

Table 4

Fractional contribution of descriptors to models.

Descriptor	Percent Contribution	
	(eq. 2)	(eq. 3)[a]
$LogP^2$	-	20.5
LogP	40.9	22.0
$HINT_{z3}^2$	-	12.7
$HINT_{z3}$	18.2	10.9
$HINT_{AcrB(hole)}^2$	-	4.2
$HINT_{AcrB(hole)}$	11.9	3.5
$HINT_{nB}$	2.6	0.7
$HINT_{cE}$	13.0	13.3
MolWidth	13.5	10.6

[a]See supporting information for information regarding equation 3.