

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Identification of a family of glycoside derivatives biologically active against *Acinetobacter baumannii* and other MDR bacteria using a QSPR model

Francisco José Palacios-Can ^{1,2}, Jesús Silva-Sánchez ³, Ismael León-Rivera ², Hugo Tlahuext ², Nina Pastor ¹ and Rodrigo Said Razo-Hernández ^{2,*}

¹ Centro de Investigación en Dinámica Celular, Universidad Autónoma del Estado de Morelos, Av. Universidad 1001, Col. Chamilpa, Cuernavaca, Morelos, 62209 México

² Centro de Investigaciones Químicas, Universidad Autónoma del Estado de Morelos, Av. Universidad 1001, Col. Chamilpa, Cuernavaca, Morelos, 62209 México.

³ Centro de Investigación sobre Enfermedades Infecciosas, Instituto Nacional de Salud Pública, Av. Universidad 655, Col. Sta. Ma. Ahuacatitlán, Cuernavaca, Morelos, 62100 México

* Correspondence: rodrigo.razo@uaem.mx

Abstract: As the rate of discovery of new antibacterial compounds towards multidrug resistant bacteria is declining, there is an urge for the search of molecules that could revert this tendency. *Acinetobacter baumannii* has emerged as a highly virulent Gram-negative bacterium that has acquired multiple mechanisms against antibiotics and is considered of critical priority. In this work we developed a quantitative structure-property relationship (QSPR) model with 592 compounds for the identification of structural parameters related to their property as antibacterial agents against *A. baumannii*. QSPR mathematical validation ($R^2 = 70.27$, $R^N = -0.008$, $a(R^2) = 0.014$ and $\delta K = 0.021$) and its prediction ability ($Q^2_{LMO} = 67.89$, $Q^2_{EXT} = 67.75$, $a(Q^2) = -0.068$, $\delta Q = 0.0$, $\overline{r_m^2} = 0.229$, and $\Delta r_m^2 = 0.522$) were obtained with different statistical parameters; additional validation jobs were done using three sets of external molecules ($R^2 = 72.89$, 71.64 and 71.56). We used the QSPR model to perform a virtual screening on the BIOFACQUIM natural product database. From this screening our model showed that molecules **32** to **35** and **54** to **68**, isolated from different extracts of plants of the *Ipomoea* sp., are potential antibacterial against *A. baumannii*. Furthermore, biological assays showed that molecules **56** and **60** to **64** have a wide antibacterial activity against clinical isolated strains of *A. baumannii*, as well as other multidrug resistant bacteria including *Staphylococcus aureus*, *Escherichia coli*, *Klebsiella pneumonia*, and *Pseudomonas aeruginosa*. Finally, we propose **60** as a potential lead compound due to its broad-spectrum activity and its structural simplicity. Therefore, our QSPR model can be used as a tool for the investigation and search of new antibacterial compounds against *A. baumannii*.

Keywords: QSPR model; antibacterials; *Acinetobacter baumannii*; natural products; virtual screening

1. Introduction

Opportunistic infectious diseases caused by multidrug resistant bacteria represent a world-concerning health problem that is growing at an accelerated rate. Despite the immense quantity of literature and efforts sponsored by health committees, academia, and other non-governmental organisms on antibiotic resistance [1–3], there is still a lack of real education campaigns to promote the correct use of antibiotics. In accordance with recent reports, more than 2.8 million of antibiotic-resistant infections occur in the U.S. alone, with over 35,000 deaths as a result [3,4]. It is estimated that for 2050 a stunning 10 million deaths will be caused solely by antibiotic-resistant bacteria [5]. In Mexico the number of deaths caused by septicemia in hospitals have been increasing in recent years and since 2019 it is within the 15 main causes of deaths [6–8].

From the twelve bacteria listed on the website [9] of the World Health Organization (WHO), *Acinetobacter baumannii*, *Pseudomonas aeruginosa*, and several *Enterobacteriaceae* are considered of “critical” urgency. *A. baumannii* is an opportunist Gram-negative (GN) pathogen that has gained notorious attention because of its high virulence, its multiple mechanisms against antibiotics, and great capacity for adaptation to different environments [10–13]. Its incidence has been mainly related to pneumonia (associated with the use of ventilators), septicemia (due to contamination of central and peripheral airways), and infections at the site of the injuries [14,15].

As companies have dropped-out of the research and development (R&D) of new antibacterial drugs and less molecules have been approved by the FDA [16–20], the quest for novel potential candidates has decreased. Natural products (NPs) are a promising alternative to the use of traditional drugs because of their vast scaffold diversity and structural complexity [21]. These properties can be advantageous when comparing to typical synthetic small-molecule compounds: high molecular mass [22], large number of sp^3 carbon and oxygen atoms, which also correlates with low cLogP values (or higher hydrophilicity) [23–26], and greater rigidity [27]. Nonetheless, identifying bioactive compounds of interest is challenging and often takes additional time for isolation, complete characterization and, if afforded, full synthesis [28,29]. Several analytical techniques have proven to be of relevance for this task, for example, the use of computational resources, which has reduced the amount of time and optimization of drug candidates. Quantitative Structure-Activity/Property relationships (QSAR/QSPR) have allowed the search and optimization of better bioactive molecules by determining which physicochemical and structural features (molecular descriptors) are key-points for the biological activity [30].

Virtual screening (VS) comprises the use of computational tools to search and analyze large databases of small molecules, to identify potential bioactive compounds. VS can be divided into two major categories depending on the type of information available: ligand-based virtual screening (LBVS) and structure-based virtual screening (SBVS), both of which have been reviewed elsewhere [31–33]. Nevertheless, many other types of techniques have been developed to improve the accuracy of activity prediction. In this sense, the use of QSAR/QSPR as an approach for virtual screening of large libraries of small compounds has proven to accelerate the rate of the discovery process by reducing the number of potential candidates. When comparing the hit rates of techniques like High-throughput screening (HTS) with the QSAR/QSPR-based virtual screening, it is seen that the hit rate of HTS ranges between 0.01% and 0.1%, while for the latter it spans between 1% and 40%. [34] This has found application in the search of new antimalarial [35], anti-schistosomiasis [36], anti-tuberculosis [37] and antiviral [38,39] drugs, for which several compounds proved to be active.

Due to the high resistance to different antibiotic treatments caused by *A. baumannii*, worldwide research groups have carried out important efforts in the search for compounds against this pathogen. Most of them have carried out QSAR-type studies to determine their biological properties based on the molecular structure. However, a problem regarding these QSAR models is the use of small sets of compounds, mainly those synthesized and tested in the same work with minor chemical changes at the core structure. Furthermore, small datasets considering molecules acting against multiple pathogens have the disadvantage that it is necessary to seek/use as many models as possible to determine and predict the antibacterial activity for these sets of compounds. Prado-Prado *et al.*, developed a QSAR analysis by introducing entropy-like molecular descriptors for their models to predict the antibacterial activity of several drugs against different strains of bacteria [40]. Semenyuta and collaborators established several QSAR models for the activity of imidazolium ionic liquids with the use of neural networks and random-forest regressions [41], allowing them to use multiple molecular descriptors to correlate the structure with the bioactivity of these new compounds towards *A. baumannii*. Nonetheless, a main drawback of these QSAR analyses is the use of complex molecular descriptors that are often difficult to interpret and handle, limiting their applicability and simplicity.

In the present work we have developed a QSPR model of active compounds (from synthetic to NPs) against multidrug resistance (MDR) *A. baumannii*, by means of the genetic algorithms (GA) technique, using several molecular 0D, 1D and 2D-descriptors. The QSPR model was employed to identify structural features of the bioactivity compounds, within the dataset, that can be associated with their pharmacokinetic aspects (absorption and distribution). Therefore, one of the objectives of our QSPR model is to predict the entry of the compounds into *A. baumannii* [42–44]. Then, our QSPR was used to identify potential antibacterial candidates from a NP-database. Furthermore, we obtained and carried out the biological evaluation of these candidates, corroborating the prediction of our QSPR model.

2. Results and discussion

2.1. QSPR model validation

As a first approach, regression models were built using genetic algorithms to select the most appropriate descriptors. After selection of the descriptors, multiple linear regression analysis was performed to generate suitable models that could allow us to categorize the biological activity of the dataset. The best QSPR model for antibacterial activity against *A. baumannii* consists of fifteen descriptors as follows:

$$\begin{aligned}
 p\text{MIC} = & (0.001 \pm 0.000)\text{D/Dr06} + (-0.438 \pm 0.004)\text{GATS6m} + (0.529 \pm 0.004)\text{nArCOOH} + \\
 & (1.249 \pm 0.005)\text{nRCONH2} + (0.334 \pm 0.001)\text{nROR} + (-0.429 \pm 0.006)\text{nImidazoles} + \\
 & (0.115 \pm 0.000)\text{nHDon} + (-0.204 \pm 0.001)\text{nHBonds} + (1.257 \pm 0.005)\text{C018} + \\
 & (0.476 \pm 0.001)\text{C029} + (1.149 \pm 0.004)\text{C032} + (-0.105 \pm 0.000)\text{H051} + (-0.186 \pm 0.001)\text{N075} + \\
 & (-0.555 \pm 0.001)\text{N079} + (0.025 \pm 0.000)\text{TI2} + 4.292(\pm 0.005)
 \end{aligned} \quad (1)$$

$$\begin{aligned}
 R^2 &= 70.278(\pm 0.907); R^2_{ADJ} = 69.162(\pm 0.973); a(R^2) = 0.014(\pm 0.000); \\
 s &= 0.462(\pm 0.000); F = 62.978(\pm 8.418); Q^2_{LMO} = 67.886(\pm 1.043); \\
 Q^2_{BOOT} &= 66.882(\pm 1.104); Q^2_{EXT} = 67.747(\pm 5.414); a(Q^2) = -0.068(\pm 0.000); \\
 \delta K &= 0.021 (0.000); \delta Q = 0.000 (-0.005); R^P = 0.015 (0.100); R^N = -0.008 (-0.054)
 \end{aligned}$$

All statistical parameters were obtained as their average values (see Table S3), for example, the square correlation coefficient (R^2) of $70.278(\pm 0.907)$, and the R^2_{ADJ} of $69.162(\pm 0.973)$. The Fischer F and the standard deviation (s) are $62.978(\pm 8.418)$ and $0.462(\pm 0.000)$ respectively, indicating that our model is acceptable. Also, redundancy and overfitting rules were checked to determine the nature of the descriptors used in the model. In this sense, the overfitting rule, $R^N = -0.008$ (-0.054), was approved fairly while the redundancy rule, $R^P = 0.015$ (0.100), indicated that some descriptors, nHDon and nHBonds, are correlated to the dependent variable. However, these descriptors cannot be removed as they are important for the correct description of our regression model. Furthermore, the prediction ability of the model was validated by the leave-many-out cross-validation, $Q^2_{LMO} = 67.886(\pm 1.043)$, a value indicating that the regression model has good predictive power. The robustness parameter was indicated by the high value of $Q^2_{BOOT} = 66.882(\pm 1.104)$ based on bootstrapping, which was repeated 5000 times.

An external validation was essential as a high Q^2_{LMO} only indicates a good internal validation, but it does not show a high prediction capability of the created model. Therefore, for the external validation procedure, 70% of all the molecules in the dataset were randomly selected for the training process, and the remaining 30% were used as the test set. This process was repeated six times; their plots are shown in Figure 1, with their upper/lower confidence intervals at a 95% confidence level. The Y-scrambling test was used on the training-test set, giving the new values of $a(R^2) = 0.014(\pm 0.000)$ and $a(Q^2) = -0.068(\pm 0.000)$. These new values were lower than the original ones, confirming that our model is reliable. With the same purpose, the Asymptotic Q^2 rule, $\delta Q = 0.000$ (-0.005), was employed. Therefore, the model in (1) passed all the statistical tests proposed by Roy *et al.* [45–47], as an average value derived from ten experiments shows: (a) $Q^2 = 67.88(\pm 1.043)$; (b) $r^2 = 0.679(\pm 0.000)$; (c) $\frac{r^2 - r_0^2}{r^2} = 0.001(\pm 0.000)$; (d) $k = 0.999(\pm 0.000)$ (or $k' = 0.991(\pm 0.000)$); (e) $|r_0^2 - r'^2_0| = 0.127(\pm 0.000)$. For an acceptable prediction, the

value of Δr_m^2 should preferably be less than 0.2, while $\overline{r_m^2}$ should be greater than 0.5. In our model, Δr_m^2 presents a value of 0.229(± 0.000), while $\overline{r_m^2}$ has an average value of 0.552(± 0.000). A complete list for each evaluation can be seen in Table S4.

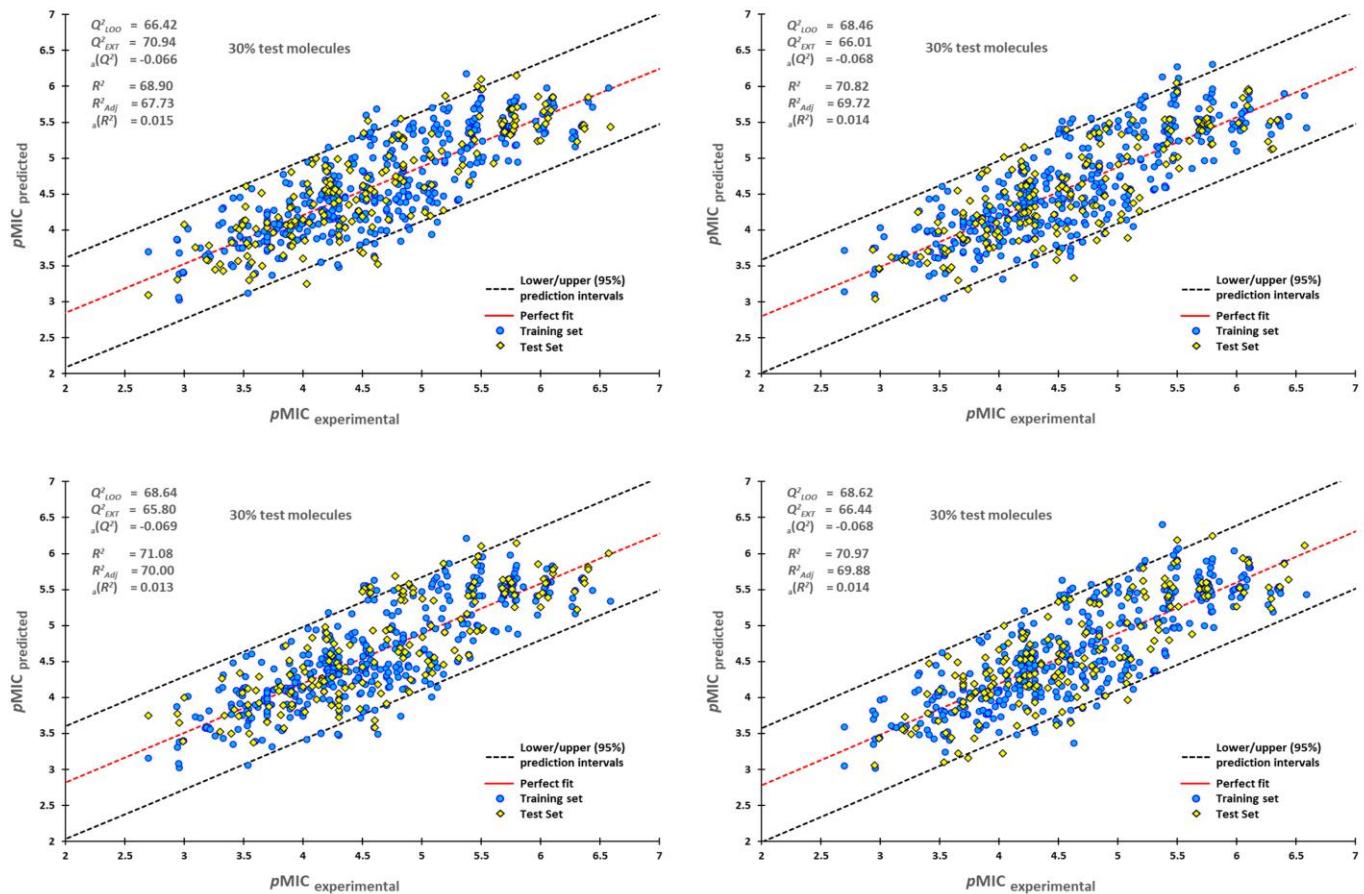


Figure 1. Scatterplots of predicted $p\text{MIC}$ against experimental $p\text{MIC}$ values. Blue dots represent molecules of the training set (70%), and yellow diamonds depict molecules used for the test set (30%). For each plot, the percentage of molecules used in the training and test datasets were randomly chosen.

The applicability domain is graphically depicted by the Williams plot in Figure 2. For each compound, the leverage values can be calculated and, by plotting these values against the standardized residuals, it is possible to establish the applicability domain of the developed model [48]. This allows detection of molecules that our model cannot predict adequately, thus considered as outliers [49,50], molecules with distinctive structures (high leverage outliers, $h > h^*$), or those associated to the response (predicted residuals $> 3^*\text{SDEC}$). All compounds that are outside the limits established by the leverage warning and three times the standard deviation in error calculation are outliers.

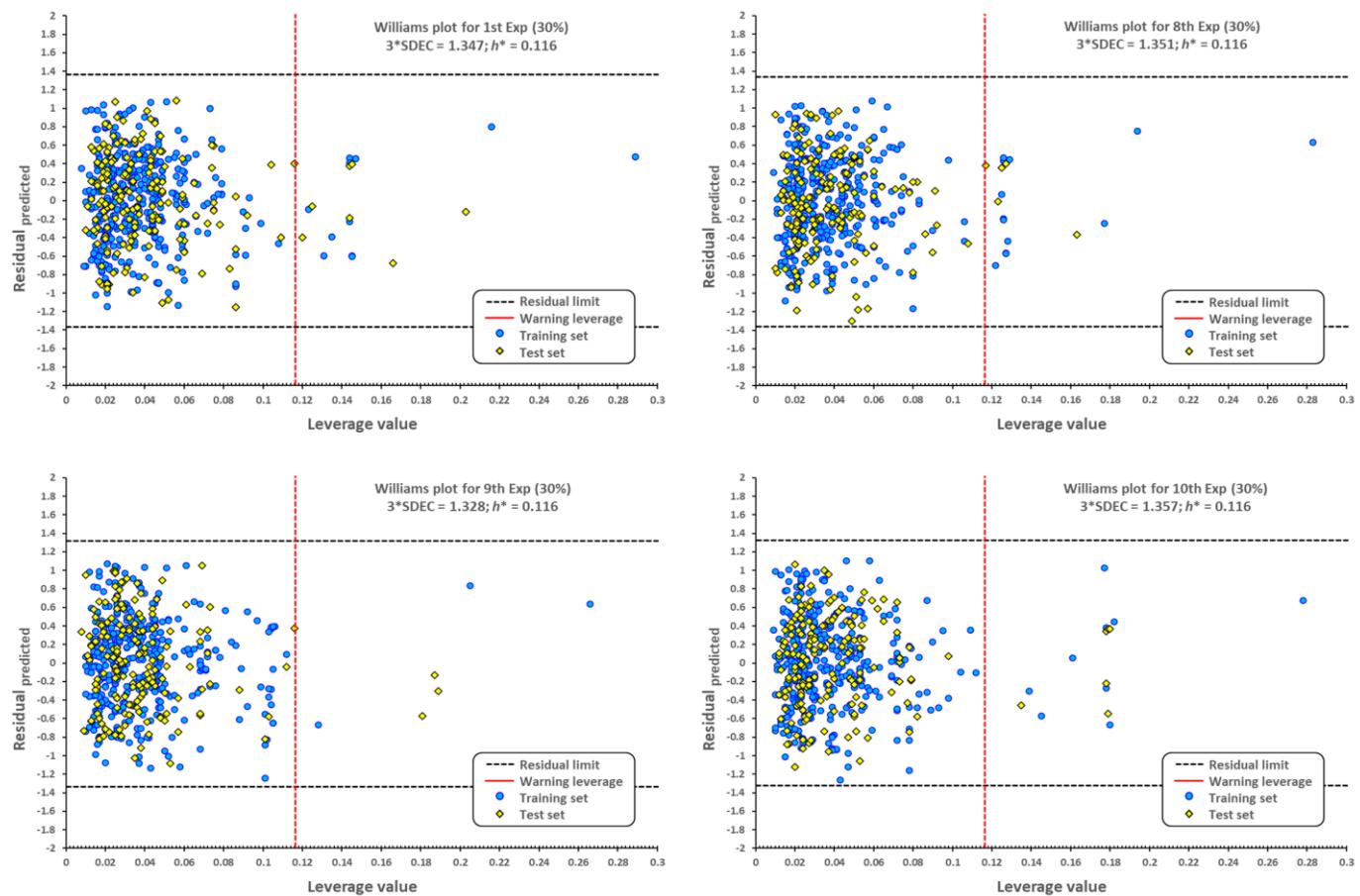


Figure 2. Williams plots for molecules with antibacterial activity against *A. baumannii*. The dotted vertical line in red indicates the warning leverage limit ($h^* = 3p/n$, where n is the number of molecules and p is the number of descriptors in the model plus one). The upper/lower dotted horizontal lines in black represents the boundaries for which the triple of the standard deviation (3^*SDEC) value is used.

