Chemoinformatics in antibacterial drug discovery: Simultaneous modeling of anti-enterococci activities and ADMET profiles through the use of probabilistic quadratic indices

Alejandro Speck-Planche,[a]* Valeria V. Kleandrova,[b] M. N. D. S. Cordeiro,[a]

[a] *REQUIMTE/Department of Chemistry and Biochemistry, University of Porto, 4169-007 Porto, Portugal.*
[b] *Faculty of Technology and Production Management, Moscow State University of Food Production, Volokolamskoe shosse 11, Moscow, Russia.*

ABSTRACT
Enterococci are Gram-positive bacteria responsible for causing multiple nosocomial infections in humans. Chemoinformatics could be a great ally of medicinal chemistry in the search for efficacious anti-enterococci drugs. Current methods cannot model the anti-enterococci activity and ADMET (absorption, distribution, metabolism, elimination, toxicity) properties at the same time. We create the first multitasking model for quantitative-structure biological effect relationships (mtk-QSBER), focused on the simultaneous prediction of anti-enterococci activities and ADMET profiles of compounds. The mtk-QSBER model was constructed by using a large and heterogeneous dataset of chemicals, and exhibited accuracy higher than 95% in both training and prediction sets. We provided the physicochemical interpretations of the molecular descriptors (probabilistic quadratic indices) that entered in the model. In order to demonstrate the practical utility of our model, we predicted multiple biological profiles of the investigational antibacterial drug oritavancin, and the results of the virtual predictions strongly converged with the experimental evidences. To date, this is one the most promising attempts to use a unified *in silico* model to guide drug discovery in antimicrobial research by predicting the antibacterial potency against enterococci, as well as the safety in laboratory animals and humans.

Keywords: anti-enterococci activity, ADMET, mtk-QSBER, TOMOCOMD, moving average approach, probabilistic quadratic indices.

To whom correspondence should be addressed:
E-mail: alejspivanovich@gmail.com; alejspivanovich@yahoo.es

Introduction
Enterococci belong to a group of Gram-positive, facultative anaerobic bacteria that can occur both, as single cocci and in chains.[1] These bacteria constitute dangerous opportunistic pathogens, which are major causes of nosocomial infections such as bacteremia, bacterial endocarditis, diverticulitis, meningitis and urinary tract infections.[1-3] When compared with other Gram-positive cocci such as bacteria of the genera *Staphylococcus* and *Streptococcus*, enterococci exhibit a lower degree of pathogenicity in terms of mortality, but they are reservoirs of antibiotic resistance genes.[4] Thus, multidrug resistant (MDR) strains of bacteria belonging to *Enterococcus* have emerged in the last years.[5-6]

Two problems arise to eliminate enterococci: the lack of an efficient antimicrobial chemotherapy, and appearance of undesirable ADMET (absorption, distribution, metabolism, elimination, toxicity) profiles that the antibacterial agent may have. From one side, medicinal chemistry has brought great benefits to drug discovery, being a discipline focused on the design, identification and preparation of biologically active compounds, and studying also the mechanisms of action at molecular level, as well as

the ADMET (absorption, distribution, metabolism, elimination, toxicity) parameters.[7] At the same time, even with the alliance of medicinal chemistry with powerful experimental techniques such as high-throughput screening (HTS) and combinatorial chemistry, the chemical space to be covered is huge (10^{63} small and medium size molecules) in order to search for new therapeutic agents with the desired properties.[8] On the other hand, if a drug candidate is discovered, serious concern is expected due to the possible lack of adequate ADMET properties,[9-11] which remains as one of the principal causes of disapproval of drugs.

Chemoinformatics has offered very useful theoretical/computational tools in drug discovery,[12-15] contributing to rationalize the chemical synthesis, as well as the evaluation of biological and/or ADMET profiles, and consequently, strengthening its link with medicinal chemistry. In fact, in the last years, promising chemoinformatic multi-target methodologies for quantitative-structure activity relationships (mt-QSAR) have been used for the prediction of diverse biological profiles,[16-40] against dissimilar targets (biomolecules, microorganisms, cell lines, mammals), and by using large and heterogeneous datasets of molecules. These methodologies have made possible the integration of different types of biological and chemical data under many sets of experimental conditions.[41-46]

Until now, there is no methodology or approach able to predict anti-enterococci activity and ADMET properties at the same time. The existence of such methodology would constitute a prime interest in drug discovery because the process of designing a drug would be guided in several stages: from *in vitro* assays to clinical studies. For this reason, and considering all the ideas mentioned above, this work introduces a promising chemoinformatic approach focused on constructing a multitasking model for quantitative-structure biological effect relationships (mtk-QSBER). In this context, the mtk-QSBER model is devoted to perform simultaneous predictions of anti-enterococci activities and ADMET parameters, with the aim of searching for safer antibacterial drugs against the aforementioned pathogens.

Materials and Methods
Dataset, calculation of molecular descriptors and creation of the mtk-QSBER model

The chemoinformatic methodology used for the extraction of the dataset, generation of the molecular descriptors, and construction the mtk-QSBER model has been reported in previous works of Gonzalez-Diaz and coworkers.[41] Also, similar explanations can be found in a recent report of Speck-Planche and Cordeiro.[45-47] Anyway, here, only the essential details will be given (**Fig. 1**). The dataset was composed by 29309 different drugs/chemicals that were extracted from CHEMBL,[48] being this a public source available at https://www.ebi.ac.uk/chembl/. The 29309 drugs/chemicals were assayed by considering at least 1 out of 19 measures of biological effect (m_e), against at least 1 out of 163 biological targets (b_t). These biological targets include biomacromolecules, bacterial strains, cell lines, and superior organisms such as mice, rats, and humans. All the experiments were carried out by considering at least 1 out of 3 types of assay information (a_i), with at least 1 out of 9 categories of target mapping (t_m), and where at least 1 out of 3 levels of curation/reliability of the assays (l_c) were taking into account. Notice that here, the combination of the elements (m_e), (b_t), (a_i), (t_m), and (l_c) represents a unique experimental condition, which can be defined by the ontology $c_j \rightarrow (m_e, b_t, a_i, t_m, l_c)$. In our dataset, several chemicals were tested in more than one experimental condition. For this the dataset contained 46822 cases as results of the combination of the aforementioned elements of c_j.
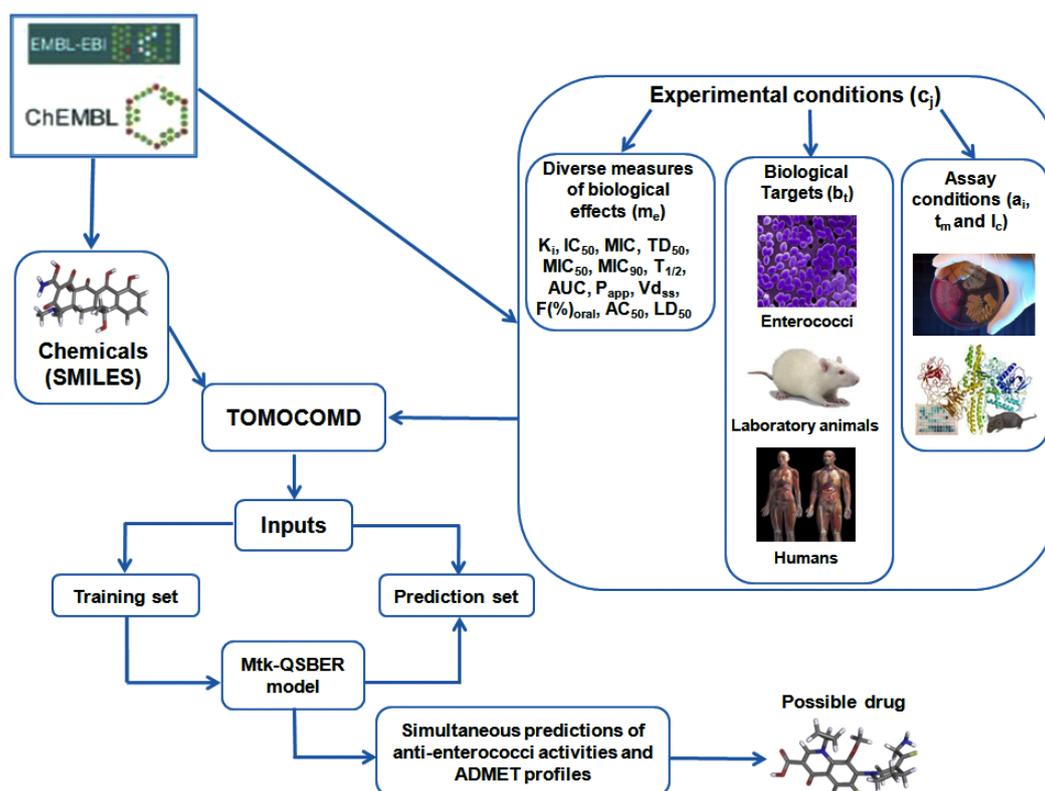
Fig. 1. Descriptive overview of the main steps involved in the development of the mtk-QSBER model.

Each of the 46822 cases were annotated as positive ($[BE_i(c_j) = 1]$) or negative [$BE_i(c_j) = -1$] according to certain cutoff values (**Table 1**), being $BE_i(c_j)$ a binary (categorical) variable, which defines the specific biological effect of a compound i under experimental condition c_j. On the other hand, a *txt file containing the SMILES codes of the cases was manually changed to *.smi, and after transformed to *.sdf by employing the program OpenBabel 2.3.0.[49] For the calculation of the molecular descriptors from the *.sdf file, the software TOMOCOMD-CARDD was used.[50] Descriptors calculated here, from order 1 to 5, and weighted by the hydrophobicity, the polar surface area, and the refractivity, are derived from the classical quadratic indices that have been widely reported in the literature.[51-56] They can be defined according to the following expression:

$$PTD_k(\overline{x}, \overline{y}) = [X^T][P]^k[Y] = \sum_{i=1}^{n} \sum_{j=1}^{n} {}^k p_{ij} \times x_i \times y_j \qquad (1)$$

In the Eq. 1, $Pq_k(x)$ is the mutual probability quadratic index of order k, weighted by the atomic physicochemical property x. Here, x is a molecular vector, $[X^T]$ is the transpose of $[X]$, being the latter, a column vector ($n \times 1$ matrix) with components $x_1,..., x_n$ (different atomic physicochemical properties). In addition, the terms ${}^k p_{ij}$ represent the mutual probabilities of the adjacent vertices (atoms) i and j belonging to the kth power of the matrix **P**. The element can be calculated according to the following expression:

$$p_{ij} = \frac{{}^k m_{ij}}{\sum_{i,j} {}^k m_{ij}} \qquad (2)$$

being ${}^k m_{ij}$ the elements of the classical matrix **M** previously reported.[51-56] Descriptors of the type $Pq_k(x)$ can only take into account the chemical structures of the molecules. More than 40 years ago, the statisticians George Box and Gwilym Jenkins published a

promising approach, which was devoted to the analysis, modeling, and forecasting of time series.[57-58]

Table 1. Summary of the cutoff values for the different measures of biological effects.