As seen from the Williams plot, outliers are correlated to the structure of molecules. Due to the relative wide variety of molecular structures used in our model, detected outliers both from the training and test sets are very different (Figure 3).

In compound **1**, although it shares a similar structure with those of the Batzelladine alkaloids' family used in this model [51], the two cyclic ether-like motifs at the central positively charged nitrogen core, as well as two pendant primary aliphatic amine arms, are distinctively different from the rest of the analyzed molecules. Compounds **2** and **5** to **13** are considered as outliers because of the many positively charged nitrogen atoms present at the molecules. Two fluoroquinolone derivatives are present as outliers: compound **3** possesses a 3,5-difluoro-substituted pyridine instead of the common cyclopropyl or ethyl groups at nitrogen while compound **4** has a pyridine-type structure at the core as in nalidixic acid. These two features are unique among the set of fluoroquinolones used in our model. Compounds **14** and **15**, being both aminoglycosides, are seen as outliers from our model as it is suggested that amino groups are responsible for this distinction. Compound **16** is a flavanone-7-O-glycoside. Although there are many flavanones in the dataset, none of them present a disaccharide (or any mono- or polysaccharide), which makes **16** unique. On the other hand, many examples of substituted triazoles are seen in our model, but molecule **17** has a benzotriazole which unique, thus, it is considered as an outlier. Even though there are many compounds with aromatic alcohols, **18** (gallic acid) possess a benzenetriol motif which is not encountered in any other molecule. Structure **19** has the hydantoin functional group, which is unique among the set of active molecules against *A. baumannii*.

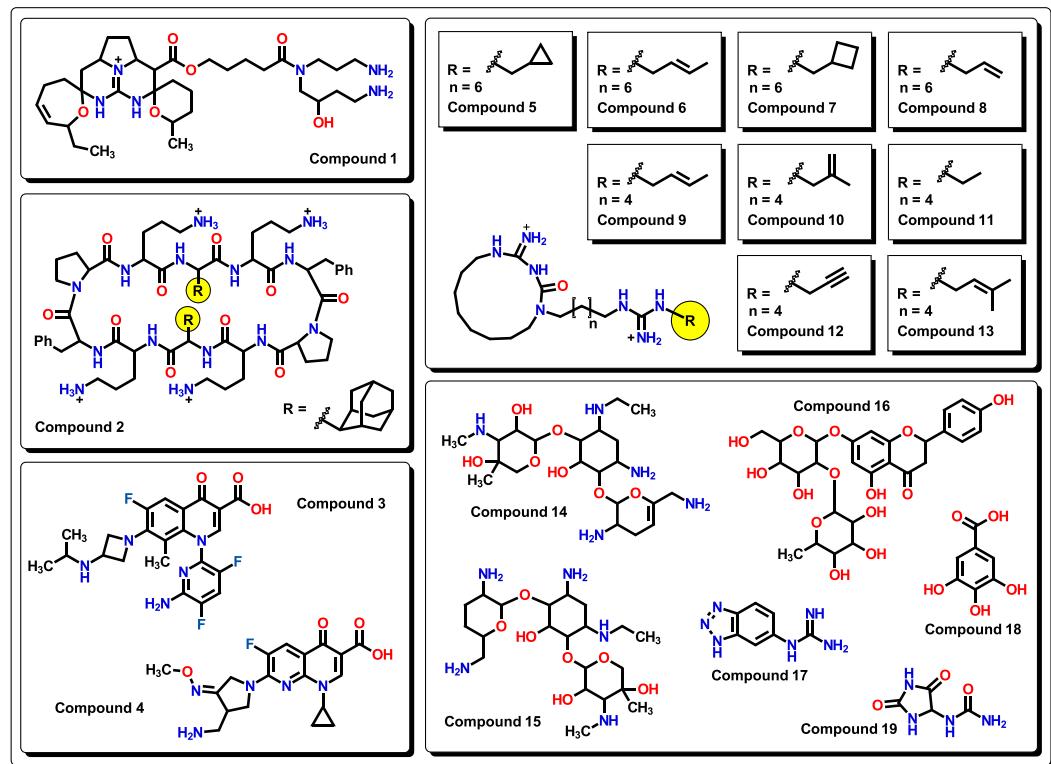


Figure 3. Chemical outliers obtained from the analysis of the Williams plots.

To test the reliability of our QSPR model, molecules which were not introduced in our initial dataset were employed as an external validation set to obtain their predicted $p\text{MIC}$ values. Three sets of compounds were used as follows: a) the first set from molecules reported from Matsingos *et al.* [52]; b) compounds reported by Singh [53], Wang [54], and Zhou [55] as the second set, and finally, c) chemical structures described by Lyons and collaborators [56]. For the three sets of data there is a good correlation between experimental and predicted $p\text{MIC}$ values, with R^2 values of 72.89, 71.64 and 71.56 respectively. On the other hand, compounds that exhibit, for example, positively charged nitrogen atoms like those reported by Vereshchagin and co-workers [57], are not well predicted by our model in accordance with the results of the outliers analyzed previously.

Our model applied to the first set of linezolid analogues with different C5-acylamino substituents gives an insight into their structural features. An increase in the $p\text{MIC}$ values is seen when moving from small-chain alkyl groups to cyclic non-aromatic and finally to aromatic substituents. This increase is shown in Figure 4.

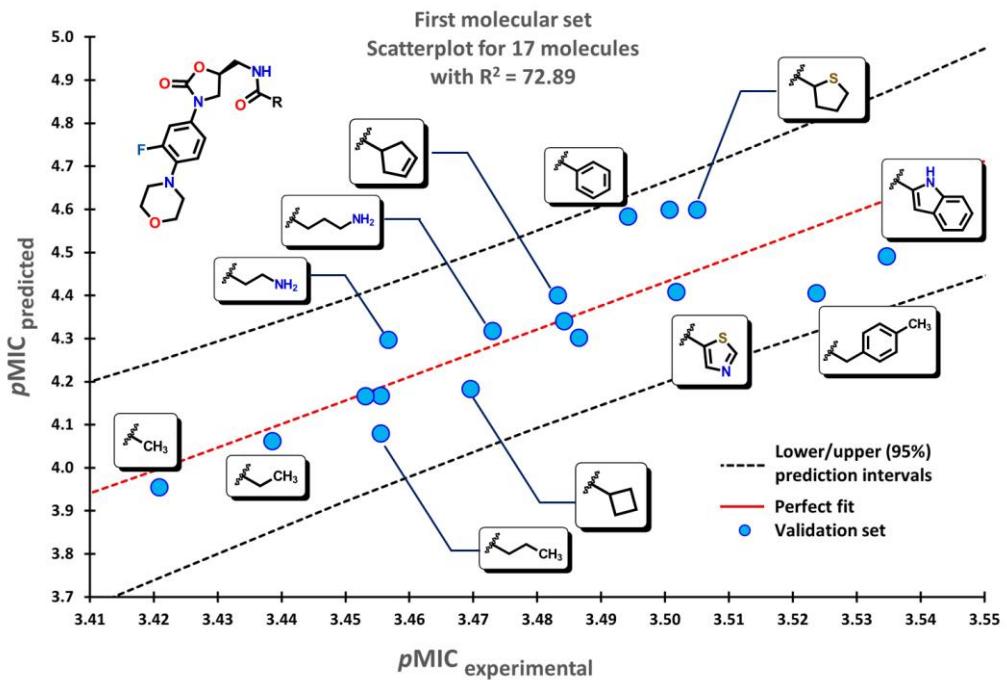


Figure 4. Scatterplot for molecular data set of linezolid analogues. Selected molecules are displayed within the plot showing the change of substituent.

The second set of compounds comprises three different groups of molecules for which our model classifies first the divin derivatives, moving into pyrazinoindole analogues, and finally with the subset of 2-aminothiazole sulfanilamide oximes, as seen in Figure 5.

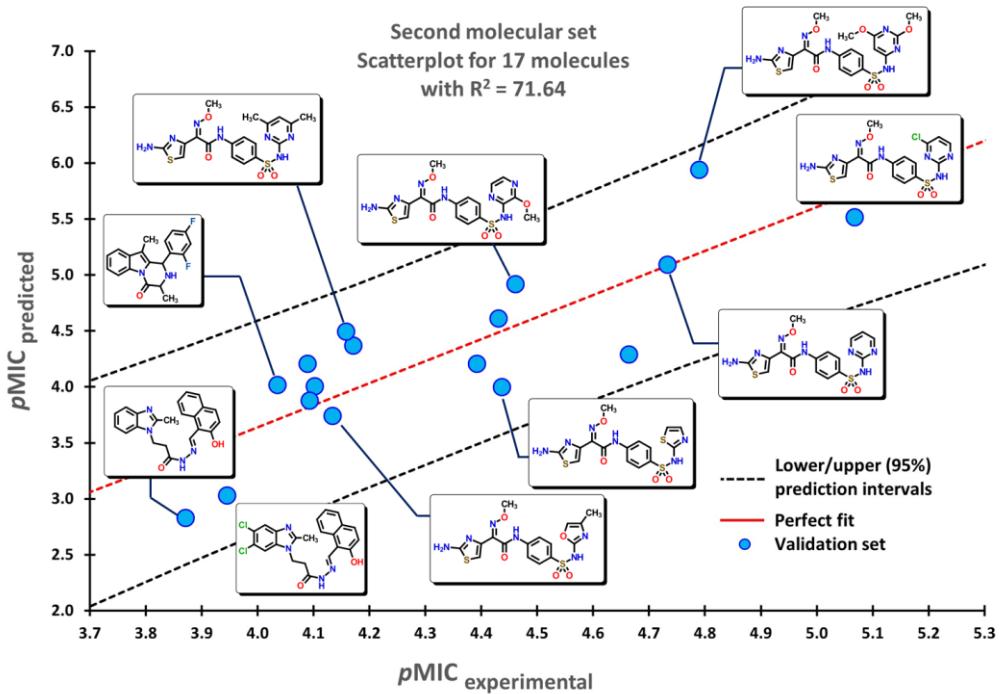


Figure 5. Scatterplot for molecular data set used the validation for the QSPR model. Selected molecules are displayed within the plots.

The last set of compounds comprises several oxazolidinone derivatives in Figure 6. The first molecules are classified in accordance with the structure of the 1,5-naphthyridin-

2(1*H*)-one, while the last ones have a 1,8-naphthyridin-2(1*H*)-one. Molecules at the center possess the nitrogen atom at different positions of the quinolin-2(1*H*)-one core.

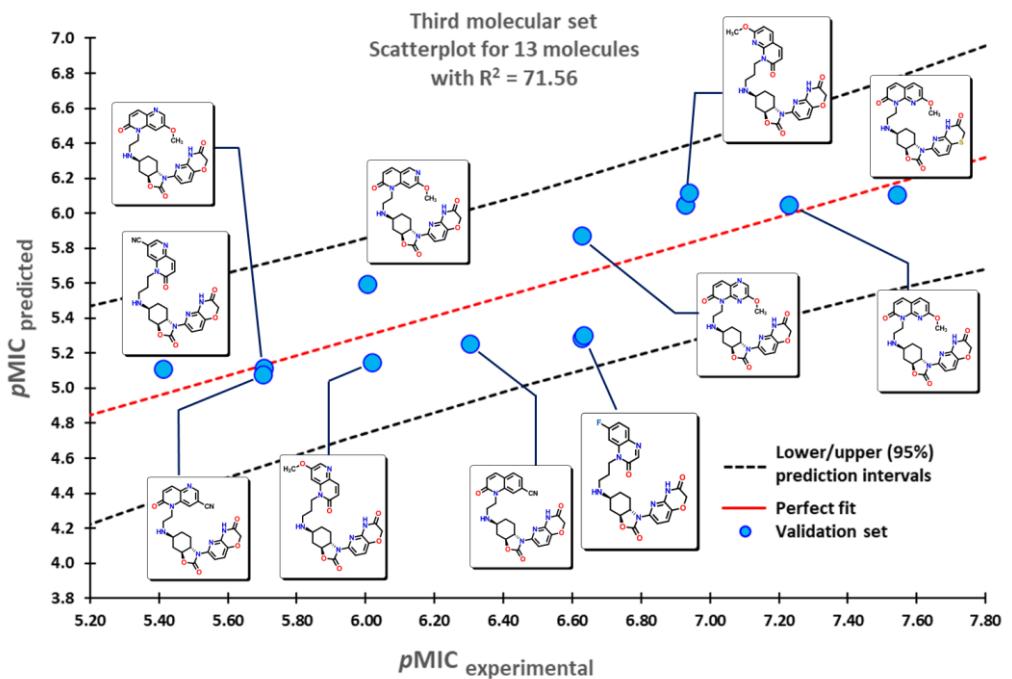


Figure 6. Scatterplots for molecular data sets used the validation for the QSPR model. Selected molecules are displayed within the plots.

2.2. QSPR interpretation

The understanding of the descriptors presented by the QSPR model allows us to gain some insights into chemical features of the molecules used in the model that are relevant for their antibacterial activity towards *A. baumannii*. Equation (1) displays two topological descriptors (D/Dr06 and TI2), one 2D-autocorrelation (GATS6m), six functional group counts (nArCOOH, nRCONH2, nROR, nImidazoles, nHDon and nHBonds), and six atom-centred fragments (C-018, C-029, C-032, H-051, N-075 and N-079), all of them being 2D-dimensional descriptors.

The first descriptor in the model is D/Dr06, a topological descriptor. Distance/detour ring indices (D/Drk) are calculated by summing up distance/detour quotient matrix row sums of vertices belonging to single rings in the molecule. These descriptors can be considered special substructure descriptors reflecting local geometrical environments in complex cyclic systems [58]. D/Dr06 displays a positive coefficient value, indicating that the presence of this descriptor enhances the activity of the molecule. This descriptor appears when a 6-membered cyclic structure is present in the molecule. From the set of compounds, most of the cyclic structures belongs to benzene type rings (both carbocyclic and heterocyclic). D/Dr06 has been used in similar way for the description of the anticancer activity of aromatic molecules [59]. The highest D/Dr06 value belongs to compound 2 where two adamantyl moieties are present in the molecule. Values of zero correspond to molecules which do not display any 6-membered cyclic system, such as compounds 22 to 25, seen from Figure 7. Furthermore, molecules that display high values of D/Dr06 also show high *pMIC* values.

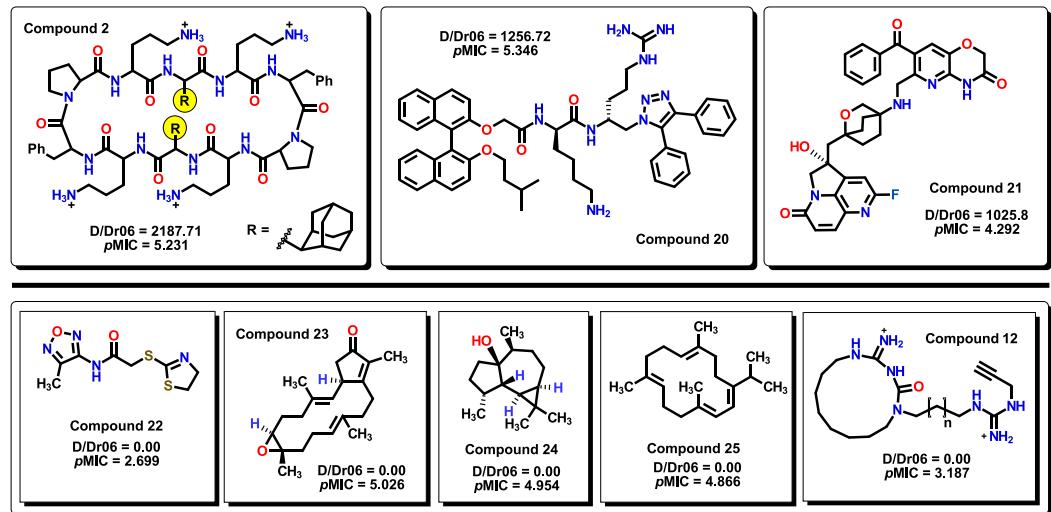


Figure 7. On the top row, the highest D/Dr06 values are displayed for compounds. Below, molecules which do not have any 6-membered ring in their structures display a zero value of the descriptor.

The second Mohar index [60], (TI2), is calculated from the eigenvalues of the Laplacian matrix as shown:

$$TI2 = \frac{4}{n_{SK} - \lambda_{n_{SK}-1}} \quad (2)$$

Where the n_{SK} is the number of non-H atoms and $\lambda_{n_{SK}-1}$ is the first non-zero eigenvalue. TI2 is a topological descriptor and belongs to the Mohar indices that are related to solubility of compounds. In general, it is associated with the size, shape, symmetry, as well with branching or cyclicity of the molecule. TI2 shows a positive coefficient value, indicating that by increasing the value of the descriptor, the expected $pMIC$ values will also increase. This descriptor has been used in the explanation of the activity of diaryl urea derivatives [61] and in the QSAR analysis of aminomethyl-piperidones [62].