Measure of effect (Units)	Biological Profile	Concept	Cutoff value[a]
AC_{50} (nM)	ADMET (Metabolism)	Concentration to elicit the 50% of the maximal effect	≤6309.58
AUC (µM.hr)iv-H	ADMET (Bioavailability, elimination)	Area under the curve after intravenous administration in humans	≥88.57
AUC (µM.hr)iv-LA	ADMET (Bioavailability, elimination)	Area under the curve after intravenous administration in laboratory animals (*Mus musculus* and/or *Rattus norvegicus*)	≥10.00
AUC (µM.hr)oral-H	ADMET (Bioavailability, elimination)	Area under the curve after oral administration in humans	≥64.39
AUC (µM.hr)oral-LA	ADMET (Bioavailability, elimination)	Area under the curve after oral administration in laboratory animals (*Mus musculus* and/or *Rattus norvegicus*)	≥15.00
F (%)oral	ADMET (Bioavailability)	Oral bioavailability assessed as the fraction of an oral administered drug that reaches systemic circulation	≥60.00
IC_{50} (nM)ime	ADMET (Metabolism)	Concentration required for 50% inhibition of the activity of a protein involved in the metabolism	≤2400
K_i (nM)ime	ADMET (Metabolism)	Inhibition constant associated to a protein involved in the metabolism	≤2200
LD_{50} (µmol/kg)ip	ADMET (Toxicity)	Lethal dose at 50% after intraperitoneal administration	≥1110.00
LD_{50} (µmol/kg)oral	ADMET (Toxicity)	Lethal dose at 50% after oral administration	≥1322.88
MIC (nM)	Antibacterial activity	Minimum inhibitory concentration against enterococci.	≤17000.00
MIC_{50} (nM)	Antibacterial activity	Minimum inhibitory concentration at which 50% of isolates belonging to the genus *Enterococcus* are inhibited	≤6529.34
MIC_{90} (nM)	Antibacterial activity	Minimum inhibitory concentration at which 90% of isolates belonging to *Enterococcus* are inhibited	≤16541.75
Papp (nm/s)	ADMET (Absorption)	Permeability	≥60.00
$T_{1/2}$ (hr)iv-LA	ADMET (Elimination)	Half life after intravenous administration in laboratory animals (*Mus musculus* and/or *Rattus norvegicus*)	≥1.00
$T_{1/2}$ (hr)oral-H	ADMET (Elimination)	Half life after oral administration in humans	≥4.00
$T_{1/2}$ (hr)oral-LA	ADMET (Elimination)	Half life after oral administration in laboratory animals (*Mus musculus* and/or *Rattus norvegicus*)	≥1.60
TD_{50} (µmol/kg)ip	ADMET (Toxicity)	Toxic dose at 50% after intraperitoneal administration	≥416.69
Vd_{ss} (L/kg)iv	ADMET (Distribution)	Volume of distribution at steady state after intravenous administration	≥1.00

[a] The cutoff values represent the conditions under which a compound/case was assigned to the group of the positive cases.

By definition, a time series is a sequence of data points, which are typically measured over an interval of time. In this context, some formulations based on the Box-Jenkins approach transform any series by subtracting the mean of the series from the value of each data point. Therefore, the deviation term, *i.e.*, the difference between any value of a data point and the mean of the group (to which that data point belongs) is a Box-Jenkins moving average.[57-58] The application of the aforementioned approach permits to create new molecular descriptors that can account for both the chemical structure, and diverse elements of the experimental condition/ontology c_j under which compounds have been assayed. Thus, we can write an equation with the following form:

$$avgPq_k(\mathrm{x})c_j = \frac{1}{n(c_j)} \sum_{m=1}^{n(c_j)} Pq_k(\mathrm{x})_m \qquad (3)$$

In **Eq. 3**, $avgPq_k(x)c_j$ is the arithmetic mean of the $Pq_k(x)_m$ descriptors for all the mth compounds in a subset $n(c_j)$. In this context, $n(c_j)$ is the number of compounds assayed by considering the same element of the experimental condition c_j, being also annotated as positive. It is important to emphasize that **Eq. 3** was applied to each element of c_j. The subsequent equation can be written according to the following formalism:

$$\Delta Pq_k(\mathrm{x})c_j = Pq_k(\mathrm{x}) - avgPq_k(\mathrm{x})c_j \qquad (4)$$

Here, **Eq. 4** shows the deviation terms $\Delta Pq_k(x)$, which are the Box-Jenkins moving averages, and they take into account both the chemical structure of a compound and the biological target against which the compound was assayed. The training set was used to construct the mtk-QSBER model, containing 35212 cases, with 18347 positive and 16865 negative. The prediction (test) set was employed to validate the model, being formed by 11610 cases, 6052 positive and 5558 negative. Linear discriminant analysis (LDA) was used as pattern classification technique to find the best model, using a forward step-wise procedure as variable selection strategy. In order to accomplish this task, the program STATISTICA was used.[59] The mtk-QSBER model follows the expression of the form:

$$BE_i(c_j) = a_0 + \sum [b \times \Delta Pq_k(\mathrm{x})c_j] \qquad (5)$$

In **Eq. 5**, a_0 is the constant, and b_i represents the coefficients of the variables. It should be emphasized that the program STATISTICA takes the categorical variable $BE_i(c_j)$, and transforms it into a real score that predict the propensity of a drug/ chemical i to exhibit certain biological effect under the experimental condition c_j. After, that score is transformed to the predicted categorical value of $BE_i(c_j)$. In order to analyze the quality and predictive power of the model, several statistical indices such as Wilks's lambda (λ), chi-square (χ^2), p-value sensitivity, specificity, accuracy, Matthews correlation coefficient (MCC), and the receiver operating characteristic (ROC) curves were calculated.[60-62] The last five statistical indices were determined for both training and prediction (test) sets.

Results and discussion
Mtk-QSBER model
With the aim of finding the most appropriate model, the principle of parsimony was applied. This means that the model exhibiting the highest statistical quality, but with few descriptors as possible was selected. In this sense, the best mtk-QSBER model found by us had five descriptors:

$$BE_i(c_j) = 0.409\Delta Pq_2(R)m_e - 6.233\Delta Pq_0(H)b_t$$
$$+ 4.405\Delta Pq_2(H)a_i - 0.019\Delta Pq_2(PSA)t_m$$
$$+ 0.022\Delta Pq_5(PSA)l_c + 1.102$$

$$N = 35212 \quad \lambda = 0.431 \quad \chi^2 = 29662.10 \quad p\text{-}value < 0.001 \quad (6)$$

All the symbols of the molecular descriptors together with their corresponding definitions can be found in **Table 2**. The relatively small values of λ and p-level, and the large χ^2, demonstrate the statistical quality of our mtk-QSBER model.[60]

Table 2. List of molecular descriptors which entered in the final mtk-QSBER model.

| Descriptor	Concept
$\Delta Pq_2(R)m_e$	Deviation of the mutual probability quadratic index of order 2, weighted by the refractivity, depending on the molecular structure and the measure of biological effect
$\Delta Pq_0(H)b_t$	Deviation of the mutual probability quadratic index of order 0, weighted by the hydrophobicity, depending on the molecular structure and the biological target
$\Delta Pq_2(H)a_i$	Deviation of the mutual probability quadratic index of order 2, weighted by the hydrophobicity, depending on the molecular structure and the assay information
$\Delta Pq_2(PSA)t_m$	Deviation of the mutual probability quadratic index of order 2, weighted by the polar surface area, depending on the molecular structure and the target mapping
$\Delta Pq_5(PSA)_c$	Deviation of the mutual probability quadratic index of order 5, weighted by the polar surface area, depending on the molecular structure and the level of curation of the experimental information

The mtk-QSBER model correctly classified 33755 out of 35212 cases, with an accuracy of 95.86% in the training set, while in the prediction (test) set, 11098 out of 11610 cases were correctly classified, with an accuracy of 95.59%. Specific details regarding the percentages of correct classification are depicted in **Table 3**, while other important information concerning the chemical and biological data of all molecules, as well as their respective classifications can be found in Supplementary Information 1 (**Suppl. Inf. 1**) upon request to the authors.

Table 3. Performance of the mtk-QSBER model.

CLASSIFICATION[a,b]	Training set	Prediction set
NC_{Total}	35212	11610
$NC_{Positive}$	18347	6052
$CCC_{Positive}$	17580	5791
Sens (%)[c]	95.82	95.69
$NC_{Negative}$	16865	5558
$CCC_{Negative}$	16175	5307
Spec (%)[d]	95.91	95.48
Acc (%)[e]	95.86	95.59
MCC[f]	0.917	0.912

[a]NC – Number of cases. [b]CCC – Correctly classified cases. [c]Sensitivity. [d]Specificity. [e]Accuracy. [f]Mathew's correlation coefficient.

Furthermore, the descriptors of type *avgPq$_k$(x)c$_j$* depending on the elements m_e, b_t, a_i, t_m, and l_c appear in Supplementary Information 2 (**Suppl. Inf. 2**) upon request to the authors. As final evidences of the quality and predictive power of the mtk-QSBER model, the areas under the ROC curves were determined, showing a value of 0.994 for both training and prediction sets (**Fig. 2**). This value demonstrates that our mtk-QSBER model is very different from a random classifier (area = 0.5) because the areas under the ROC curves are much larger. By analyzing **Table 3**, the ROC curves, as well as **Suppl. Inf. 1**, it is intuitive to see that the mtk-QSBER model developed here has an excellent performance, which is comparable with other reports of literature devoted to the use of mt-QSAR/mtk-QSBER models.[41-47]
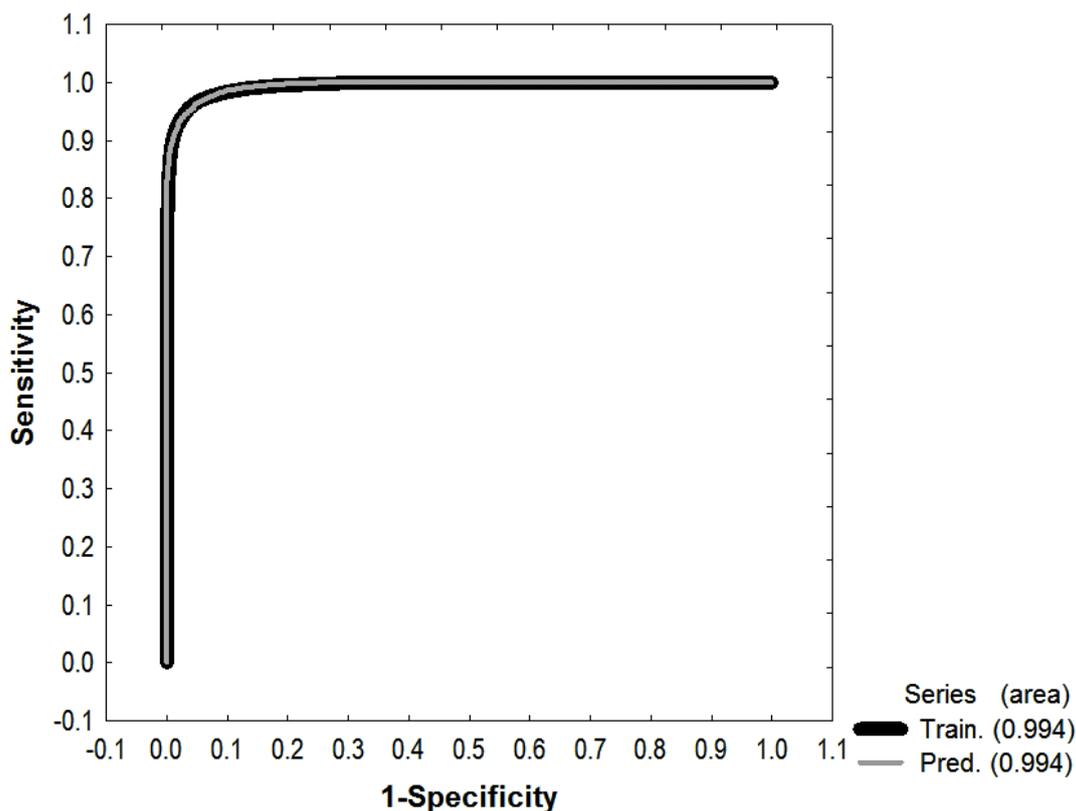
Fig. 2. Pictorial representation of the areas under the ROC curves.