The GATS6m [63,64] descriptor belongs to the 2D autocorrelation indices where the Geary coefficient is a distance-type function that can be any physicochemical property (w), calculated for each atom, such as atomic mass, polarizability, or volume, among others, and is represented by (3). By summing the products of a certain property of two atoms located at a certain distance or spatial lag (k), a spatial autocorrelation can be obtained.

$$GATS(k, w) = \frac{\left(\frac{1}{2\Delta k}\right) \cdot \sum_{i=1}^A \sum_{j=1}^A (w_i - w_j)^2 \cdot \delta(d_{ij}; k)}{\left(\frac{1}{A-1}\right) \cdot \sum_{i=1}^A (w_i - \bar{w})^2} \quad (3)$$

Where A is the number of non-hydrogen atoms, \bar{w} is the average of the w_i atomic property value, $\delta(d_{ij}; k)$ is the Kronecker delta, and Δk is the number of vertex pairs at distance equal to k . GATS6m is the mean Geary autocorrelation of lag 6/weighted by atomic mass, which means that this descriptor considers the atomic mass of any atom in the structure at different path lengths (lag) of 6. Strong spatial autocorrelation between pair of atoms produces low values of this index. Also, symmetric, or low-branched structures as well as molecules with low number of heteroatoms (atoms besides C and H) are expected to produce low to zero values. The GATS6m descriptor displays a negative coefficient in (1), which indicates that by increasing the autocorrelation between pairs of atoms considering their atomic masses at a distance of 6 between them, the value of this descriptor will increase, causing a reduction in its $pMIC$ value. As seen from Figure 8a, there is a homogeneous distribution of the data when plotting the GATS6m descriptor against the corresponding $pMIC$ values. Eight molecules from the dataset have a zero

value of GATS6m; their structures are displayed in Figure 8b. Furthermore, these molecules are seen to have medium-interval of *p*MIC (between 3.5 to 5) relative to their location in the scatterplot. In Figure 8c, for the molecule with the highest GATS6m-value, selected pathways are shown for which the sum of their atomic masses produces the final value.

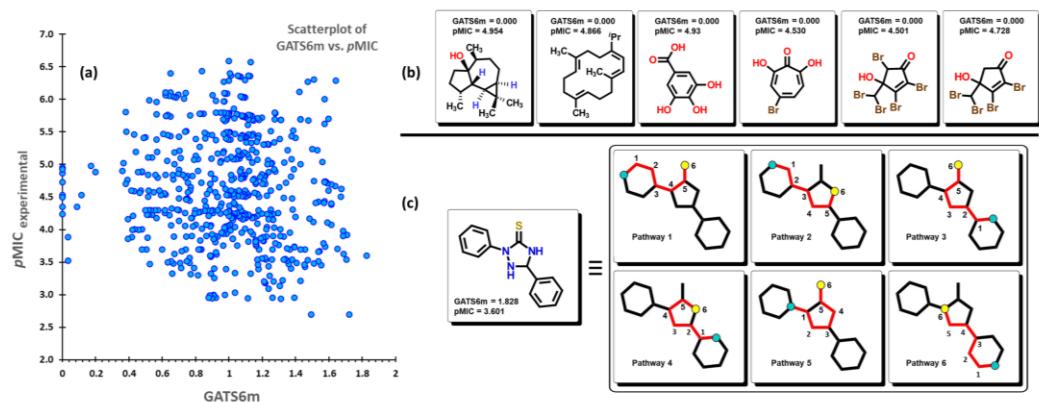


Figure 8. (a) Scatterplot of the GATS6m descriptor vs. the experimental *p*MIC value of the 592 molecules. (b) Molecular structures of compounds with zero value of GATS6m. (c) Selected pathways used for the calculation of the descriptor.

The next six descriptors belong to the functional-group counts (FGC), which are considered as indicator variables. Their value will depend on the number of functional groups present or absent from the molecule, meaning that not all compounds will feature them. The FGC have been used to identify structural features that are important for a property of particular interest. Therefore, their presence or absence can significantly alter the predicted activity in the model. Each FGC descriptor can be easily understood in terms of the nature of functional groups. For example, nArCOOH, nRCONH2, nROR, and nImidazoles accounts for the number of aromatic carboxylic acids, the number of aliphatic primary amides, the number of aliphatic ethers and the number of imidazole moieties, respectively (Figure 9).

The nHDon indicates the number of hydrogen donor atoms (-NH₂ and -OH) for which the formation of hydrogen bonds is possible; in the same manner, nHBonds accounts for the number of intramolecular hydrogen bonds that are possible when there are acceptor atoms like N, O, or F, as shown in Figure 10. Intramolecular hydrogen bonds are crucial for the biological activity of many compounds. It is well established that intramolecular hydrogen bond formation can lead to temporarily closed ring systems which are more lipophilic in nature, while open forms are exposed to solvent, lending more hydrophilic character to the molecule [65]. For example, small hydrophilic molecules, such as β -lactams, use the pore-forming porins to enter cytoplasm/periplasm [66], while hydrophobic drugs like macrolides diffuse across the lipid bilayer [67]. In our model, the nHDon descriptor displays a positive value, indicating that a high number of donor atoms leads to an increase in the biological activity. However, as the nHBonds descriptor possesses a negative coefficient, it indicates that as the number of intramolecular hydrogen bonds increases, in part due to a high number of hydrogen donor atoms, the biological activity decreases, which is correlated to a more lipophilic nature of the molecules.

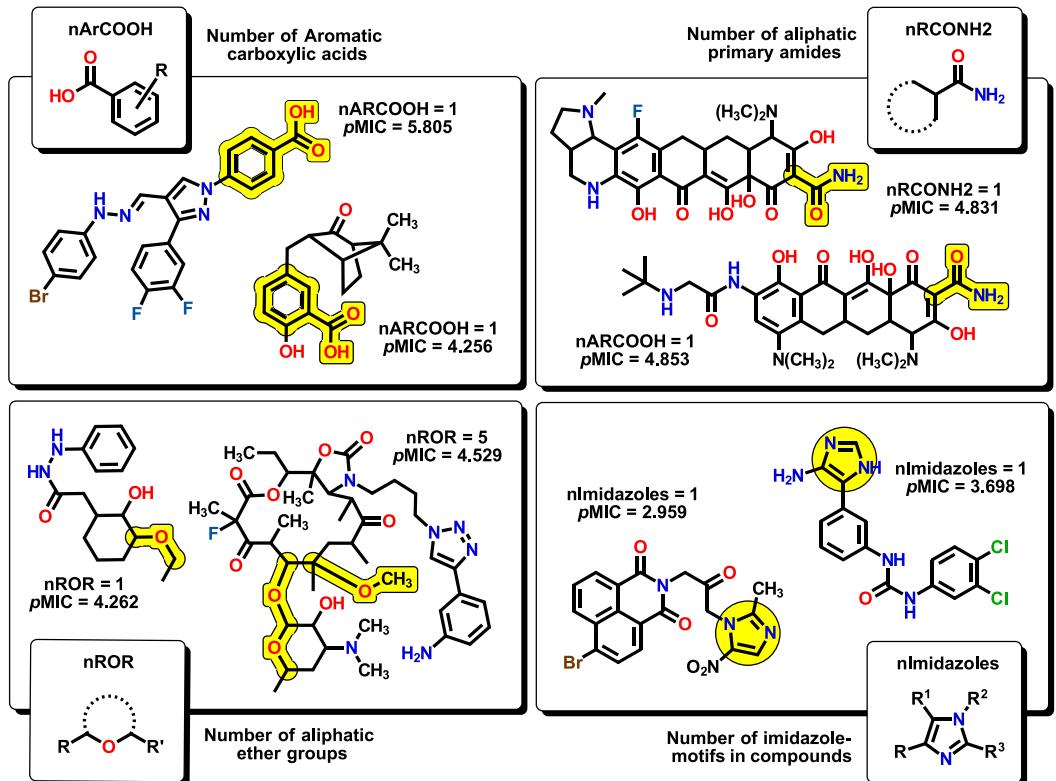


Figure 9. Functional-group count (FGC) descriptors with some representative molecules for each nArCOOH, nRCONH2, nROR, and nImidazoles. The corresponding functional groups are highlighted in yellow.

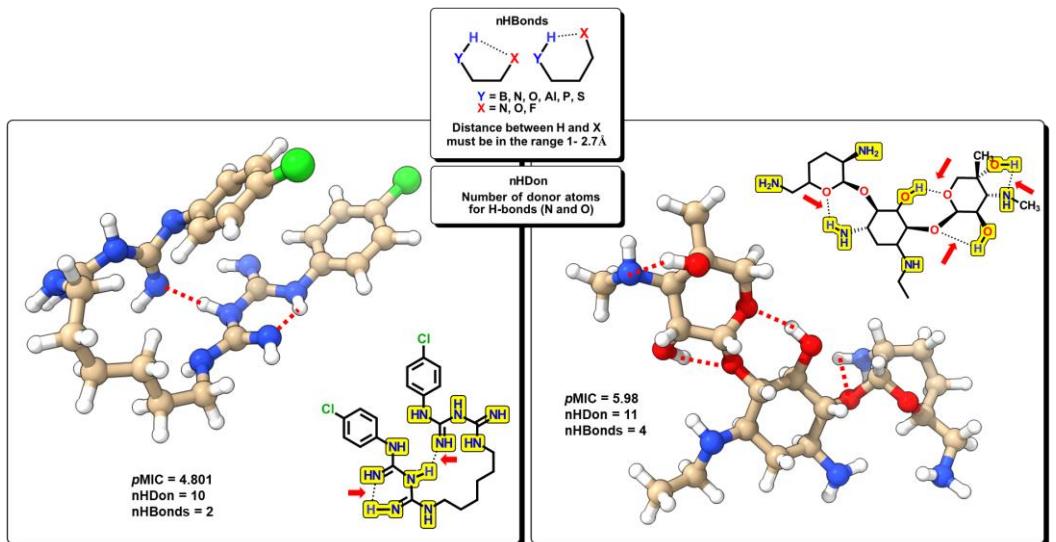


Figure 10. Functional-group count (FGC) descriptors with some representative molecules for nHDon and nHBonds. Highlighted in yellow are groups for which hydrogen donor atoms are counted (nHDon). Red arrows indicate the groups where intramolecular hydrogen bonds are possible (nHBonds).

Six atom-centred fragment (ACF) descriptors are present. ACF descriptors are based on structural fragments which contain the information of the central atom and their bonding neighbors [68–70]. Each ACF is defined by the type of bonding, as well as the number and nature of the neighbors bounded to the centered atom. For example, C018 (=CHX) corresponds to a sp^2 C atom which is single-bonded to a hydrogen and to any electronegative atom (such as N, O, S, etc). The C029 (R–CX–X) descriptor, for which the “–” represents an aromatic bond (e.g., benzene) or delocalized bonds (as in the N–O bond in a NO_2

group), corresponds to a central sp^2 C atom that is single-bonded to an electronegative X atom, and also both double-bonded to a X atom and a R group, in which their bonds are delocalized. The C032 (X--CX--X) descriptor behaves in a similar fashion to C029, but instead of a R group it is replaced by a third X atom. This descriptor has been used also for the analysis of chemical features essential for antiviral activity [71]. The H051 descriptor stands for the environment in which a hydrogen atom is bonded. It is defined as a hydrogen that is attached to an alpha-C atom; an alpha-C may be defined as a carbon connected through a single bond with $-C=X$ (double bond), $-C\equiv X$ (triple bond), or $-C-X$ (aromatic bond), where X represents any electronegative atom, like in the case of alpha-hydrogens in carbonyl compounds. This descriptor has been used to explain the activity of a series of molecules containing nitroaromatics motifs as radiosensitizers [72]. The next two descriptors, N075 and N079, are nitrogen based structural fragments. The first one is defined as a central sp^2 N atom which is bonded to two R groups or to one R and X groups (R--N--R or R--N--X), like in pyridine type motifs. This descriptor is particularly important as many molecules in our set present these kinds of motifs. The second descriptor is related to any nitrogen atom which bears a positive charge. Representative examples for each of the ACF descriptors are presented in Figure 11.

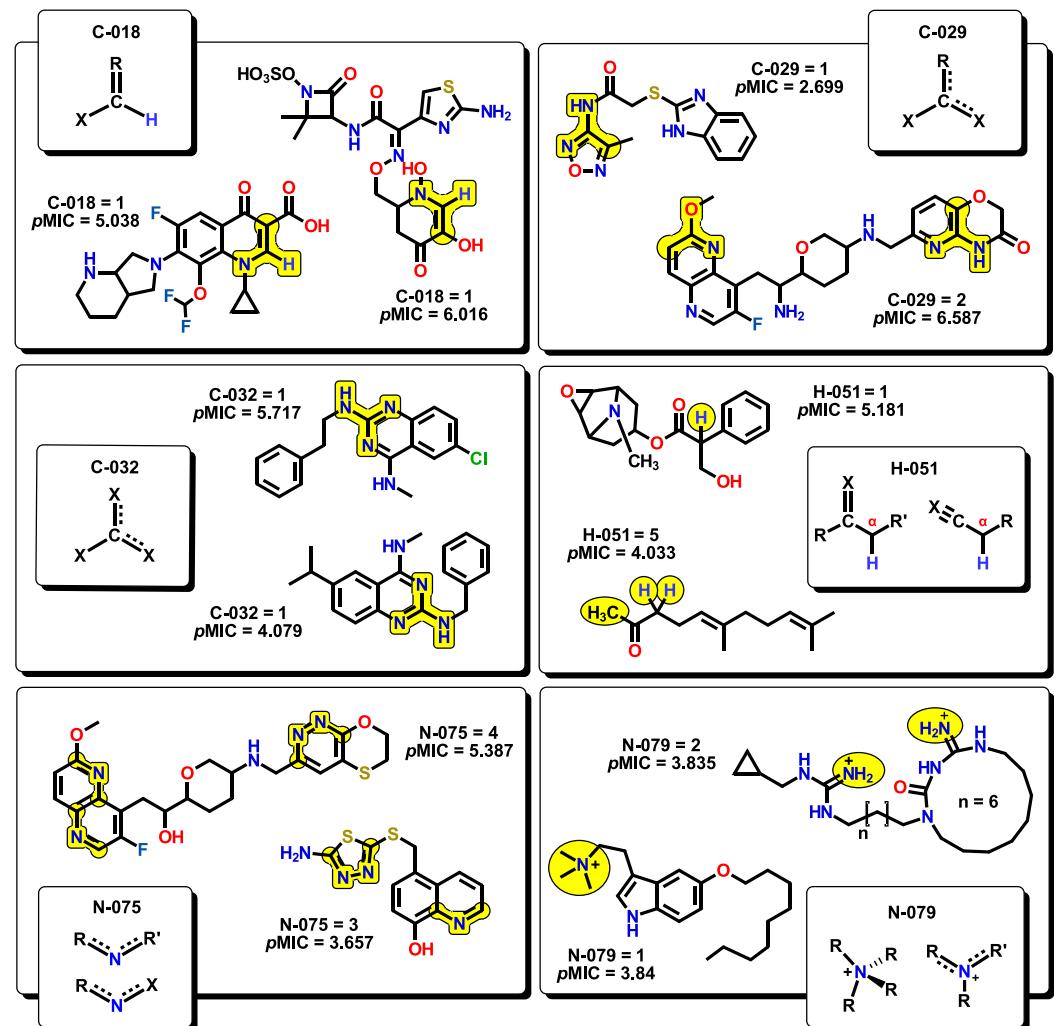


Figure 11. Atom-centred fragments (ACF) descriptors with representative molecules which incorporate them within their structures.

From a general view, descriptors in Equation (1) can be classified into global and indicator variables. Global terms like GATS6m and TI2 are present in the molecule and give information of the whole structure, while indicator variables only appear if the molecular structure contains the motif. Furthermore, descriptors can be associated to the

steric and electronic properties of the molecule (D/Dr06, GATS6m, and nImidazoles, as well as the six-ACF descriptors), while others are more related to solubility of compounds, like in the case of nHDon, nHBonds, TI2, as well as functional groups like nArCOOH, nRCONH2, and nROR. Electronic parameters can be associated to atom-centred fragments which indicates the distribution of substituents around a specific atom. As many molecules include within their structure specific ACF moieties, their inclusion will lead to an increase or decrease in the predicted *p*MIC value. For example, the three ACF based on central carbon (C-018, C-029 and C-032) are positive in their signs indicating that their presence enhance bioactivity. Furthermore, as they are carbon ACF descriptors, they can be associated with core-structure features. However, H-051, N-075, and N-079 ACF descriptors lead to a decrease in the activity. H-051 recall hydrogen atoms which are reactive and hence, they are prone to be abstracted by the use of bases. Nitrogen atoms like those described by the N-075 descriptor are good hydrogen bond acceptors, leading to the generation of inter- and intramolecular interactions by the use of their lone pairs of electrons, which decreases the solubility of molecules, as stated by the nHBonds descriptor. On the other hand, molecules which are well solvated in aqueous media are expected to be high in *p*MIC values.

Figure 12 shows the percentage of distribution of the descriptors for molecules in the dataset. 95.9% of the molecules (568) have the nHDon functional group and almost all other descriptors fall within this category. The second major descriptor that appears in the dataset is nHBonds with 53.9% of the molecules (319), followed by the N-075 descriptor is in 278 molecules of the subset (47%). Considering the high number of bioactive compounds which includes a pyridine-fused or pyridine-containing heterocycles, as well as their tendency to participate in hydrogen bonding, the presence of this descriptor in great percentage is important to account the description of the activity of molecules [73,74].

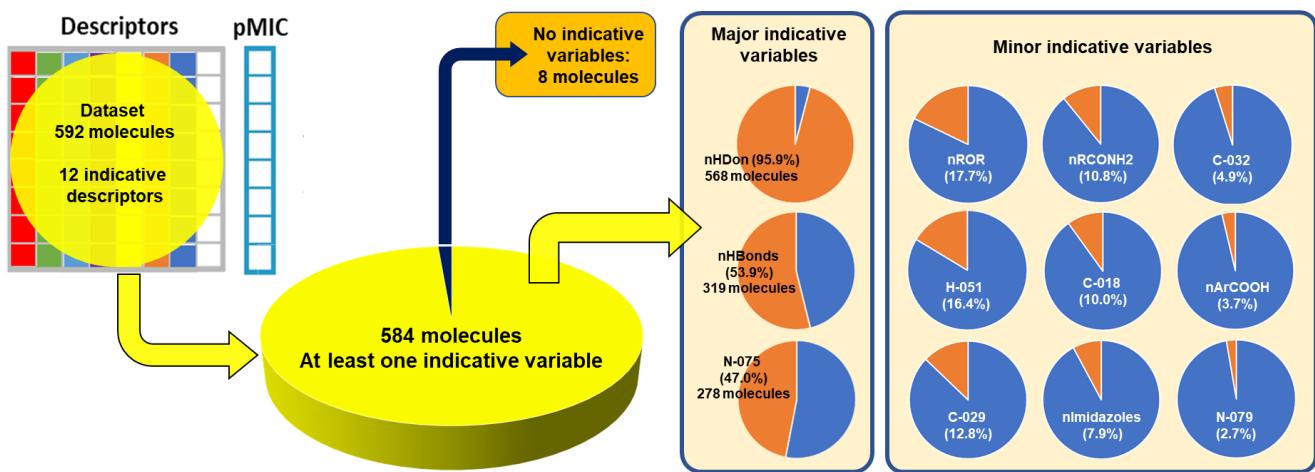


Figure 12. Local distribution of descriptors in molecules used in the model.