An interesting and peculiar detail of our mtk-QSBER model represented by **Eq. 5** is that the molecular descriptors can be interpreted in terms of simple physicochemical and/or structural properties. First, it is necessary to emphasize that all these molecular descriptors based on mutual probabilities indicate that the global property of a molecule will be influenced by the atomic contributions expressed as quadratic functions of the physicochemical properties, depending on the number of times (occurrences) in which certain connections (bonds) appear in the whole molecule. For this reason, it is intuitive to see that the molecular descriptors involved in the construction of the mtk-QSBER model consider the atoms with their corresponding chemical environments. Bond multiplicity is also accounted for by these descriptors. Thus, **ΔPq$_2$(R)m$_e$** describes the increment in the molecular refractivity, and consequently the increment in the molecular polarizability and/or size depending on both the chemical structure and the measure of the biological effect. This increment in the physicochemical property mentioned above should occur in regions where atoms are placed at topological distances equal to 2 (two bonds between the atoms). The descriptor **ΔPq$_0$(H)b$_t$** represents the diminution of the global hydrophobicity, taking into consideration the structure of the molecules, and the biological targets against which they were tested. This last descriptor is constrained

7

by the variable $\Delta Pq_2(H)a_i$ which expresses the increment of the hydrophobicity in regions where atoms are placed at topological distances equal to 2. Thus, $\Delta Pq_2(H)a_i$ characterizes the structure of the molecules and the assay information, which means that the experiments exhibit different hydrophobicity requirements when they are carried out by assessing the affinity (binding), measuring effects of the compound on a pathway, system or whole organism (functional), or by determining ADMET properties that involving key metabolic enzymes, cells, tissues and even organisms.

The increment in the hydrophobicity explained by $\Delta Pq_2(H)a_i$ is consistent with the diminution of the polar surface area in the same molecular regions expressed by the descriptor $\Delta Pq_2(PSA)t_m$, which depends on the chemical structure and the target mapping, *i.e.*, the degree of knowledge whether an assay is intended to a general type of biological target. Finally, the diminution of the polar surface area is confirmed by the descriptor $\Delta Pq_5(PSA)l_c$, which characterizes the molecular regions where any two atoms are placed at topological distances equal to 5. This descriptor provides information about the variation in the molecular structure, and the level of curation of the biological tests.

Oritavancin. Prediction of multiple biological effects

Until now, from the analysis of **Eq. 5**, tables, and supplementary materials, we have demonstrated that our mtk-QSBER model can integrate different kinds of chemical and biological data. In fact, in the dataset used to construct the model, there are many chemical families of compounds, where multiple biological effects associated with anti-enterococci activities and ADMET profiles have been predicted by considering dissimilar experimental conditions c_j. Anyway, the purpose here is to show the practical applicability of the mtk-QSBER model. In this sense, we performed simultaneous prediction of many biological effects for the investigational antibacterial drug oritavancin (**Fig. 3**), which was originally discovered and developed by Eli Lilly. After, The Medicines Company was running clinical trials for a possible new FDA (Food and Drug Administration) application in 2013.[63] Very recently, this antibacterial drug has been approved by FDA in 2014 for treatment of skin infections in the United States.
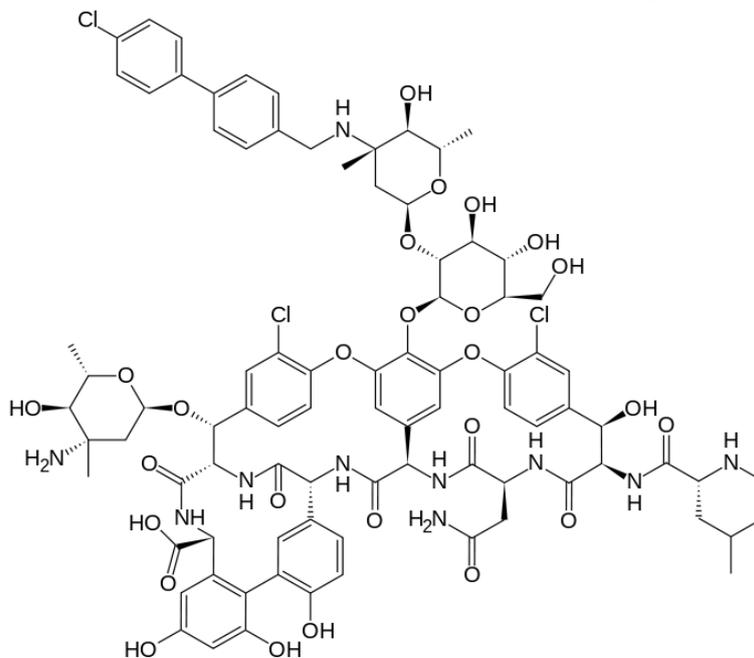
Fig. 3. Chemical structure of oritavancin.

Oritavancin has exhibited high antibacterial activity against sensitive, drug resistant, MDR strains. More specifically, this antibacterial drug has been reported with values of MIC_{90} = 1.00 µg/ml (557.69 nM) against different strains belonging to the genus *Enterococcus*, containing dissimilar degrees of resistance against vancomycin.[64] The same value of has been reported against *Enterococcus faecium* exhibiting resistance to vancomycin and ciprofloxacin.[65] Other reports indicate that oritavancin showed MIC_{50} = 0.25 µg/ml (139.42 nM) and MIC_{90} = 0.50 µg/ml (278.85 nM) against diverse enterococci strains.[66] According to all these experimental evidences, if the cutoff values depicted in **Table 1** are used, then, oritavancin should be classified as positive (active against enterococci).