Some molecules are observed to be outside the boundaries of the nHDon/nHBonds descriptors, which agrees with the presence of compounds without donor groups such as hydroxyls (-OH), or amines (-NH₂), like in natural products. The rest of the molecules are located into these major categories, which can be seen adequately in the Venn diagram [75-80] in Figure 13. It is also seen that eight molecules lack of the rest of molecular descriptors used in the model. Thus, they are depicted outside the Venn diagram as a sole group.

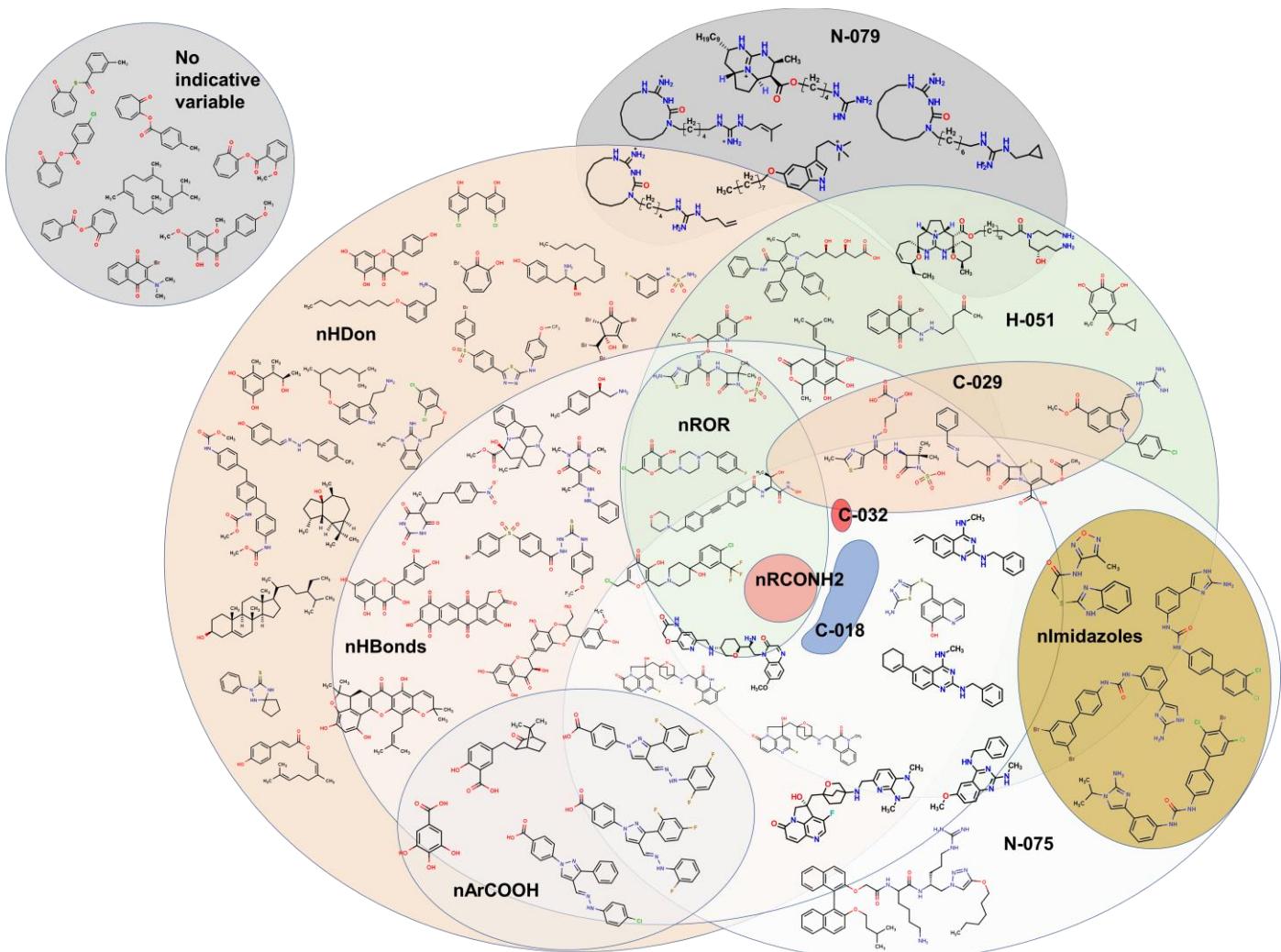


Figure 13. Venn diagram showing representative molecules from the dataset classified by the presence of, at least, one molecular descriptor and their correlations. Molecular descriptors used for Venn diagram are nHDOn, nRBonds, nArCOOH, nRCONH2, nROR, nImidazoles, C018, C029, C032, H051, N075, N079.

2.3. Virtual screening using BIOFACQUIM dataset

Once we fully validated our model for antibacterial activity against *A. baumannii*, we proceeded to search for new molecular candidates in an online database of molecular compounds. BIOFACQUIM [81] is a Mexican natural product database which comprises 528 compounds isolated from many plants and other organisms from Mexico. After careful curation of the database and calculation of their descriptors (Table S5), we performed the analysis of the molecules using our QSPR model. The predicted *pMIC* values from molecules of the database range between 1.65 and 11.24. Table 1 shows these values for the most active molecules, suggested by our model, and depicted for some structures in Figure 14. As stated in Equation 1, a high value of the calculated *pMIC* implies a small concentration of the compound, which correlates to an increase in its potency. In this sense, desirable molecules should exhibit high *pMIC* values.

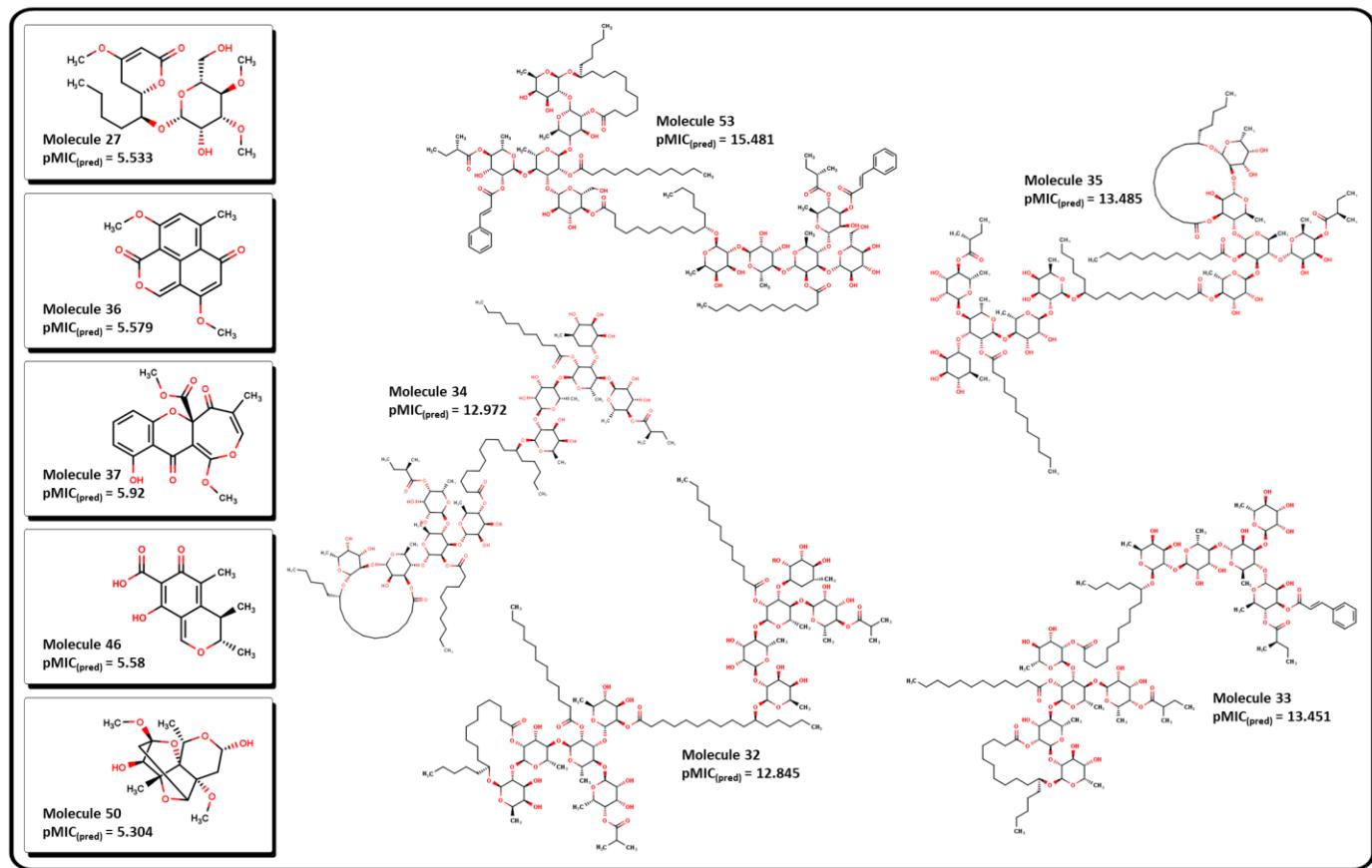


Figure 14. Molecular structure of selected natural products. Molecules 32 to 35 and 53 exhibit the highest predicted *p*MIC value. Molecules 27, 36, 37, 46, and 50 show low predicted values according to Equation 12.

Table 1. Molecular descriptors values for natural products 26 to 53 with their predicted *p*MIC value. These molecules show the highest values from the BIOFACQUIM database.

| MolID | MW | D/Dr06 | GATS6m | nROR | nHDon | nHBonds | C018 | H051 | N075 | TI2 | pMIC |
|-------|---------|----------|--------|------|-------|---------|------|------|------|--------|--------|
| 26 | 274.24 | 101.297 | 0.399 | 1 | 1 | 0 | 1 | 0 | 0 | 1.208 | 5.910 |
| 27 | 404.51 | 208.41 | 1.05 | 5 | 2 | 2 | 0 | 0 | 0 | 3.678 | 5.533 |
| 28 | 1167.41 | 1692.143 | 0.951 | 12 | 16 | 8 | 0 | 2 | 0 | 8.304 | 9.049 |
| 29 | 1195.47 | 1739.777 | 0.958 | 12 | 16 | 8 | 0 | 2 | 0 | 8.397 | 9.076 |
| 30 | 1341.63 | 2360.184 | 0.965 | 13 | 19 | 10 | 0 | 2 | 0 | 10.723 | 9.756 |
| 31 | 1690.16 | 2257.102 | 1.01 | 14 | 13 | 8 | 0 | 9 | 0 | 6.18 | 8.879 |
| 32 | 2473.43 | 6328.172 | 1.071 | 19 | 16 | 11 | 0 | 10 | 0 | 20.998 | 12.845 |
| 33 | 2449.3 | 7015.971 | 1.082 | 19 | 16 | 11 | 0 | 8 | 0 | 21.234 | 13.451 |
| 34 | 2445.37 | 6531.924 | 1.071 | 19 | 16 | 11 | 0 | 10 | 0 | 21.416 | 12.972 |
| 35 | 2501.49 | 6713.547 | 1.076 | 19 | 16 | 9 | 0 | 10 | 0 | 21.566 | 13.485 |
| 36 | 272.27 | 101.297 | 0.894 | 1 | 0 | 0 | 1 | 0 | 0 | 1.208 | 5.579 |
| 37 | 346.31 | 89.966 | 1.133 | 2 | 1 | 0 | 1 | 0 | 0 | 1.397 | 5.920 |
| 38 | 560.71 | 440.654 | 0.966 | 4 | 8 | 4 | 0 | 0 | 0 | 5.474 | 5.695 |
| 39 | 250.27 | 80.687 | 0.942 | 1 | 2 | 1 | 1 | 0 | 0 | 1.546 | 5.580 |
| 40 | 1151.41 | 1669.117 | 0.921 | 11 | 16 | 9 | 0 | 2 | 0 | 8.266 | 8.511 |
| 41 | 1179.47 | 1715.713 | 0.923 | 11 | 16 | 9 | 0 | 2 | 0 | 8.347 | 8.539 |
| 42 | 869.18 | 889.623 | 0.957 | 7 | 10 | 5 | 0 | 2 | 0 | 8.006 | 6.834 |
| 43 | 1035.28 | 1396.119 | 0.967 | 10 | 14 | 8 | 0 | 2 | 0 | 8.601 | 7.984 |
| 44 | 1165.44 | 1695.422 | 0.921 | 11 | 15 | 10 | 0 | 2 | 0 | 8.575 | 8.216 |
| 45 | 1193.5 | 1742.017 | 0.923 | 11 | 15 | 9 | 0 | 2 | 0 | 8.623 | 8.446 |

| | | | | | | | | | | | |
|----|---------|----------|-------|----|----|---|---|----|---|--------|--------|
| 46 | 250.27 | 80.687 | 0.942 | 1 | 2 | 1 | 1 | 0 | 0 | 1.546 | 5.580 |
| 47 | 1199.65 | 1656.598 | 1.024 | 10 | 10 | 7 | 0 | 4 | 0 | 8.71 | 7.644 |
| 48 | 1223.67 | 1189.333 | 1.054 | 10 | 8 | 6 | 0 | 5 | 0 | 5.649 | 7.156 |
| 49 | 512.56 | 409.063 | 0.805 | 3 | 6 | 2 | 1 | 2 | 1 | 4.621 | 6.430 |
| 50 | 302.36 | 106.192 | 1.753 | 5 | 2 | 1 | 0 | 0 | 0 | 1.027 | 5.304 |
| 51 | 1369.82 | 1718.197 | 1.079 | 10 | 8 | 6 | 0 | 5 | 0 | 6.149 | 7.460 |
| 52 | 1383.85 | 1778.609 | 1.097 | 10 | 8 | 6 | 0 | 6 | 0 | 6.324 | 7.386 |
| 53 | 2795.76 | 9251.423 | 1.099 | 20 | 16 | 8 | 0 | 10 | 0 | 22.254 | 15.481 |

Table 1 shows that molecules with the highest predicted *p*MIC values are compounds **32** to **35** and **53**, which were isolated from several plants of the genus *Ipomoea* [82,83]. Their molecular structures contain several functional groups that contribute to their predicted activity. Three important features are observed: 1) all of them have a high number of pyranose-like rings, which may contribute to their hydrophilicity properties; 2) most of them contain large aliphatic side chains and/or macrocyclic lactone rings, which may contribute to their lipophilicity; 3) all of them present at least one terminal ester group which may be prone to cleavage by hydrolysis in aqueous media. Molecules **59** to **62** and **64** also exhibit terminal carboxylic acid fragments.

Analyzing these characteristics in our model we can obtain some insights regarding the structural information that correlates to the predicted values. For example, all the molecules exhibit a great number of aliphatic ether groups and, according to our model in Equation 12, as the number of aliphatic ether motifs (nROR) increases, the greater their activity will be. This is highly correlated to the large number of donor atoms (oxygens) and therefore, as the number of nHDon increases, so does the predicted bioactivity. Nonetheless, a great number of donor atoms also increases the possible number of intramolecular hydrogen bonds (nHBonds) which, according to our model, diminishes the predicted values. Another descriptor that appears to affect the predicted values is H-051 that implies the presence of hydrogens attached to alpha-carbon atoms, known as alpha-hydrogens (α -H). As the number of α -H increases, the bioactivity tends to decrease. In those molecules that are predicted with the highest *p*MIC values, ester and carboxylic acid groups appear in great numbers, suggesting that this kind of functional groups are not adequate for their pharmacokinetic profile, as all of them exhibit α -H. Another feature is the presence of a high number of pyranose-like rings which are 6-membered rings, thus, the high D/Dr06 value displayed. Furthermore, because of their structure, these molecules are highly branched which is seen in their high TI2 values. GATS6m is complex in nature but correlates well with the molecules under analysis. As the average number of possible 6-pathways for which heavy atoms can be included, there is a decrease in the predicted bioactivity. There are many known bioactive compounds for which their molecular masses are substantially high, for example the macrolides and some other natural products like digitoxin [21,22,84,85], thus violating one of the Lipinski's rules used for the evaluation of possible new drugs [86]. Molecules **31**, **48** and **53** present relative medium high GATS6m values, hence high molecular mass; however, our model predicts elevated *p*MIC for these compounds. This can suggest that there could be a limit in the mass of the molecule and the number of oxygen atoms or any other heavy element that will cause molecules to be less active.

2.4. Antibacterial activity evaluation

Having identified molecular properties with potential high activity against *A. baumannii* from plants, we searched for similar molecules from the same genus of plants *Ipomoea*. Several isolated molecules from plants of the species *I. stans*, *I. purga*, *I. muricoides*, and *I. tyrianthina* [87–89] were subject to treatment with our model to obtain their predicted values prior to experimental work. Results are shown in Table 2 and Figure 15 depicts molecules with the highest and lowest predicted *p*MIC values.

Table 2. Molecular descriptors for natural products **54** to **68** with their predicted pMIC value.

| MolID | MW | D/Dr06 | GATS6m | nROR | nHDon | nHBonds | H051 | TI2 | pMIC |
|-----------|---------|----------|--------|------|-------|---------|------|-------|-------|
| 54 | 1139.49 | 637.544 | 1.05 | 8 | 8 | 4 | 5 | 4.452 | 6.553 |
| 55 | 1095.43 | 603.051 | 1.06 | 8 | 7 | 4 | 5 | 4.134 | 6.406 |
| 56 | 1155.49 | 645.374 | 1.047 | 8 | 9 | 4 | 5 | 4.4 | 6.672 |
| 57 | 1225.64 | 1192.314 | 1.016 | 10 | 9 | 5 | 7 | 5.92 | 7.289 |
| 58 | 1107.49 | 600.453 | 1.072 | 8 | 6 | 5 | 5 | 4.095 | 6.080 |
| 59 | 1153.38 | 1669.117 | 0.957 | 12 | 16 | 11 | 2 | 8.266 | 8.422 |
| 60 | 334.46 | 90.65 | 0.904 | 2 | 4 | 1 | 2 | 4.886 | 4.781 |
| 61 | 326.33 | 164.888 | 1.054 | 1 | 5 | 1 | 0 | 4.222 | 4.733 |
| 62 | 342.33 | 166.979 | 1.09 | 1 | 6 | 2 | 0 | 3.98 | 4.624 |
| 63 | 1646.15 | 2393.949 | 1.019 | 13 | 14 | 12 | 6 | 5.644 | 8.223 |
| 64 | 1019.28 | 1307.203 | 0.962 | 10 | 13 | 6 | 2 | 9.197 | 8.242 |
| 65 | 855.1 | 437.562 | 0.955 | 8 | 9 | 5 | 2 | 3.704 | 6.688 |
| 66 | 1093.46 | 601.959 | 1.069 | 8 | 6 | 4 | 5 | 4.081 | 6.285 |
| 67 | 1037.39 | 565.718 | 1.019 | 8 | 7 | 4 | 5 | 4.448 | 6.410 |
| 68 | 1123.49 | 624.213 | 1.054 | 8 | 7 | 4 | 5 | 4.266 | 6.424 |

Compounds **59**, **63**, and **64** exhibit the highest predicted values of *p*MIC of 8.422, 8.223 and 8.242. From our QSPR model, we can observe some important features which are present in these compounds. First, molecules from **54** to **68** present many aliphatic ether groups from the pyranose-type rings, consequently a great number of hydrogen donors (nHDon descriptor), which contributes to an increase in its antibacterial activity. As the pyranose-type rings are six-membered structures, the D/Dr06 descriptor also promotes a rise in the expected *p*MIC. However, because of the large number of oxygen atoms and carbonyl motifs, the nHBonds and H-051 descriptors have a considerable effect in decreasing the predicted *p*MIC. Furthermore, the value of the calculated GATS6m, compared to other molecules, implies a small negative contribution to the predicted activity which is balanced by the contribution of the D/Dr06 descriptor. Compounds **60**, **61** and **62** (Figure 15) are predicted to have the lowest *p*MIC values (4.781, 4.733 and 4.264, respectively). This situation is due to the presence of only one pyranose-type ring in each structure, hence, only one aliphatic ether group and a reduced number of oxygen atoms. Moreover, given their molecular structure (low symmetry), their GAST6m values are also the highest among the compounds, thus diminishing the predicted value. Consequently, it is expected that a great number of pyranose-type rings which do not form intramolecular hydrogen bonds are valuable for the antibacterial activity of these compounds.

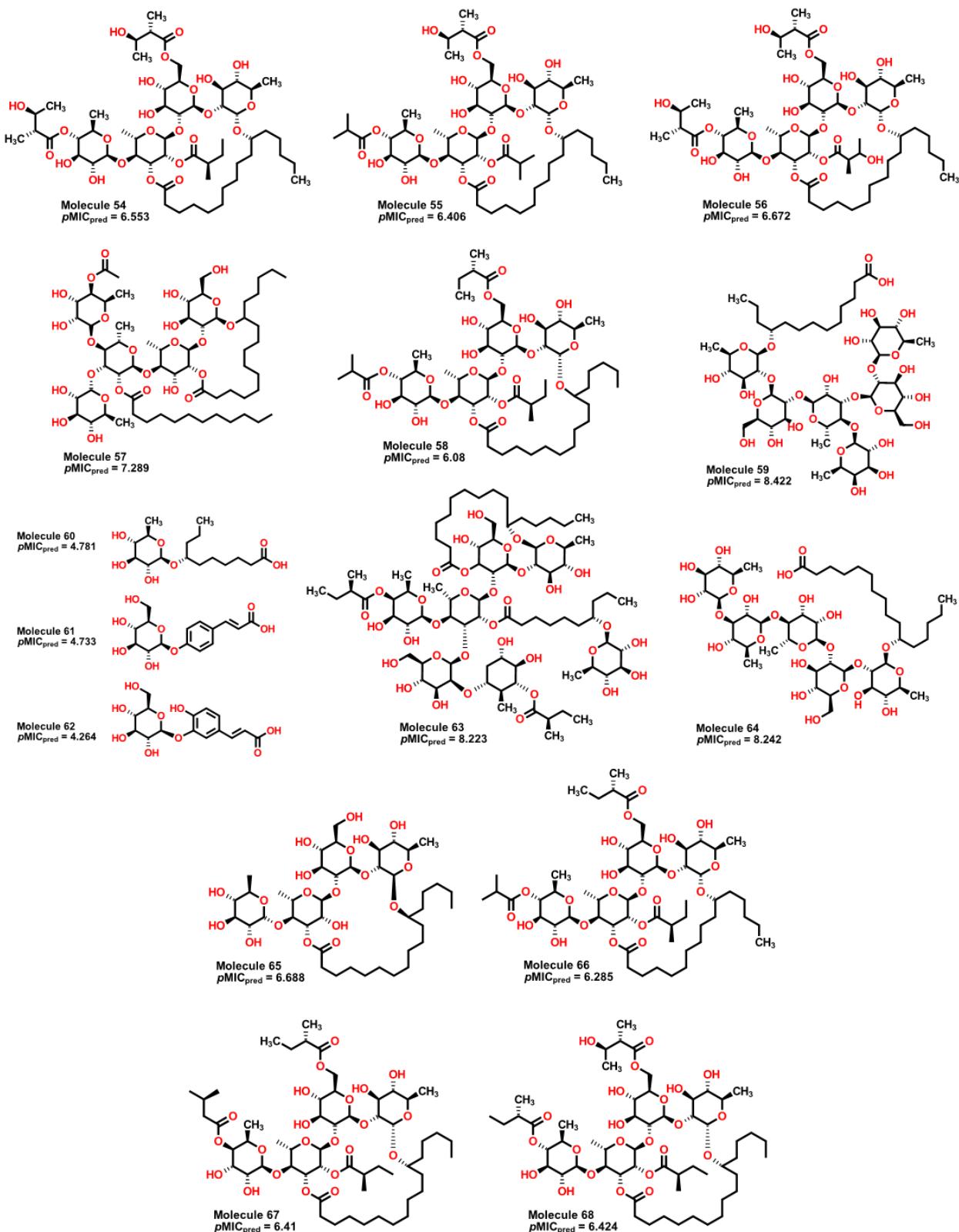


Figure 15. Molecular structure of natural products isolated from different species of *Ipomoea*. Molecules 59, 63, and 64 display high values of predicted $p\text{MIC}$, while compounds 60 to 62 show the lowest values.

2.5. Glycoside SAR analysis

As stated before, the increasing number of multidrug-resistant bacteria represents an important risk to human health worldwide. Although *A. baumannii* represents a serious threat, the search for wide-spectrum antibiotics for the treatment of infections caused by several of the ESKAPE pathogens is crucial. To determine if the proposed molecules display antibacterial activity towards this bacterial critical group, the corresponding bioassays were tested using clinical isolates which are metallo- β -lactamase producers and resistant to betalactam antibiotics (Table 3).

Table 3. Biological essays for compounds isolated from *Ipomoea* sp. towards different bacterial strains. For each bacterial strain, essays marked as (+) were positive and (-) negative in susceptibility tests

| ID sample | Bacterial strains | | | | | | | | |
|--------------|---------------------------------|--------------------------|---------------------------------|------------------------------|-------------------------|------------------------------|--------------------------------|------------------------------|------------------------------|
| | <i>E. coli</i> ATCC 25922 | <i>S. aureus</i> ATCC | <i>A. baumannii</i> 9736 (1) | <i>A. baumannii</i> 10324 | <i>E. coli</i> 10225 | <i>K. pneumoniae</i> 6411 | <i>K. pneumoniae</i> 3407-2 | <i>P. aeruginosa</i> 4899 | <i>P. aeruginosa</i> 4677 |
| | - | - | - | - | - | + | - | - | - |
| 54 | - | - | - | - | - | + | - | - | - |
| 55 | - | + | - | - | - | + | - | + | + |
| 56 | + | + | + | - | - | + | + | + | + |
| 58 | - | - | - | - | - | - | + | - | - |
| 60 | + | + | + | + | + | + | + | + | + |
| 61 | - | - | - | + | - | - | - | - | - |
| 62 | - | - | - | + | - | - | - | - | - |
| 63 | - | - | + | + | - | - | - | - | - |
| 64 | - | - | + | - | - | - | - | - | - |

From the results, important features arise from the molecular structures of the glycosides. First, molecular structures of compounds 54, 55, 56 and 58 contain the same tetrasaccharide core which is connected by a macrolactone ring. From 54 to 55, removal of one carbon atom, from central 2-methylbutyrate to 2-methylpropionate, increases the activity of the glycoside, being active not only to *K. pneumonia* but also now to *P. aeruginosa* and *S. aureus*. In compound 56, reinsertion of the carbon atom but with the addition of a hydroxyl group at position three of the 2-methylbutyrate group reinforces the activity spectrum by being active to *A. baumannii*, as seen in Figure 16. However, removal of the hydroxyl group of the central and outer 2-methylbutyrate groups and addition of one carbon atom of the macrolactone ring (from ten atoms to eleven) causes molecule 58 to lose wide spectrum activity and to be only active against *K. pneumoniae*, this suggest that hydroxyl groups, located in specific regions of this molecular core, enhance the bioactivity of this set of glycosides.

Compounds 60 to 62 are the smallest compounds. They share in common a terminal carboxylic acid alongside a pyranose-ring. Although 60 has wide antibacterial activity against multidrug resistant bacteria, 61 and 62 only display activity against *A. baumannii*. This important loss of activity may be attributed to the removal of the aliphatic chain connecting the pyranose ring and the terminal carboxylic acid, being replaced by a more rigid phenyl core. Close inspection of compound 63 reveals the structure of compound 60 within it forming an ester bond at the terminal carboxylic acid group. This feature could explain the retained activity against *A. baumannii*. Similar to this, molecules 54 to 56 share common structural features, like at the macrolactone ring with the same set of atoms, the lack of hydroxyl groups at the outer methylbutyrate may affect the expected activity.

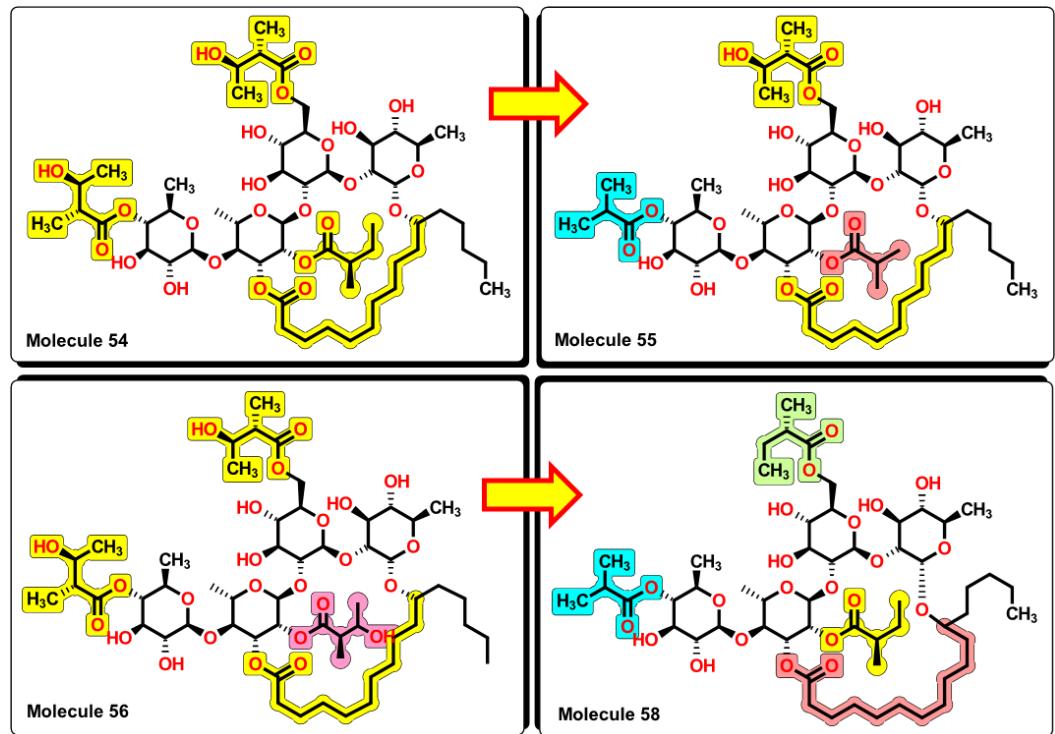


Figure 16. SAR analysis of compounds **54**, **55**, **56** and **58**. Small changes in the structure expand the antibacterial activity from **54** to **56**. Removal of -OH groups and elongation of the alkyl chain in the macrocyclic ring decrease the bioactivity of the molecule.

An insight into the chemical structures of **32** to **35** and **53**, the most potent molecules according to model in Equation 12, reveals that the core of **60** is present (Figure 17). Furthermore, the macrolactone ring alongside the chiral carbon is also a common feature, with the cycle formed of ten or eleven methylene groups as in **54** to **58**. This could suggest that molecules of the BIOFACQUIM database would also exert antibacterial activity towards *A. baumannii* and other resistant bacteria.

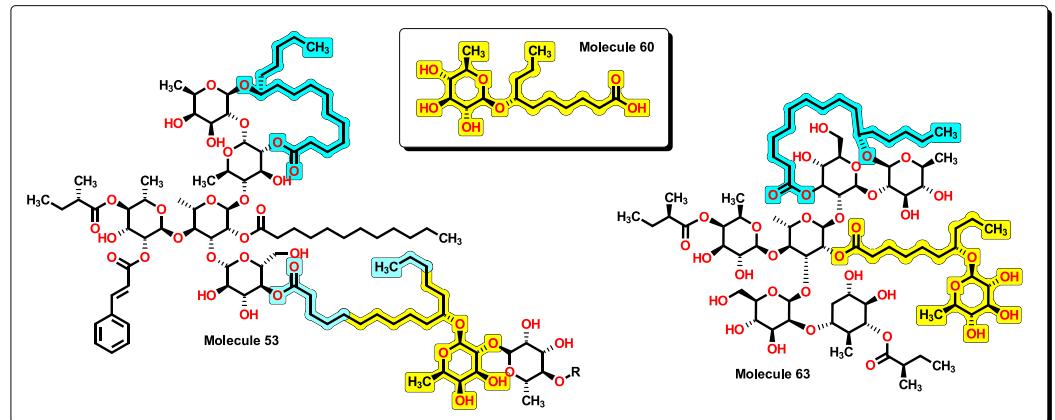


Figure 17. Partial chemical structure of **53** and complete molecule **63**. Chemical core of **60** is displayed within the other structures. Furthermore, the macrolactone ring is shared between compounds.

One of our remaining questions is which action mechanisms can exert these molecules? In order to propose one, we constructed a simplified version of the Venn diagram in which is possible to observe the correlation between the H-051 and the nROR descriptors seen in the isolated molecules. The purpose of this diagram in Figure 18 is to identify molecules with known action mechanism and with structural similarity (same molecular descriptors) to our compounds. Furthermore, other type of compounds used are also part of the inner set of molecules. These compounds have different structural

motifs when compared to compounds **54** to **68** and, they present different mechanisms of action

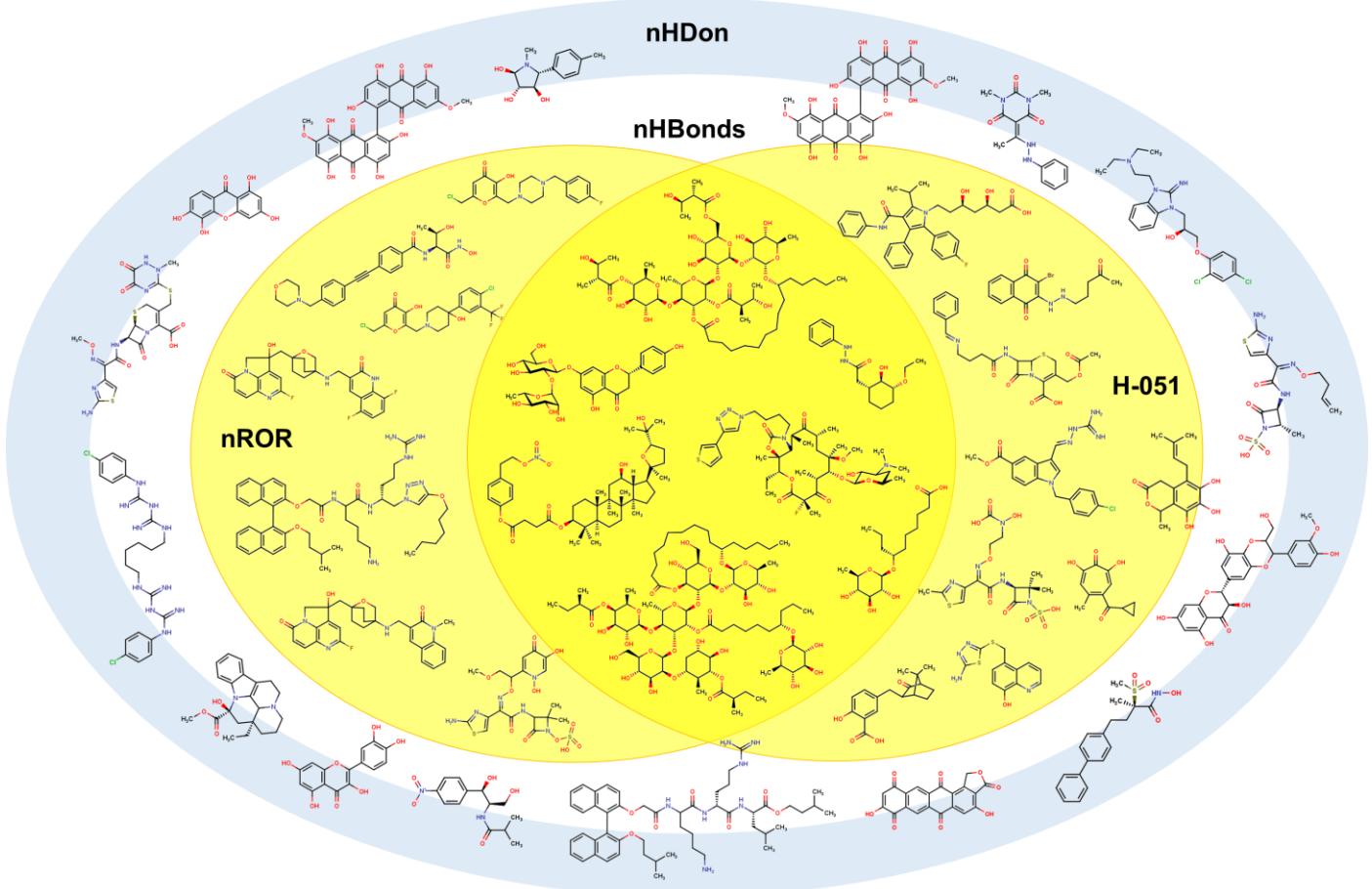


Figure 18. Simplified Venn diagram with representative molecules showing the nHDon, nHBonds, nROR and H-051 descriptors which appear in molecules **54** to **68**.