In the case of preclinical studies, experimental results show that the half-life ($t_{1/2}$) of oritavancin in rats with venous catheter-associated infection was 10 hours.[67] At the same time, the $t_{1/2}$ in neutropenic mouse infection model was of 33 hours, and the area under the curve (AUC) of 562.17 µg.hr/ml (313.52 µM.hr).[68] All these parameters were determined after intravenous administration in laboratory animals. We did not find data reported for oritavancin in healthy laboratory animals. However, by analyzing the experimental evidences, and the criteria regarding the use of cutoff values in **Table 1**, we can assume that with the large values of half-life times and AUC, oritavancin should be classified as positive (safe) in the case of the ADMET parameters mentioned above in healthy mice and rats. On the other hand, a review devoted to clinical studies indicated that after intravenous administration in healthy human volunteers, oritavancin had a value of volume of distribution at steady state (V_{ss}) as large as 1.92 L/kg, with $t_{1/2}$ = 356 hours, and AUC = 1111 µg.hr/ml (619.60 µM.hr).[69] These experimental ADMET values clearly demonstrate that according to the cutoff values, the investigational antibacterial drug mentioned above can be considered as positive, *i.e.*, oritavancin can be considered as a safe drug.

Predictions of diverse biological effects of oritavancin were realized by using the mtk-QSBER model, where 1311 different experimental conditions (combinations of the elements of c_i) were considered. Results of these predictions are depicted in Supplementary Information 3 (**Suppl. Inf. 3**) upon request to the authors. These results suggest that oritavancin is very active against drug sensitive, and MDR strains belonging to different enterococci, including *Enterococcus faecium* and *Enterococcus faecalis*. Regarding the ADMET profiles in preclinical studies, the analysis of our predictions confirm that oritavancin exhibits desirable pharmacokinetic parameters, which strongly converge with the experimental reports. We could not find toxicological data for this drug, but the virtual assessment of several measures of toxicity such as LD_{50} and TD_{50} confirmed that the appearance of toxic effect depends on the breed of mice and/or rats, which were used in the assays.

Finally, very useful information was extracted from the predictions realized for ADMET profiles in clinical studies. Thus, our analysis demonstrates the safety of oritavancin because this drug was predicted to have good absorption and bioavailability (including measures such as Papp, F(%), and AUC), with excellent volume of distribution, and good elimination. In terms of metabolism, the predictions indicate that oritavancin is not metabolized major cytochromes P450 (CYPs) such as CYP1A2, CYP2C19, CYP2C9, CYP2D6 and CYP3A4. By taking into account our predictions, only few CYPs and other metabolizing enzymes are inhibited by oritavancin. These theoretical results complement the experiments, which have demonstrated that the metabolism of the drug mentioned above is very limited.[70] Consequently, metabolism of oritavancin should be analyzed with caution.

Conclusion

With the fast emergence of drug resistant enterococci strains, more innovative approaches for rational discovery of antibacterial agents are needed. As an alternative to overcome this problem, we have applied a chemoinformatic methodology through the generation of an mtk-QSBER model based on probabilistic quadratic indices. Our model was devoted to the virtual search for potent and safer anti-enterococci agents. The predictions of multiple biological effects performed over the antibacterial drug oritavancin demonstrate that the present mtk-QSBER model can serve as a guide for pharmaceutical and medicinal chemists through the different stages in drug discovery: from *in vitro* tests to preclinical and clinical studies. Our work also suggests the possibility of extending the present chemoinformatic methodology to integrate other pharmacological activities with dissimilar ADMET profiles. This constitutes a new horizon regarding the application of promising and innovative *in silico* tools in medicinal chemistry to support the design of new molecular entities with desired properties.

Acknowledgments

References

[1]   K. Fisher, C. Phillips, *Microbiology* **2009**, *155*, 1749-1757.
[2]   K. J. Ryan, C. G. Ray, *Sherris Medical Microbiology. An Introduction to infectious diseases*, McGraw-Hill Companies, Inc, Arizona, **2004**.
[3]   P. S. Brachman, E. Abrutyn, *Bacterial Infections of Humans: Epidemiology and Control*, Springer Science+Business Media, LLC, New York, NY, **2009**.
[4]   L. Rigottier-Gois, A. Alberti, A. Houel, J. F. Taly, P. Palcy, J. Manson, D. Pinto, R. C. Matos, L. Carrilero, N. Montero, M. Tariq, H. Karsens, C. Repp, A. Kropec, A. Budin-Verneuil, A. Benachour, N. Sauvageot, A. Bizzini, M. S. Gilmore, P. Bessieres, J. Kok, J. Huebner, F. Lopes, B. Gonzalez-Zorn, A. Hartke, P. Serror, *PLoS One* **2011**, *6*, e29023.
[5]   C. A. Arias, G. A. Contreras, B. E. Murray, *Clin Microbiol Infect* **2010**, *16*, 555-562.
[6]   S. S. Printsevskaya, M. I. Reznikova, A. M. Korolev, G. B. Lapa, E. N. Olsufyeva, M. N. Preobrazhenskaya, J. J. Plattner, Y. K. Zhang, *Future Med Chem* **2013**, *5*, 641-652.
[7]   T. L. Lemke, D. A. Williams, V. F. Roche, S. W. Zito (Eds.), *Foye's Principles of Medicinal Chemistry*, Lippincott Williams & Wilkins, Baltimore, Philadelphia, **2008**.
[8]   W. Jahnke, D. A. Erlanson, *Fragment-based approaches in drug discovery*, Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany, **2006**.
[9]   R. T. Borchardt, E. H. Kerns, M. J. Hageman, D. R. Thakker, J. L. Stevens (Eds.), *Optimizing the "Drug-Like" Properties of Leads in Drug Discovery*, Springer Science+Business Media, LLC, New York, NY, **2006**.
[10] S. Croes, A. H. Koop, S. A. van Gils, C. Neef, *Eur J Pharm Sci* **2012**, *45*, 90-100.
[11] J. Hau, S. J. Schapiro, *Handbook of Laboratory Animal Science: Essential Principles and Practices*, CRC Press, Taylor & Francis Group, LLC, Boca Raton, FL, **2011**.
[12] J. Gasteiger, *Handbook of chemoinformatics*, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, **2003**.
[13] T. Oprea, *Chemoinformatics in drug discovery*, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, **2005**.