From a structural point of view, compounds **56** and **63** resemble those of the macrolide antibiotics [85]. Examples of macrolides are Erythromycin A, oleandomycin, josamycin, and spiramycin, isolated from different microorganisms, as well as many semisynthetic derivatives like clarithromycin, flurithromycin, and other unique compounds like azithromycin. Also, the latest new members, the ketolides and fluoroketolides are also structurally related to the macrolide family. As stated above, when comparing the new molecules with macrolides, several features are shared (Figure 19). Macrolides are well characterized by the presence of 14- to 16-membered macrocyclic lactone ring to which one or more deoxy sugars are attached. In the case of compounds like **56** and **63**, the macrocyclic ring is shown connecting two or three sugar-type rings. Furthermore, because of the relative high number of carbonyl motifs in macrolides, α -H are also present in great numbers. This is also true for many compounds from **54** to **58** and **65** to **68**, where the ester group is observed. Moreover, a great number of aliphatic ether groups, and a great number of oxygen atoms present at the hydroxyl groups and other motifs are also features which are in common. Macrolides are potential bacteriostatic compounds, for which one mechanism of action relies on binding to the P site on the 50S subunit of the bacterial ribosome. Because of this, we can suggest that compounds **56** and **64**, among others, could exhibit a similar action on bacteria, thus acting as protein synthesis inhibitors

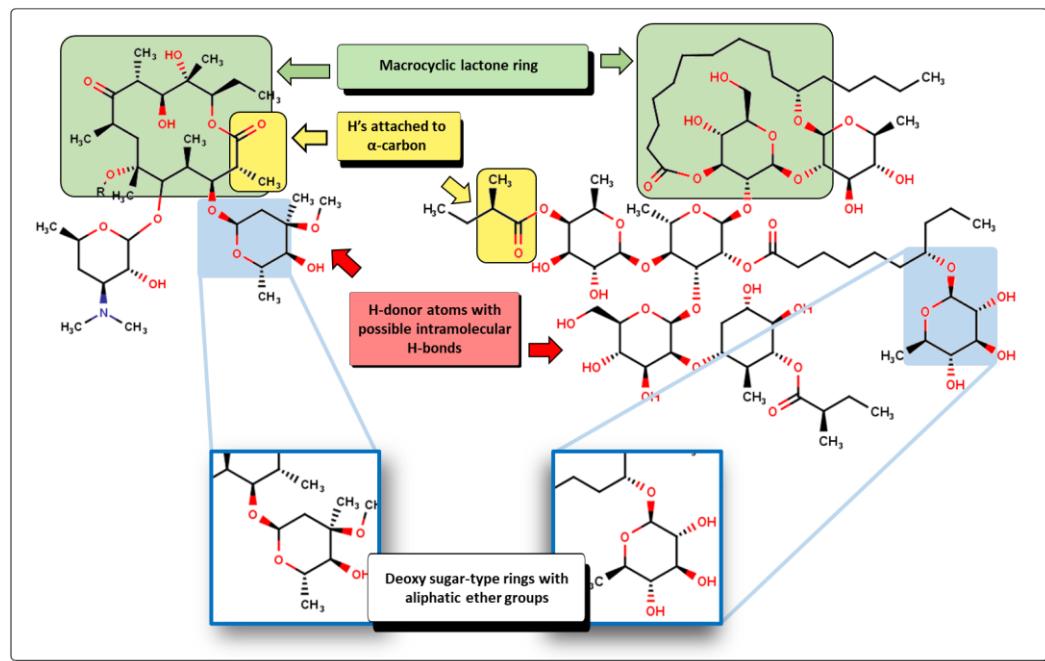


Figure 19. Structural comparison between a macrolide (clarithromycin) and compound 63, where arrows indicate the descriptors that are shared.

Finally, compound **60** as a small molecule can be considered as a lead compound for which specific chemical transformations could improve its efficacy. Moreover, molecules isolated from *Ipomoea* also share within their structures a deoxy-sugar moiety that could be relevant for their activity. By close inspection of the fragment, we search for molecules in the ChEMBL database for bioactive compounds which incorporate the deoxy sugar in their structures. A wide variety of molecules possesses the motif, from anticancer to anti-allergenic [90–98]. Chemical structures for these compounds can be seen in Figure S3.

In summary, the model was validated statistically by internal and external parameters, showing good predictive power. This was demonstrated using the model, first applied to the BIOFACQUIM natural products' database in the search of potential candidates and finally, by exploring the properties of isolated natural products from *Ipomoea* sp. We observed wide antibacterial spectra activity of compounds **56**, **60**, and **63** against several isolated bacterial strains, which agrees with the properties calculated by the model.

3. Materials and methods

3.1. Data set

An initial dataset of 944 compounds was obtained from the literature between 1995 and 2020. These compounds shared the same evaluation method, as follows. To improve the reliability of the data, all the compounds were curated [99–101] to point-out outliers, uncertainties and potential errors that could affect the models generated at later stages, which included: (1) removal of mixtures, salts, and inorganic/organometallic compounds; (2) ring aromatization as well as standardization of the carboxyl, nitro, and sulfonyls groups; (3) deletion of duplicates and exclusion of stereoisomers of the same compound, as 3D-molecular descriptors are not used in this work (see below). After data curation, compounds with undefined MIC values and values greater than 300 $\mu\text{g}/\text{mL}$ were removed [102], leaving a final set of 592 molecules for the generation of the models. Finally, logarithmic transformation of MIC values was achieved to normalize the experimental information; a conversion of MIC values from $\mu\text{g}/\text{mL}$ to molar concentration ($M = \text{mol}/\text{L}$) was done, followed by a transformation to $p\text{MIC}$ according to:

$$p\text{MIC} = -\log_{10}[\text{MIC}] \quad (4)$$

3.2. Calculation of molecular descriptors

The structures of the molecules of interest were drawn in Avogadro [103,104] and MarvinSketch [105] (ChemAxon, Budapest). For the calculation of molecular descriptors, the Dragon [106] computational package was employed. For most of the molecules, their action mechanism is unknown, as in the case of many natural products. Therefore, because molecular conformation is not considered, only zero-, mono- and bi-dimensional descriptors were calculated. The number of descriptors employed per family for the Genetic Algorithms (GA) technique were as follows: 45 constitutional, 105 topological, 33 connectivity indices, 96 2D autocorrelations, 21 topological charge indices, 93 functional groups, 88 atom-centred fragments, and ten molecular properties. A complete list with the molecular descriptors and biological activities reported as Minimum Inhibitory Concentration (MIC) in $\mu\text{g/mL}$ are found in the Supplementary Information (SI, Tables S1 and S2).

3.3. Generation of the mathematical model

The regression models were built using GA techniques with the Mobydigs software [107]. GA are a statistical method that can be employed for analyzing complex systems that correlate with multiple variables. In an analogous manner to genetic evolution, this approximation allows the selection of the most suitable mathematical models from a large set [108]. Molecular descriptors were used as independent variables and the experimental MIC value expressed as $p\text{MIC}$ was used as the dependent variable. The selection of the best model was based on parameters values such as the coefficient of determination (R^2); additionally, the standard deviation (s) and the Fischer test (F) were employed. The Y-scrambling test was used to guarantee that QSPR model was built adequately in terms of correlation obtained by chance. This was performed first by randomly permuting the $p\text{MIC}$ values of the data set and then using the new column of values with the same variables to generate new models. The procedure was repeated 300 times, and the quality parameters of these new models were compared to the original values of the QSPR model: if the original model has no chance correlation, the new R^2 and Q^2 values calculated for the permuted $p\text{MIC}$ QSPR models will have a significant difference with respect of the original values, otherwise, the model is rejected. Non-collinearity between descriptors is determined using the QUIK rule. Accordingly, the QUIK rule is based on the K multivariate correlation index that measures the total correlation of a set of variables as follows:

$$K = \frac{\sum_j \left| \frac{\lambda_j}{\sum_j \lambda_j} - \frac{1}{p} \right|}{\frac{2(p-1)}{p}} \quad (5)$$

where $j = 1, \dots, p$ and $0 \leq K \leq 1$

λ are the eigenvalues obtained from the correlation matrix of the data set $X(n, p)$, n represents the number of compounds and p the number of variables (descriptors). The total correlation in the set given by the model descriptors X plus the response $Y(K_{XY})$ should always be greater than that measured only in the set of descriptors (K_X). In other words, if $K_{XY} - K_X < \delta K$ then the model is rejected. The typical δK threshold values for models are between 0.01 – 0.05. Models that have negative values are not allowed. In order to detect models with an excess of “good” or “bad” descriptors, the redundancy (R^P) and overfitting (R^N) rules were applied. R^P is defined as:

$$R^P = \prod_{j=1}^{p^+} \left(1 - M_j \left(\frac{p}{p-1} \right) \right) \quad (6)$$

with $M_j > 0$ and $0 \leq R^P \leq 1$

While R^N is defined as:

$$R^N = \sum_{j=1}^{p^-} M_j \quad (7)$$

where $M_j < 0$ and $-1 < R^N \leq 0$

Given a regression model with p variables, R_{jy} is the absolute value of the regression coefficient between the j th descriptors and the response Y . In this sense, M_j can be calculated as follows:

$$M_j = \frac{R_{jy}}{R} - \frac{1}{p} \quad \text{and} \quad -\frac{1}{p} \leq M_j \leq \frac{p-1}{p} \quad (8)$$

The redundancy rule establishes that if $R^P < t^P$, then the model is rejected, where depending on the data, t^P , which is a user-defined threshold, can range from 0.01 and 0.1, with a suggested value of 0.05. The overfitting rule specifies that if $R^N < t^N(\varepsilon)$, then the model is rejected. Calculating $t^N(\varepsilon)$ follows:

$$t^N(\varepsilon) = \frac{p\varepsilon - R}{pR} \quad (9)$$

Where values of ε can range from 0.01 to 0.1 and p is the number of variables.

3.4. QSPR validation of prediction capability

The model reported herein was validated internally by the leave-many-out cross-validation method (Q^2_{LMO}) for which the data set was randomly divided into a training set (415 molecules) and a test set (117 molecules) which represented 70% and 30%, respectively of the complete data set. The robustness of the model was further evaluated by bootstrap (Q^2_{BOOT}) and Q^2_{EXT} . The predictive ability validation was performed by applying the Asymptotic Q^2 rule (δQ). It is assumed that a good model should have a small difference between fitting and predictive ability, in which significant variations between the R^2 and Q^2 values can be due to overfitting or to some not predictable samples [109,110]. The Asymptotic Q^2 rule evaluates the asymptotic Q^2 versus the Q^2 values of the model:

$$Q^2_{LMO} - Q^2_{ASYM} < \delta Q \quad (10)$$

If the difference is less than the threshold, typically $\delta Q = -0.005$, then the model is rejected. As Q^2_{LMO} is asymptotically related to the value of R^2 , it is possible to calculate the Q^2_{ASYM} by using the following expression:

$$Q^2_{ASYM} = 1 - (1 - R^2) \left(\frac{n}{n - p'} \right)^2 \quad (11)$$

Where n is the number of objects and p' the number of model parameters. To further evaluate the predictive applicability of the model, some statistical parameters developed by Roy *et al.* were used [45–47]. According to the statistical parameters, the following criteria must be present for each evaluation as shown: (i) $Q^2 > 0.5$; (ii) $r^2 > 0.6$; (iii) $(r^2 - r_0^2)/r^2 < 0.1$ (or $(r^2 - r'_0)^2/r^2 < 0.1$); (iv) $0.85 \leq k \leq 1.15$ (or $0.85 \leq k' \leq 1.15$) and (v) $|r_0^2 - r'^2| < 0.3$. Additionally, two parameters derived from the above, $\overline{r_m^2}$ and $\Delta\overline{r_m^2}$ were also used to evaluate the predictive power of the model [111]. According to, $\overline{r_m^2}$ follows that:

$$\overline{r_m^2} = \frac{(r_m^2 + r'^2_m)}{2} \quad (12)$$

Where $\overline{r_m^2}$ is calculated as:

$$r_m^2 = r^2 \left(1 - \sqrt{r^2 - r_0^2} \right) \quad (13)$$

and where:

$$r'^m^2 = r^2 \left(1 - \sqrt{r^2 - r'_0^2} \right) \quad (14)$$

While according to (15), Δr_m^2 is obtained by the following expression:

$$\Delta r_m^2 = |r_m^2 - r'^m^2| \quad (15)$$

The calculation of the r^2 , r_0^2 , r'_0^2 , k and k' are shown in the SI.

3.5. External validation

The generated model was validated externally by the prediction of different sets of molecules which were not included in the generation of the model with the following specifications: (1) only molecules with reported MIC values and active towards *A. baumannii* were used; and (2) molecules above 300 μ g/mL were excluded. Data curation, as stated above, was performed on a total of 98 molecules, which were drawn in Avogadro and their molecular descriptors obtained from the Dragon software package. A complete list of descriptors and references can be found in the SI, Table S5.

3.6. Virtual screening

528 natural products were obtained from the molecular database BIOFACQUIM. To improve the consistency of the data, all the compounds were curated by (1) ring aromatization, (2) standardization of the carboxyl, nitro and sulfonyls groups if present, and (3) addition of missing bonds where required. The structures of the molecules of interest were drawn in Avogadro and their molecular descriptors were obtained from the Dragon software package. A complete list of descriptors can be found in Table S6.

3.7. Plant material

Roots of *Ipomoea stans* were collected in the state of Puebla, México. Botanical classification was carried out by M. Sc. Abigail Aguilar, Head of the Instituto Mexicano del Seguro Social Herbarium in Mexico City (IMSSM), and a voucher specimen (number 15077) is deposited at IMSSM. Exudates from the bark of *Ipomoea mururoides* were collected manually in the campus of the Universidad Autónoma del Estado de Morelos (UAEM), in Cuernavaca, Morelos, México. The plant material was identified by Biol. Alejandro Flores, and a voucher specimen (No. 22444) was deposited at the Herbarium of the Centro de Investigación en Biodiversidad y Conservación, UAEM. Roots of *Ipomoea purga* were authenticated and donated by M. Sc. Abigail Aguilar, Head of the Instituto Mexicano del Seguro Social Herbarium in Mexico City (IMSSM). A voucher specimen (number 16180) is deposited at IMSSM.

3.8. Extraction and isolation of compounds

The dried, powdered roots of *I. purga* and *I. stans* (250.0 g each one) were extracted by maceration with MeOH (500 ml \times 3), to obtain a dark syrup (25.0 g *I. purge*, and 20.4 g *I. stans*). The dark syrups were extracted with distilled water (3 \times 50 ml) and dichloromethane (DCM, 3 \times 50 ml), to afford a dark solid (9.3 g *I. purga*, and 7.6 g *I. stans*). The dark solids (1.0 g *I. purga*, 1.1 g *I. stans*) were submitted to a C18 column (Supelco, 10 \times 15 mm) with a gradient of MeOH:H₂O (0:100 to 100:0, at increments of 10%), fractions were collected and (0.7 g *I. purga*, 0.6 g *I. stans*) was obtained. The resinous solids were percolated on an activated charcoal column, eluting with MeOH. Fractions of 5 ml were collected and reunited giving the convolvulin (0.42 g *I. purga*, and 0.32 g *I. stans*). Convolvulin of *I. purga* was chromatographed on normal and inverse phase silica gel column, using as mobile phase DCM/MeOH/H₂O (84:14:2) respectively with MeOH gradient, yielding 80 mg IP-JALB (compound 63). From the convolvulin of *I. stans* in the same conditions, 30 mg of ISACAF (compound 62) and 27 mg ISACAR (compound 61) were obtained. Exudates from the bark of *Ipomoea mururoides* (15 g) were dried, ground, and dissolved in MeOH to

give, after filtration and removal of the solvent, a brown solid material (10 g). The brown solid was dissolved in a mixture of CHCl₃:MeOH (9:1). This solution was subsequently subjected to passage over a silica gel column eluted with a gradient system of CHCl₃:MeOH (from 9:1 to 7:3), leading to the separations of two chromatographic fractions. Purification of the less polar chromatographic fraction was carried out by preparative HPLC. Eluates with retention time, t_R, value of 23.5 min were collected and reinjected into the HPLC system to achieve pure IM620 (compound 57). The rest of compounds tested for biological essays were given by Dr. Ismael León, purified by similar methods, and used as received.

3.9. Bacterial strains

Escherichia coli ATCC 25922, and *Staphylococcus aureus* ATCC 29213 were purchased from the American Type Culture Collection. MDR clinical isolates, which are non-susceptible to at least one agent in three or more antimicrobial categories and cause nosocomial infections, were obtained from the Center for Research on Infectious Diseases collection of the National Institute of Public Health (Instituto Nacional de Salud Pública), Cuernavaca, Morelos, Mexico. The various strains include the following isolates: *A. baumannii* 9736 and 10324, *E. coli* 10225, *K. pneumoniae* 6411 and 3407-2 and two *P. aeruginosa* 4899 and 4677. These isolates are metallo-β-lactamase-producers and are resistant to all beta-lactam antibiotics including cephalosporins and carbapenems.

3.10. Antibacterial assays

The antibacterial activity of the compounds was qualitatively measured following the Kirby–Bauer method (1996), according to the CLSI (Clinical and Laboratory Standards Institute) recommendations [112]. Briefly, Petri dishes containing Müller–Hinton agar were sown with bacteria inoculums from 1 to 2×10⁸ colony-forming units (CFU)/ml, and then 3 µl of compound solution was placed over the agar. Incubation time was from 16 to 19 h at 35 ± 2 °C. A halo of growth inhibition was observed as a positive result. Two reference susceptible strains were used: *E. coli* ATCC25922 and *S. aureus* ATCC 29213.

4. Conclusions

There is a great number of compounds that have been biological tested as antibacterial against *A. baumannii*. Nevertheless, a careful selection of them needs to be done before their use for the generation of a QSPR/QSAR model. Our QSPR model comprises fifteen 2D-dimensional descriptors: one 2D-autocorrelation, two topological, six functional group counts, and six atom-centred fragments descriptors. These molecular descriptors were used to describe their suitability as antibacterial compounds against *A. baumannii*. Additionally, our QSPR model prediction ability, which was fully evaluated by means of different test and validation sets of molecules, allowed us the identification of antibacterial compounds against *A. baumannii* by means of a virtual screening of the BIOFACQUIM database, an interesting source for potential bioactive compounds. The identified compounds, isolated from *Ipomoea* sp., indicated specific molecular features consistent with antibacterial activity. Furthermore, our model proved to be predictively reliable by identifying compounds isolated from local collections of *Ipomoea* sp. that showed promising wide antibacterial spectrum. Upon experimental testing, compound 60 showed wide antibacterial activity against clinically isolated multidrug resistant bacteria. Its structure can be found in other compounds also isolated from *Ipomoea*, as in the case of molecule 63. Molecule 60 could serve as a lead compound for the development of new compounds with possible wide spectrum antimicrobial activity.

Supplementary Materials: The following supporting information can be downloaded at: www.mdpi.com/xxx/s1, **Figure S1.** Scatterplots and Williams plots for each training-test experiment; **Figure S2.** Scatterplots for each of the molecular descriptors against experimental pMIC; **Figure S3.** Chemical structures from them ChEMBL database which incorporates the deoxy-sugar moiety found in compound 60; **Table S1.** Complete list of descriptors from Dragon package; **Table S2.** List of molecules used for the generation of the QSPR mode; **Table S3.** Average values for each

molecular descriptor; **Table S4**. Statistical parameters for the evaluation of the predictive power; **Table S5**. List of molecules used for the validation tests sets, their calculated molecular descriptors and predicted *p*MIC values; **Table S6**. List of molecules from the BIOFACQUIM database and their calculated molecular descriptors and predicted *p*MIC values.

Author Contributions: Conceptualization, F.J.P.C, and R.S.R.H.; methodology, F.J.P.C; software, R.S.R.H; validation, F.J.P.C, R.S.R.H. and N.P.; formal analysis, F.J.P.C; investigation, F.J.P.C; resources, I.L.R., J.S.S.; data curation, F.J.P.C; writing—original draft preparation, F.J.P.C; writing—review and editing, R.S.R.H., I.L.R, J.S.S., H.T, and N.P.; visualization, N.P.; supervision, R.S.R.H., H.T., and N.P.; project administration, R.S.R.H.; funding acquisition, R.S.R.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by CONACyT, Mexico, projects number 256927 and 320243.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article and supplementary material.

Acknowledgments: We would like to thank CONACyT for a PhD grant (F.J.P.C.). In addition, R.S.R.H. thanks Dr. Manuel Villanueva of the Universidad de Guanajuato for the software facilities. Authors acknowledge to Alejandro Sanchez-Pérez for technical support in the susceptibility assays. R.S.R.H. thanks Andelija M. Malenović and Ana Protic for kindly inviting him to contribute a full research article to the Special Issue of Pharmaceuticals on “The Age of In-Silico Methods in Drug Discovery, Development, Manufacture and Quality Control”.

Conflicts of Interest: The authors declare that they have no conflict of interest regarding the publication of this article, financial, and/or otherwise.