[14] R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics*, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, **2009**.

[15] I. Mitra, A. Saha, K. Roy, *Future Med Chem* **2013**, *5*, 261-280.

[16] G. Marzaro, A. Chilin, A. Guiotto, E. Uriarte, P. Brun, I. Castagliuolo, F. Tonus, H. Gonzalez-Diaz, *Eur J Med Chem* **2011**, *46*, 2185-2192.

[17] H. Gonzalez-Diaz, F. Prado-Prado, X. Garcia-Mera, N. Alonso, P. Abeijon, O. Caamano, M. Yanez, C. R. Munteanu, A. Pazos, M. A. Dea-Ayuela, M. T. Gomez-Munoz, M. M. Garijo, J. Sansano, F. M. Ubeira, *J Proteome Res* **2011**, *10*, 1698-1718.

[18] A. Speck-Planche, V. V. Kleandrova, F. Luan, M. N. D. S. Cordeiro, *Anticancer Agents Med Chem* **2012**, *12*, 678-685.

[19] A. Speck-Planche, V. V. Kleandrova, F. Luan, M. N. D. S. Cordeiro, *Eur J Med Chem* **2011**, *46*, 5910-5916.

[20] F. J. Prado-Prado, F. M. Ubeira, F. Borges, H. Gonzalez-Diaz, *J Comput Chem* **2010**, *31*, 164-173.

[21] F. J. Prado-Prado, X. Garcia-Mera, H. Gonzalez-Diaz, *Bioorg. Med. Chem.* **2010**, *18*, 2225-2231.

[22] I. Garcia, Y. Fall, G. Gomez, H. Gonzalez-Diaz, *Mol Divers* **2011**, *15*, 561-567.

[23] A. Speck-Planche, V. V. Kleandrova, F. Luan, M. N. D. S. Cordeiro, *Curr Alzheimer Res* **2013**, *10*, 117-124.

[24] A. Speck-Planche, F. Luan, M. N. D. S. Cordeiro, *Curr Top Med Chem* **2012**, *12*, 2745-2762.

[25] A. Speck-Planche, V. V. Kleandrova, *Mol Divers* **2012**, *16*, 183-191.

[26] F. J. Prado-Prado, E. Uriarte, F. Borges, H. Gonzalez-Diaz, *Eur J Med Chem* **2009**, *44*, 4516-4521.

[27] F. J. Prado-Prado, F. M. Ubeira, F. Borges, H. Gonzalez-Diaz, *J Comput Chem* **2009**, *31*, 164-173.

[28] C. R. Munteanu, A. L. Magalhaes, E. Uriarte, H. Gonzalez-Diaz, *J Theor Biol* **2009**, *257*, 303-311.

[29] P. Riera-Fernandez, R. Martin-Romalde, F. J. Prado-Prado, M. Escobar, C. R. Munteanu, R. Concu, A. Duardo-Sanchez, H. Gonzalez-Diaz, *Curr Top Med Chem* **2012**, *12*, 927-960.

[30] F. Prado-Prado, X. Garcia-Mera, P. Abeijon, N. Alonso, O. Caamano, M. Yanez, T. Garate, M. Mezo, M. Gonzalez-Warleta, L. Muino, F. M. Ubeira, H. Gonzalez-Diaz, *Eur J Med Chem* **2011**, *46*, 1074-1094.

[31] F. J. Prado-Prado, F. Borges, E. Uriarte, L. G. Perez-Montoto, H. Gonzalez-Diaz, *Anal. Chim. Acta* **2009**, *651*, 159-164.

[32] R. Concu, M. A. Dea-Ayuela, L. G. Perez-Montoto, F. Bolas-Fernandez, F. J. Prado-Prado, G. Podda, E. Uriarte, F. M. Ubeira, H. Gonzalez-Diaz, *J Proteome Res* **2009**, *8*, 4372-4382.

[33] F. J. Prado-Prado, H. Gonzalez-Diaz, O. M. de la Vega, F. M. Ubeira, K. C. Chou, *Bioorg. Med. Chem.* **2008**, *16*, 5871-5880.

[34] A. Speck-Planche, F. Luan, M. N. D. S. Cordeiro, *Curr Med Chem* **2012**, *19*, 1635-1645.

[35] A. Speck-Planche, V. V. Kleandrova, M. T. Scotti, *Chemometr Intell Lab Syst* **2012**, *111*, 39-45.

[36] A. Speck-Planche, V. V. Kleandrova, F. Luan, M. N. D. S. Cordeiro, *Eur J Pharm Sci* **2012**, *47*, 273-279.

[37] A. Speck-Planche, V. V. Kleandrova, F. Luan, M. N. D. S. Cordeiro, *Mol Biosyst* **2012**, *8*, 2188-2196.

[38] A. Speck-Planche, V. V. Kleandrova, *Curr Top Med Chem* **2012**, *12*, 1734-1747.