References

1. Barlam, T. F.; Cosgrove, S. E.; Abbo, L. M.; MacDougall, C.; Schuetz, A. N.; Septimus, E. J.; Srinivasan, A.; Dellit, T. H.; Falck-Ytter, Y. T.; Fishman, N. O.; Hamilton, C. W.; Jenkins, T. C.; Lipsett, P. A.; Malani, P. N.; May, L. S.; Moran, G. J.; Neuhauser, M. M.; Newland, J. G.; Ohl, C. A.; Samore, M. H.; Seo, S. K.; Trivedi, K. K., Implementing an Antibiotic Stewardship Program: Guidelines by the Infectious Diseases Society of America and the Society for Healthcare Epidemiology of America. *Clin Infect Dis* **2016**, 62 (10), e51-77. <https://doi.org/10.1093/cid/ciw118>
2. Pollack, L. A.; Srinivasan, A., Core elements of hospital antibiotic stewardship programs from the Centers for Disease Control and Prevention. *Clin Infect Dis* **2014**, 59 Suppl 3, S97-100. <https://doi.org/10.1093/cid/ciu542>
3. Kadri, S. S., Key Takeaways From the U.S. CDC's 2019 Antibiotic Resistance Threats Report for Frontline Providers. *Crit Care Med* **2020**, 48 (7), 939-945. <https://doi.org/10.1097/CCM.0000000000004371>
4. Antibiotic resistance threats in the United States, 2019. 2019.
5. de Kraker, M. E.; Stewardson, A. J.; Harbarth, S., Will 10 Million People Die a Year due to Antimicrobial Resistance by 2050? *PLoS Med* **2016**, 13 (11), e1002184. <https://doi.org/10.1371/journal.pmed.1002184>
6. Novoa-Farias, O.; Frati-Munari, A. C.; Peredo, M. A.; Flores-Juarez, S.; Novoa-Garcia, O.; Galicia-Tapia, J.; Romero-Carpi, C. E., Susceptibility to rifaximin and other antimicrobials of bacteria isolated in patients with acute gastrointestinal infections in Southeast Mexico. *Rev Gastroenterol Mex* **2017**, 82 (3), 226-233. <https://doi.org/10.1016/j.rgmxen.2017.03.012>
7. Ramirez-Castillo, F. Y.; Moreno-Flores, A. C.; Avelar-Gonzalez, F. J.; Marquez-Diaz, F.; Harel, J.; Guerrero-Barrera, A. L., An evaluation of multidrug-resistant *Escherichia coli* isolates in urinary tract infections from Aguascalientes, Mexico: cross-sectional study. *Ann Clin Microbiol Antimicrob* **2018**, 17 (1), 34. <https://doi.org/10.1186/s12941-018-0286-5>
8. Sosa-Hernandez, O.; Matias-Tellez, B.; Estrada-Hernandez, A.; Cureno-Diaz, M. A.; Bello-Lopez, J. M., Incidence and costs of ventilator-associated pneumonia in the adult intensive care unit of a tertiary referral hospital in Mexico. *Am J Infect Control* **2019**, 47 (9), e21-e25. <https://doi.org/10.1016/j.ajic.2019.02.031>
9. Asokan, G. V.; Vanitha, A., WHO global priority pathogens list on antibiotic resistance: an urgent need for action to integrate One Health data. *Perspectives in public health* **2018**, 138 (2), 87-88. <https://doi.org/10.1177/1757913917743881>
10. Custovic, A.; Smajlovic, J.; Tihic, N.; Hadzic, S.; Ahmetagic, S.; Hadzagic, H., Epidemiological monitoring of nosocomial infections caused by *Acinetobacter baumannii*. *Med Arch* **2014**, 68 (6), 402-6. <https://doi.org/10.5455/medarh.2014.68.402-406>
11. Dijkshoorn, L.; Nemec, A.; Seifert, H., An increasing threat in hospitals: multidrug-resistant *Acinetobacter baumannii*. *Nat Rev Microbiol* **2007**, 5 (12), 939-51. <https://doi.org/10.1038/nrmicro1789>
12. Geisinger, E.; Huo, W.; Hernandez-Bird, J.; Isberg, R. R., *Acinetobacter baumannii*: envelope determinants that control drug resistance, virulence, and surface variability. *Annual review of microbiology* **2019**, 73, 481-506. <https://doi.org/10.1146/annurev-micro-020518-115714>
13. Montefour, K.; Frieden, J.; Hurst, S.; Helmich, C.; Headley, D.; Martin, M.; Boyle, D. A., *Acinetobacter baumannii*: an emerging multidrug-resistant pathogen in critical care. *Crit Care Nurse* **2008**, 28 (1), 15-25; quiz 26.
14. Sebely, P. J.; Riddle, M. S.; Petersen, K., *Acinetobacter baumannii* skin and soft-tissue infection associated with war trauma. *Clin Infect Dis* **2008**, 47 (4), 444-9. <https://doi.org/10.1086/590568>

15. O'Toole, G.; Kaplan, H. B.; Kolter, R., Biofilm formation as microbial development. *Annual review of microbiology* **2000**, 54, 49. <https://doi.org/10.1146/annurev.micro.54.1.49>
16. Butler, M. S.; Gigante, V.; Sati, H.; Paulin, S.; Al-Sulaiman, L.; Rex, J. H.; Fernandes, P.; Arias, C. A.; Paul, M.; Thwaites, G. E., Analysis of the clinical pipeline of treatments for drug-resistant bacterial infections: despite progress, more action is needed. *Antimicrobial agents and chemotherapy* **2022**, 66 (3), e01991-21. <https://doi.org/10.1128/aac.01991-21>
17. Dheman, N.; Mahoney, N.; Cox, E. M.; Farley, J. J.; Amini, T.; Lanthier, M. L., An Analysis of Antibacterial Drug Development Trends in the United States, 1980-2019. *Clin Infect Dis* **2021**, 73 (11), e4444-e4450. <https://doi.org/10.1093/cid/ciaa859>
18. Luepke, K. H.; Suda, K. J.; Boucher, H.; Russo, R. L.; Bonney, M. W.; Hunt, T. D.; Mohr, J. F., 3rd, Past, Present, and Future of Antibacterial Economics: Increasing Bacterial Resistance, Limited Antibiotic Pipeline, and Societal Implications. *Pharmacootherapy* **2017**, 37 (1), 71-84. <https://doi.org/10.1002/phar.1868>
19. Vila, J.; Moreno-Morales, J.; Balleste-Delpierre, C., Current landscape in the discovery of novel antibacterial agents. *Clin Microbiol Infect* **2020**, 26 (5), 596-603. <https://doi.org/10.1016/j.cmi.2019.09.015>
20. Jackson, N.; Czaplewski, L.; Piddock, L. J. V., Discovery and development of new antibacterial drugs: learning from experience? *J Antimicrob Chemother* **2018**, 73 (6), 1452-1459. <https://doi.org/10.1093/jac/dky019>
21. Feher, M.; Schmidt, J. M., Property distributions: differences between drugs, natural products, and molecules from combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **2003**, 43 (1), 218-227. <https://doi.org/10.1021/ci0200467>
22. Whitty, A.; Zhong, M.; Viarengo, L.; Beglov, D.; Hall, D. R.; Vajda, S., Quantifying the chameleonic properties of macrocycles and other high-molecular-weight drugs. *Drug Discov Today* **2016**, 21 (5), 712-7. <https://doi.org/10.1016/j.drudis.2016.02.005>
23. Bueschleb, M.; Dorich, S.; Hanessian, S.; Tao, D.; Schenthal, K. B.; Overman, L. E., Synthetic strategies toward natural products containing contiguous stereogenic quaternary carbon atoms. *Angew. Chem. Int. Ed.* **2016**, 55 (13), 4156-4186. <https://doi.org/10.1002/anie.201507549>
24. Ganesan, A., The impact of natural products upon modern drug discovery. *Current opinion in chemical biology* **2008**, 12 (3), 306-317. <https://doi.org/10.1016/j.cbpa.2008.03.016>
25. Kong, D.-X.; Jiang, Y.-Y.; Zhang, H.-Y., Marine natural products as sources of novel scaffolds: Achievement and concern. *Drug Discov Today* **2010**, 15 (21-22), 884-886. <https://doi.org/10.1016/j.drudis.2010.09.002>
26. Wei, W.; Cherukupalli, S.; Jing, L.; Liu, X.; Zhan, P., Fsp(3): A new parameter for drug-likeness. *Drug Discov Today* **2020**, 25 (10), 1839-1845. <https://doi.org/10.1016/j.drudis.2020.07.017>
27. Furukawa, A.; Schwochert, J.; Pye, C. R.; Asano, D.; Edmondson, Q. D.; Turmon, A. C.; Klein, V. G.; Ono, S.; Okada, O.; Lokey, R. S., Drug-Like Properties in Macrocycles above MW 1000: Backbone Rigidity versus Side-Chain Lipophilicity. *Angew. Chem. Int. Ed.* **2020**, 59 (48), 21571-21577. <https://doi.org/10.1002/anie.202004550>
28. Cragg, G. M.; Scheifele, S. A.; Suffness, M.; Grever, M. R., The taxol supply crisis. New NCI policies for handling the large-scale production of novel natural product anticancer and anti-HIV agents. *J. Nat. Prod.* **1993**, 56 (10), 1657-68. <https://doi.org/10.1021/np50100a001>
29. Chhetri, B. K.; Lavoie, S.; Sweeney-Jones, A. M.; Kubanek, J., Recent trends in the structural revision of natural products. *Nat. Prod. Rep.* **2018**, 35 (6), 514-531. <https://doi.org/10.1039/C8NP00011E>
30. Muratov, E. N.; Bajorath, J.; Sheridan, R. P.; Tetko, I. V.; Filimonov, D.; Poroikov, V.; Oprea, T. I.; Baskin, I. I.; Varnek, A.; Roitberg, A., QSAR without borders. *Chem. Soc. Rev.* **2020**, 49 (11), 3525-3564. <https://doi.org/10.1039/D0CS00098A>
31. Banegas-Luna, A. J.; Ceron-Carrasco, J. P.; Perez-Sanchez, H., A review of ligand-based virtual screening web tools and screening algorithms in large molecular databases in the age of big data. *Future Med Chem* **2018**, 10 (22), 2641-2658. <https://doi.org/10.4155/fmc-2018-0076>
32. Vazquez, J.; Lopez, M.; Gibert, E.; Herrero, E.; Luque, F. J., Merging Ligand-Based and Structure-Based Methods in Drug Discovery: An Overview of Combined Virtual Screening Approaches. *Molecules* **2020**, 25 (20), 4723. <https://doi.org/10.3390/molecules25204723>
33. Wang, Z.; Sun, H.; Shen, C.; Hu, X.; Gao, J.; Li, D.; Cao, D.; Hou, T., Combined strategies in structure-based virtual screening. *Phys Chem Chem Phys* **2020**, 22 (6), 3149-3159. <https://doi.org/10.1039/C9CP06303J>
34. Neves, B. J.; Braga, R. C.; Melo-Filho, C. C.; Moreira-Filho, J. T.; Muratov, E. N.; Andrade, C. H., QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. *Front Pharmacol* **2018**, 9, 1275. <https://doi.org/10.3389/fphar.2018.01275>
35. Zhang, L.; Fourches, D.; Sedykh, A.; Zhu, H.; Golbraikh, A.; Ekins, S.; Clark, J.; Connelly, M. C.; Sigal, M.; Hodges, D.; Guiguenme, A.; Guy, R. K.; Tropsha, A., Discovery of novel antimalarial compounds enabled by QSAR-based virtual screening. *J. Chem. Inf. Model* **2013**, 53 (2), 475-92. <https://doi.org/10.1021/ci300421n>
36. Neves, B. J.; Dantas, R. F.; Senger, M. R.; Melo-Filho, C. C.; Valente, W. C.; de Almeida, A. C.; Rezende-Neto, J. M.; Lima, E. F.; Paveley, R.; Furnham, N.; Muratov, E.; Kamentsky, L.; Carpenter, A. E.; Braga, R. C.; Silva-Junior, F. P.; Andrade, C. H., Discovery of New Anti-Schistosomal Hits by Integration of QSAR-Based Virtual Screening and High Content Screening. *J. Med. Chem.* **2016**, 59 (15), 7075-88. <https://doi.org/10.1021/acs.jmedchem.5b02038>
37. Gomes, M. N.; Braga, R. C.; Grzelak, E. M.; Neves, B. J.; Muratov, E.; Ma, R.; Klein, L. L.; Cho, S.; Oliveira, G. R.; Franzblau, S. G.; Andrade, C. H., QSAR-driven design, synthesis and discovery of potent chalcone derivatives with antitubercular activity. *Eur. J. Med. Chem.* **2017**, 137, 126-138. <https://doi.org/10.1016/j.ejmech.2017.05.026>
38. Lian, W.; Fang, J.; Li, C.; Pang, X.; Liu, A.-L.; Du, G.-H., Discovery of Influenza A virus neuraminidase inhibitors using support vector machine and Naïve Bayesian models. *Molecular diversity* **2016**, 20 (2), 439-451. <https://doi.org/10.1007/s11030-015-9641-z>

39. Kurczyk, A.; Warszycki, D.; Musiol, R.; Kafel, R.; Bojarski, A. J.; Polanski, J., Ligand-Based Virtual Screening in a Search for Novel Anti-HIV-1 Chemotypes. *J. Chem. Inf. Model.* **2015**, 55 (10), 2168-77. <https://doi.org/10.1021/acs.jcim.5b00295>

40. Prado-Prado, F. J.; González-Díaz, H.; Santana, L.; Uriarte, E., Unified QSAR approach to antimicrobials. Part 2: predicting activity against more than 90 different species in order to halt antibacterial resistance. *Bioorganic & medicinal chemistry* **2007**, 15 (2), 897-902. <https://doi.org/10.1016/j.bmc.2006.10.039>

41. Semenyuta, I. V.; Trush, M. M.; Kovalishyn, V. V.; Rogalsky, S. P.; Hodyna, D. M.; Karpov, P.; Xia, Z.; Tetko, I. V.; Metelytsia, L. O., Structure-Activity Relationship Modeling and Experimental Validation of the Imidazolium and Pyridinium Based Ionic Liquids as Potential Antibacterials of MDR *Acinetobacter baumannii* and *Staphylococcus aureus*. *Int. J. Mol. Sci.* **2021**, 22 (2), 563. <https://doi.org/10.3390/ijms22020563>

42. Lucas, A. J.; Sproston, J. L.; Barton, P.; Riley, R. J., Estimating human ADME properties, pharmacokinetic parameters and likely clinical dose in drug discovery. *Expert Opin. Drug Discov.* **2019**, 14 (12), 1313-1327. <https://doi.org/10.1080/17460441.2019.1660642>

43. Currie, G. M., Pharmacology, Part 2: Introduction to Pharmacokinetics. *J. Nucl. Med. Technol.* **2018**, 46 (3), 221-230. <https://doi.org/10.2967/jnmt.117.199638>

44. Vinarov, Z.; Abdallah, M.; Agundez, J. A. G.; Allegaert, K.; Basit, A. W.; Braeckmans, M.; Ceulemans, J.; Corsetti, M.; Griffin, B. T.; Grimm, M.; Keszhelyi, D.; Koziolek, M.; Madla, C. M.; Matthys, C.; McCoubrey, L. E.; Mitra, A.; Reppas, C.; Stappaerts, J.; Steenackers, N.; Trevaskis, N. L.; Vanuytsel, T.; Vertzoni, M.; Weitschies, W.; Wilson, C.; Augustijns, P., Impact of gastrointestinal tract variability on oral drug absorption and pharmacokinetics: An UNGAP review. *Eur. J. Pharm. Sci.* **2021**, 162, 105812. <https://doi.org/10.1016/j.ejps.2021.105812>

45. Roy, K.; Chakraborty, P.; Mitra, I.; Ojha, P. K.; Kar, S.; Das, R. N., Some case studies on application of "rm2" metrics for judging quality of quantitative structure-activity relationship predictions: emphasis on scaling of response data. *J. Comput. Chem.* **2013**, 34 (12), 1071-1082. <https://doi.org/10.1002/jcc.23231>

46. Roy, K.; Das, R. N., QSTR with extended topochemical atom (ETA) indices. 16. Development of predictive classification and regression models for toxicity of ionic liquids towards *Daphnia magna*. *J. Hazard Mater.* **2013**, 254-255, 166-178. <https://doi.org/10.1016/j.jhazmat.2013.03.023>

47. Roy, P. P.; Roy, K., On some aspects of variable selection for partial least squares regression models. *QSAR Comb. Sci.* **2008**, 27 (3), 302-313 <https://doi.org/10.1002/qsar.200710043>

48. Kar, S.; Roy, K.; Leszczynski, J., Applicability domain: a step toward confident predictions and decidability for QSAR modeling. In *Computational Toxicology*, Springer: 2018; pp 141-169. https://doi.org/10.1007/978-1-4939-7899-1_6

49. Netzeva, T. I.; Worth, A.; Aldenberg, T.; Benigni, R.; Cronin, M. T.; Gramatica, P.; Jaworska, J. S.; Kahn, S.; Klopman, G.; Marchant, C. A.; Myatt, G.; Nikolova-Jeliazkova, N.; Patlewicz, G. Y.; Perkins, R.; Roberts, D.; Schultz, T.; Stanton, D. W.; van de Sandt, J. J.; Tong, W.; Veith, G.; Yang, C., Current status of methods for defining the applicability domain of (quantitative) structure-activity relationships. The report and recommendations of ECVAM Workshop 52. *Altern. Lab. Anim.* **2005**, 33 (2), 155-73. <https://doi.org/10.1177/026119290503300209>

50. Roy, K.; Kar, S.; Ambure, P., On a simple approach for determining applicability domain of QSAR models. *Chemometrics and Intelligent Laboratory Systems* **2015**, 145, 22-29. <https://doi.org/10.1016/j.chemolab.2015.04.013>

51. Lin, Y. C.; Ribaucourt, A.; Moazami, Y.; Pierce, J. G., Concise Synthesis and Antimicrobial Evaluation of the Guanidinium Alkaloid Batzelladine D: Development of a Stereodivergent Strategy. *J. Am. Chem. Soc.* **2020**, 142 (21), 9850-9857. <https://doi.org/10.1021/jacs.0c04091>

52. Matsingos, C.; Al-Adhami, T.; Jamshidi, S.; Hind, C.; Clifford, M.; Mark Sutton, J.; Rahman, K. M., Synthesis, microbiological evaluation and structure activity relationship analysis of linezolid analogues with different C5-acylamino substituents. *Bioorg. Med. Chem.* **2021**, 49, 116397. <https://doi.org/10.1016/j.bmc.2021.116397>

53. Singh, A.; Kumar, N.; Singh, S.; Sewariya, S.; Sharma, M. K.; Chandra, R., High-valued pyrazinoindole analogues: Synthesis, antibacterial activity, structure activity relationship and molecular dynamics analyses. *Results in Chemistry* **2021**, 3, 100194. <https://doi.org/10.1016/j.rechem.2021.100194>

54. Wang, J.; Zhang, P. L.; Ansari, M. F.; Li, S.; Zhou, C. H., Molecular design and preparation of 2-aminothiazole sulfanilamide oximes as membrane active antibacterial agents for drug resistant *Acinetobacter baumannii*. *Bioorg. Chem.* **2021**, 113, 105039. <https://doi.org/10.1016/j.bioorg.2021.105039>

55. Zhou, M.; Eun, Y. J.; Guzei, I. A.; Weibel, D. B., Structure-activity studies of divin: an inhibitor of bacterial cell division. *ACS Med. Chem. Lett.* **2013**, 4 (9), 880-885. <https://doi.org/10.1021/ml400234x>

56. Lyons, A.; Kirkham, J.; Blades, K.; Orr, D.; Dauncey, E.; Smith, O.; Dick, E.; Walker, R.; Matthews, T.; Bunt, A.; Finlayson, J.; Morrison, I.; Savage, V. J.; Moyo, E.; Butler, H. S.; Newman, R.; Ooi, N.; Smith, A.; Charrier, C.; Ratcliffe, A. J.; Stokes, N. R.; Best, S.; Salisbury, A. M.; Craighead, M.; Cooper, I. R., Discovery and structure-activity relationships of a novel oxazolidinone class of bacterial type II topoisomerase inhibitors. *Bioorg. Med. Chem. Lett.* **2022**, 65, 128648. <https://doi.org/10.1016/j.bmcl.2022.128648>

57. Vereshchagin, A. N.; Frolov, N. A.; Konyuhova, V. Y.; Kapelistaya, E. A.; Hansford, K. A.; Egorov, M. P., Investigations into the structure-activity relationship in gemini QACs based on biphenyl and oxydiphenyl linker. *RSC Adv.* **2021**, 11 (6), 3429-3438. <https://doi.org/10.1039/D0RA08900A>

58. Randić, M., On characterization of cyclic structures. *J. Chem. Inf. Comput. Sci.* **1997**, 37 (6), 1063-1071. <https://doi.org/10.1021/ci9702407>

59. Helguera, A. M.; Combes, R. D.; Gonzalez, M. P.; Cordeiro, M. N., Applications of 2D descriptors in drug design: a DRAGON tale. *Curr. Top. Med. Chem.* **2008**, 8 (18), 1628-55. <https://doi.org/10.2174/156802608786786598>

60. Trinajstic, N.; Babic, D.; Nikolic, S.; Plavsic, D.; Amic, D.; Mihalic, Z., The Laplacian matrix in chemistry. *J. Chem. Inf. Comput. Sci.* **1994**, 34 (2), 368-376. <https://doi.org/10.1021/ci00018a023>

61. Sadeghian-Rizi, S.; Sakhteman, A.; Hassanzadeh, F., A quantitative structure-activity relationship (QSAR) study of some diaryl urea derivatives of B-RAF inhibitors. *Research in pharmaceutical sciences* **2016**, 11 (6), 445-453. <https://doi.org/10.4103/1735-5362.194869>

62. Amini, Z.; Fatemi, M. H.; Gharaghani, S., Hybrid docking-QSAR studies of DPP-IV inhibition activities of a series of aminomethyl-piperidones. *Comput. Biol. Chem.* **2016**, 64, 335-345. <https://doi.org/10.1016/j.compbiochem.2016.08.003>

63. Geary, R. C., The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician* **1954**, 5 (3), 115-146. <https://doi.org/10.2307/2986645>

64. Velázquez-Libera, J. L.; Caballero, J.; Toropova, A. P.; Toropov, A. A., Estimation of 2D autocorrelation descriptors and 2D Monte Carlo descriptors as a tool to build up predictive models for acetylcholinesterase (AChE) inhibitory activity. *Chemosometrics and Intelligent Laboratory Systems* **2019**, 184, 14-21. <https://doi.org/10.1016/j.chemolab.2018.11.008>

65. Kuhn, B.; Mohr, P.; Stahl, M., Intramolecular hydrogen bonding in medicinal chemistry. *J Med Chem* **2010**, 53 (6), 2601-11. <https://doi.org/10.1021/jm100087s>

66. Yang, N. J.; Hinner, M. J., Getting Across the Cell Membrane: An Overview for Small Molecules, Peptides, and Proteins. In *Site-Specific Protein Labeling: Methods and Protocols*, Gautier, A.; Hinner, M. J., Eds. Springer New York: New York, NY, 2015; pp 29-53. https://doi.org/10.1007/978-1-4939-2272-7_3

67. Khondker, A.; Bider, R. C.; Passos-Gastaldo, I.; Wright, G. D.; Rheinstadter, M. C., Membrane interactions of non-membrane targeting antibiotics: The case of aminoglycosides, macrolides, and fluoroquinolones. *Biochim. Biophys. Acta Biomembr.* **2021**, 1863 (1), 183448. <https://doi.org/10.1016/j.bbamem.2020.183448>

68. Ghose, A. K.; Crippen, G. M., Atomic Physicochemical Parameters for Three-Dimensional Structure-Directed Quantitative Structure-Activity Relationships I. Partition Coefficients as a Measure of Hydrophobicity. *J. Comput. Chem.* **1986**, 7 (4), 565-577. <https://doi.org/10.1021/ci00063a006>

69. Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J., Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. *J. Phys. Chem. A* **1998**, 102 (21), 3762-3772. <https://doi.org/10.1021/jp980230o>

70. Viswanadhan, V. N.; Ghose, A. K.; Revankar, G. R.; Robins, R. K., Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics. *J. Chem. Inf. Model.* **1989**, 29 (3), 163-172. <https://doi.org/10.1021/ci00063a006>

71. De, P.; Kumar, V.; Kar, S.; Roy, K.; Leszczynski, J., Repurposing FDA approved drugs as possible anti-SARS-CoV-2 medications using ligand-based computational approaches: sum of ranking difference-based model selection. *Struct. Chem.* **2022**, 33 (5), 1741-1753. <https://doi.org/10.1007/s11224-022-01975-3>

72. De, P.; Roy, K., Nitroaromatics as hypoxic cell radiosensitizers: A 2D-QSAR approach to explore structural features contributing to radiosensitization effectiveness. *European Journal of Medicinal Chemistry Reports* **2022**, 4, 100035. <https://doi.org/10.1016/j.ejmc.2022.100035>

73. Alizadeh, S. R.; Ebrahimzadeh, M. A., Antiviral Activities of Pyridine Fused and Pyridine Containing Heterocycles, A Review (from 2000 to 2020). *Mini Rev. Med. Chem.* **2021**, 21 (17), 2584-2611. <https://doi.org/10.2174/138955752166210126143558>

74. Altaf, A. A.; Shahzad, A.; Gul, Z.; Rasool, N.; Badshah, A.; Lal, B.; Khan, E., A review on the medicinal importance of pyridine derivatives. *J. Drug Des. Med. Chem.* **2015**, 1 (1), 1-11. https://www.sciencepublishinggroup.com/journal/paper_info?journalid=329&doi=10.11648/j.ddmc.20150101.11

75. Huang, Z.; Yoon, S., Integration of Time-Series Transcriptomic Data with Genome-Scale CHO Metabolic Models for mAb Engineering. *Processes* **2020**, 8 (3), 331. <https://doi.org/10.3390/pr8030331>

76. Schmidt, S.; Schindler, M.; Eriksson, L., Block-wise Exploration of Molecular Descriptors with Multi-block Orthogonal Component Analysis (MOCA). *Mol Inform* **2022**, 41 (5), e2100165. <https://doi.org/10.1002/minf.202100165>

77. Bai, L. Y.; Dai, H.; Xu, Q.; Junaid, M.; Peng, S. L.; Zhu, X.; Xiong, Y.; Wei, D. Q., Prediction of Effective Drug Combinations by an Improved Naive Bayesian Algorithm. *Int. J. Mol. Sci.* **2018**, 19 (2), 467. <https://doi.org/10.3390/ijms19020467>

78. McEuen, K.; Borlak, J.; Tong, W.; Chen, M., Associations of Drug Lipophilicity and Extent of Metabolism with Drug-Induced Liver Injury. *Int. J. Mol. Sci.* **2017**, 18 (7), 1335. <https://doi.org/10.3390/ijms18071335>

79. Nachege, J. B.; Ishoso, D. K.; Otokoye, J. O.; Hermans, M. P.; Machekano, R. N.; Sam-Agudu, N. A.; Bongo-Pasi Nswe, C.; Mbala-Kingebeni, P.; Madinga, J. N.; Mukendi, S.; Kolie, M. C.; Nkwembe, E. N.; Mbuiy, G. M.; Nsio, J. M.; Mukeba Tshialala, D.; Tshiasuma Pipo, M.; Ahuka-Mundeke, S.; Muyembe-Tamfum, J. J.; Mofenson, L.; Smith, G.; Mills, E. J.; Mellors, J. W.; Zumla, A.; Mavungu Landu, D. J.; Kayembe, J. M., Clinical Characteristics and Outcomes of Patients Hospitalized for COVID-19 in Africa: Early Insights from the Democratic Republic of the Congo. *Am J. Trop. Med. Hyg.* **2020**, 103 (6), 2419-2428. <https://doi.org/10.4269/ajtmh.20-1240>

80. Lee, S.; Barron, M. G., A mechanism-based 3D-QSAR approach for classification and prediction of acetylcholinesterase inhibitory potency of organophosphate and carbamate analogs. *J. Comput. Aided Mol. Des.* **2016**, 30 (4), 347-63. <https://doi.org/10.1007/s10822-016-9910-7>

81. Pilon-Jimenez, B. A.; Saldivar-Gonzalez, F. I.; Diaz-Eufracio, B. I.; Medina-Franco, J. L., BIOFACQUIM: A Mexican Compound Database of Natural Products. *Biomolecules* **2019**, 9 (1), 31. <https://doi.org/10.3390/biom9010031>

82. Munoz-Rodriguez, P.; Carruthers, T.; Wood, J. R. I.; Williams, B. R. M.; Weitemier, K.; Kronmiller, B.; Goodwin, Z.; Sumadijaya, A.; Anglin, N. L.; Filer, D.; Harris, D.; Rausher, M. D.; Kelly, S.; Liston, A.; Scotland, R. W., A taxonomic

monograph of *Ipomoea* integrated across phylogenetic scales. *Nat. Plants* **2019**, *5* (11), 1136-1144. <https://doi.org/10.1038/s41477-019-0535-4>

83. Nimmakayala, P.; Vajja, G.; Reddy, U. K., *Ipomoea*. In *Wild Crop Relatives: Genomic and Breeding Resources*, Kole, C., Ed. Springer Berlin Heidelberg: Berlin, Heidelberg, **2011**; pp 123-132. https://doi.org/10.1007/978-3-642-21102-7_7

84. Protti, I. F.; Rodrigues, D. R.; Fonseca, S. K.; Alves, R. J.; Oliveira, R. B.; Maltarollo, V. G., Do Drug-likeness Rules Apply to Oral Prodrugs? *ChemMedChem* **2021**, *16* (9), 1446-1456. <https://doi.org/10.1002/cmdc.202000805>

85. Stepanic, V.; Ziher, D.; Gabelica-Markovic, V.; Jelic, D.; Nunhuck, S.; Valko, K.; Kostrun, S., Physicochemical profile of macrolides and their comparison with small molecules. *Eur. J. Med. Chem.* **2012**, *47* (1), 462-472. <https://doi.org/10.1016/j.ejmech.2011.11.016>

86. Walters, W. P., Going further than Lipinski's rule in drug design. *Expert Opinion on Drug Discovery* **2012**, *7* (2), 99-107. <https://doi.org/10.1517/17460441.2012.648612>

87. Leon-Rivera, I.; Del Rio-Portilla, F.; Enriquez, R. G.; Rangel-Lopez, E.; Villeda, J.; Rios, M. Y.; Navarrete-Vazquez, G.; Hurtado-Dias, I.; Guzman-Valdivieso, U.; Nunez-Urquiza, V.; Escobedo-Martinez, C., Hepta-, hexa-, penta-, tetra-, and trisaccharide resin glycosides from three species of *Ipomoea* and their antiproliferative activity on two glioma cell lines. *Magnetic Resonance in Chemistry* **2017**, *55* (3), 214-223. <https://doi.org/10.1002/mrc.4476>

88. Leon-Rivera, I.; Villeda-Hernandez, J.; Campos-Pena, V.; Aguirre-Moreno, A.; Estrada-Soto, S.; Navarrete-Vazquez, G.; Rios, M. Y.; Aguilar-Guadarrama, B.; Castillo-Espana, P.; Rivera-Leyva, J. C., Evaluation of the neuroprotective activity of stansin 6, a resin glycoside from *Ipomoea stans*. *Bioorg. Med. Chem. Lett.* **2014**, *24* (15), 3541-5. <https://doi.org/10.1016/j.bmcl.2014.05.055>

89. Mirón-López, G.; Herrera-Ruiz, M.; Estrada-Soto, S.; Aguirre-Crespo, F.; Vázquez-Navarrete, L.; León-Rivera, I., Resin Glycosides from the Roots of *Ipomoea tyrianthina* and Their Biological Activity. *J. Nat. Prod.* **2007**, *70* (4), 557-562. <https://doi.org/10.1021/np0604634>

90. Gonzalez, M.; Zamilpa, A.; Marquina, S.; Navarro, V.; Alvarez, L., Antimycotic spirostanol saponins from *Solanum hispidum* leaves and their structure-activity relationships. *J. Nat. Prod.* **2004**, *67* (6), 938-41. <https://doi.org/10.1021/np0305019>

91. Lee, C. L.; Hwang, T. L.; Yang, J. C.; Cheng, H. T.; He, W. J.; Yen, C. T.; Kuo, C. L.; Chen, C. J.; Chang, W. Y.; Wu, Y. C., Anti-Inflammatory Spirostanol and Furostanol Saponins from *Solanum macaonense*. *J. Nat. Prod.* **2014**, *77* (8), 1770-83. <https://doi.org/10.1021/np500057b>

92. Tezuka, Y.; Honda, K.; Banskota, A. H.; Thet, M. M.; Kadota, S., Kinmoonosides A-C, three new cytotoxic saponins from the fruits of *Acacia concinna*, a medicinal plant collected in myanmar. *J. Nat. Prod.* **2000**, *63* (12), 1658-64. <https://doi.org/10.1021/np000347f>

93. Pettit, G. R.; Schaufelberger, D. E.; Nieman, R. A.; Dufresne, C.; Saenz-Renauld, J. A., Antineoplastic agents, 177. Isolation and structure of phyllanthostatin 6. *J. Nat. Prod.* **1990**, *53* (6), 1406-1413. <https://doi.org/10.1021/np50072a002>

94. Susplugas, S.; Hung, N. V.; Bignon, J.; Thoison, O.; Kruczynski, A.; Sevenet, T.; Gueritte, F., Cytotoxic arylnaphthalene lignans from a Vietnamese acanthaceae, *Justicia patentiflora*. *J. Nat. Prod.* **2005**, *68* (5), 734-738. <https://doi.org/10.1021/np050028u>

95. Shi, D. K.; Zhang, W.; Ding, N.; Li, M.; Li, Y. X., Design, synthesis and biological evaluation of novel glycosylated diphyllin derivatives as topoisomerase II inhibitors. *Eur. J. Med. Chem.* **2012**, *47* (1), 424-31. <https://doi.org/10.1016/j.ejmech.2011.11.011>

96. Zhang, H. J.; Rumschlag-Booms, E.; Guan, Y. F.; Wang, D. Y.; Liu, K. L.; Li, W. F.; Nguyen, V. H.; Cuong, N. M.; Soejarto, D. D.; Fong, H. H. S.; Rong, L., Potent Inhibitor of Drug-Resistant HIV-1 Strains Identified from the Medicinal Plant *Justicia gendarussa*. *J. Nat. Prod.* **2017**, *80* (6), 1798-1807. <https://doi.org/10.1021/acs.jnatprod.7b00004>

97. Wang, T. M.; Hojo, T.; Ran, F. X.; Wang, R. F.; Wang, R. Q.; Chen, H. B.; Cui, J. R.; Shang, M. Y.; Cai, S. Q., Cardenolides from *Saussurea stella* with cytotoxicity toward cancer cells. *J. Nat. Prod.* **2007**, *70* (9), 1429-33. <https://doi.org/10.1021/np070150o>

98. Gelmi, M. L.; Fontana, G.; Pocar, D.; Pontremoli, G.; Pellegrino, S.; Bombardelli, E.; Riva, A.; Balduini, W.; Carloni, S.; Cimino, M., Novel 3-O-glycosyl-3-demethylthiocolchicines as ligands for glycine and gamma-aminobutyric acid receptors. *J. Med. Chem.* **2007**, *50* (9), 2245-8. <https://doi.org/10.1021/jm061056s>

99. Ambure, P.; Cordeiro, M., Importance of data curation in QSAR studies especially while modeling large-size datasets. *Ecotoxicological QSARs* **2020**, 97-109. https://doi.org/10.1007/978-1-0716-0150-1_5

100. Fourches, D.; Muratov, E.; Tropsha, A., Trust, but verify: on the importance of chemical structure curation in cheminformatics and QSAR modeling research. *J. Chem. Inf. Model.* **2010**, *50* (7), 1189-204. <https://doi.org/10.1021/ci100176x>

101. Young, D.; Martin, T.; Venkatapathy, R.; Harten, P., Are the chemical structures in your QSAR correct? *QSAR & combinatorial science* **2008**, *27* (11-12), 1337-1345. <https://doi.org/10.1002/qsar.200810084>

102. Kowalska-Krochmal, B.; Dudek-Wicher, R., The Minimum Inhibitory Concentration of Antibiotics: Methods, Interpretation, Clinical Relevance. *Pathogens* **2021**, *10* (2), 165. <https://doi.org/10.3390/pathogens10020165>

103. Hanwell, M. D.; Curtis, D. E.; Lonie, D. C.; Vandermeersch, T.; Zurek, E.; Hutchison, G. R., Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminform.* **2012**, *4* (1), 17. <https://doi.org/10.1186/1758-2946-4-17>

104. Avogadro home page. <https://avogadro.cc/>

105. MarvinSketch, C. 2018.0 program, Budapest, Hungary, 2022. <https://chemaxon.com/products/marvin>

106. Talete, S., Dragon for windows (software for molecular descriptor calculations), version 5.4. Talete, Srl, Milano, Italy, 2006.

107. Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M., MobyDigs: software for regression and classification models by genetic algorithms. *Data handling in science and technology* **2003**, *23*, 141-167. [https://doi.org/10.1016/S0922-3487\(03\)23005-7](https://doi.org/10.1016/S0922-3487(03)23005-7)

108. Holland, J. H., *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press: 1992.
109. Mattioni, B. E.; Jurs, P. C., Development of quantitative structure-activity relationship and classification models for a set of carbonic anhydrase inhibitors. *J. Chem. Inf. Comput. Sci.* **2002**, *42* (1), 94-102. <https://doi.org/10.1021/ci0100696>
110. Miller, A., *Subset selection in regression*. chapman and hall/CRC: 2002.
111. Ojha, P. K.; Mitra, I.; Das, R. N.; Roy, K., Further exploring rm2 metrics for validation of QSPR models. *Chemometrics and Intelligent Laboratory Systems* **2011**, *107* (1), 194-205. <https://doi.org/10.1016/j.chemolab.2011.03.011>
112. Schwalbe, R.; Steele-Moore, L.; Goodwin, A. C., *Antimicrobial susceptibility testing protocols*. Crc Press: 2007. <https://doi.org/10.1201/9781420014495>

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.