[39] E. Molina, E. Sobarzo-Sanchez, A. Speck-Planche, M. J. Matos, E. Uriarte, L. Santana, M. Yanez, F. Orallo, *Mini Rev Med Chem* **2012**, *12*, 947-958.

[40] A. Speck-Planche, V. V. Kleandrova, F. Luan, M. N. D. S. Cordeiro, *Bioorg. Med. Chem.* **2011**, *19*, 6239-6244.

[41] F. Luan, M. N. D. S. Cordeiro, N. Alonso, X. Garcia-Mera, O. Caamano, F. J. Romero-Duran, M. Yanez, H. Gonzalez-Diaz, *Bioorg. Med. Chem.* **2013**, *21*, 1870-1879.

[42] E. Tenorio-Borroto, C. G. Penuelas Rivas, J. C. Vasquez Chagoyan, N. Castanedo, F. J. Prado-Prado, X. Garcia-Mera, H. Gonzalez-Diaz, *Bioorg. Med. Chem.* **2012**, *20*, 6181-6194.

[43] A. Speck-Planche, V. V. Kleandrova, M. N. D. S. Cordeiro, *Eur J Pharm Sci* **2013**, *48*, 812-818.

[44] A. Speck-Planche, V. V. Kleandrova, M. N. D. S. Cordeiro, *Bioorg. Med. Chem.* **2013**, *21*, 2727-2732.

[45] A. Speck-Planche, M. N. D. S. Cordeiro, *ACS Comb Sci* **2014**, *16*, 78-84.

[46] A. Speck-Planche, M. N. D. S. Cordeiro, *Future Med Chem* **2014**, *6*, 2013-2028.

[47] A. Speck-Planche, M. N. D. S. Cordeiro, *Curr Top Med Chem* **2013**, *13*, 1656-1665.

[48] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light, S. McGlinchey, D. Michalovich, B. Al-Lazikani, J. P. Overington, *Nucleic Acids Res* **2012**, *40*, D1100-1107.

[49] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, *J Cheminform* **2011**, *3*, 33.

[50] Y. Marrero-Ponce, J. R. Valdés-Martini, C. R. García-Jacas, Y. Silveira-Vaz 'd Almeida, C. Morell, N. Cubillan, *TOMOCOMD-CARDD. Module QUBILs-MAS*, v1.0, Santa Clara, Villa Clara, **2012**.

[51] Y. Marrero-Ponce, *Molecules* **2003**, *8*, 687-726.

[52] Y. Marrero-Ponce, *J Chem Inf Comput Sci* **2004**, *44*, 2010-2026.

[53] Y. Marrero-Ponce, J. A. Castillo-Garit, E. Olazabal, H. S. Serrano, A. Morales, N. Castanedo, F. Ibarra-Velarde, A. Huesca-Guillen, E. Jorge, A. del Valle, F. Torrens, E. A. Castro, *J Comput Aided Mol Des* **2004**, *18*, 615-634.

[54] Y. Marrero-Ponce, J. A. Castillo-Garit, *J Comput Aided Mol Des* **2005**, *19*, 369-383.

[55] J. A. Castillo-Garit, Y. Marrero-Ponce, F. Torrens, *Bioorg. Med. Chem.* **2006**, *14*, 2398-2408.

[56] G. M. Casanola-Martin, Y. Marrero-Ponce, M. T. Khan, A. Ather, S. Sultan, F. Torrens, R. Rotondo, *Bioorg. Med. Chem.* **2007**, *15*, 1483-1503.

[57] T. Hill, P. Lewicki, *STATISTICS Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining*, StatSoft, Tulsa, **2006**.

[58] G. Box, G. Jenkins, *Time series analysis: Forecasting and control*, Holden-Day, San Francisco, **1970**.

[59] Statsoft-Team, *STATISTICA. Data analysis software system*, v6.0, Tulsa, **2001**.

[60] C. J. Huberty, S. Olejnik, *Applied MANOVA and discriminant analysis*, John Wiley & Sons, Inc., Hoboken, New Jersey, **2006**.

[61] A. Speck-Planche, V. V. Kleandrova, F. Luan, M. N. D. S. Cordeiro, *Bioorg. Med. Chem.* **2012**, *20*, 4848-4855.

[62] B. Hanczar, J. Hua, C. Sima, J. Weinstein, M. Bittner, E. R. Dougherty, *Bioinformatics* **2010**, *26*, 822-830.

[63] A. Belley, G. A. McKay, F. F. Arhin, I. Sarmiento, S. Beaulieu, I. Fadhil, T. R. Parr, Jr., G. Moeck, *Antimicrob. Agents Chemother.* **2010**, *54*, 5369-5371.

[64] F. F. Arhin, D. C. Draghi, C. M. Pillar, T. R. Parr, Jr., G. Moeck, D. F. Sahm, *Antimicrob. Agents Chemother.* **2009**, *53*, 4762-4771.

[65] S. Noviello, F. Ianniello, S. Esposito, *J Antimicrob Chemother* **2001**, *48*, 283-286.

[66] R. E. Mendes, L. N. Woosley, D. J. Farrell, H. S. Sader, R. N. Jones, *Antimicrob. Agents Chemother.* **2012**, *56*, 1639-1642.

[67] M. E. Rupp, P. D. Fey, G. M. Longo, *J Antimicrob Chemother* **2001**, *47*, 705-707.

[68] C. J. Boylan, K. Campanale, P. W. Iversen, D. L. Phillips, M. L. Zeckel, T. R. Parr, Jr., *Antimicrob. Agents Chemother.* **2003**, *47*, 1700-1706.

[69] D. R. Guay, *Pharmacotherapy* **2004**, *24*, 58-68.

[70] S. M. Bhavnani, J. S. Owen, J. S. Loutit, S. B. Porter, P. G. Ambrose, *Diagn Microbiol Infect Dis* **2004**, *50*, 95-102.