ACS Combinatorial Science

Research Article

pubs.acs.org/acscombsci

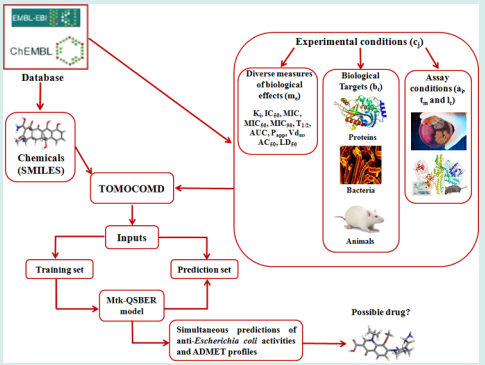
Simultaneous Virtual Prediction of Anti-*Escherichia coli* Activities and ADMET Profiles: A Chemoinformatic Complementary Approach for High-Throughput Screening

Alejandro Speck-Planche* and M. N. D. S. Cordeiro*

REQUIMTE/Department of Chemistry and Biochemistry, University of Porto, 4169-007 Porto, Portugal

Supporting Information

ABSTRACT: *Escherichia coli* remains one of the principal pathogens that cause nosocomial infections, medical conditions that are increasingly common in healthcare facilities. *E. coli* is intrinsically resistant to many antibiotics, and multidrug-resistant strains have emerged recently. Chemoinformatics has been a great ally of experimental methodologies such as high-throughput screening, playing an important role in the discovery of effective antibacterial agents. However, there is no approach that can design safer anti-*E. coli* agents, because of the multifactorial nature and complexity of bacterial diseases and the lack of desirable ADMET (absorption, distribution, metabolism, elimination, and toxicity) profiles as a major cause of disapproval of drugs. In this work, we introduce the first multitasking model based on quantitative−structure biological effect relationships (mtk-QSBER) for simultaneous virtual prediction of anti-*E. coli* activities and ADMET properties of drugs and/or chemicals under many experimental conditions. The mtk-QSBER model was developed from a large and heterogeneous data set of more than 37800 cases, exhibiting overall accuracies of >95% in both training and prediction (validation) sets. The utility of our mtk-QSBER model was demonstrated by performing virtual prediction of properties for the investigational drug avarofloxacin (AVX) under 260 different experimental conditions. Results converged with the experimental evidence, confirming the remarkable anti-*E. coli* activities and safety of AVX. Predictions also showed that our mtk-QSBER model can be a promising computational tool for virtual screening of desirable anti-*E. coli* agents, and this chemoinformatic approach could be extended to the search for safer drugs with defined pharmacological activities.



KEYWORDS: *antibacterial, ADMET, linear discriminant analysis, mtk-QSBER, TOMOCOMD*

INTRODUCTION

Resistance of bacteria to current antibiotics is one of the most alarming problems worldwide, affecting even healthcare facilities, which have been remarkably impacted. Thus, health compromising situations have emerged, nosocomial infections.[1] In this sense, one of the most common and dangerous pathogens is *Escherichia coli*, a Gram-negative bacterium that belongs to the family Enterobacteriaceae. *E. coli* can cause several diseases in humans, including peritonitis, septicemia, gastroenteritis, and urinary tract infections, the last being one of the main nosocomial diseases.[2] As in the case of many other Gram-negative bacteria, *E. coli* displays intrinsic resistance to several antibiotics because of its relatively impermeable cell wall.[2] In fact, in recent years, several resistant strains have emerged even in highly developed countries.[3] For this reason, the search for more potent antibacterial agents against the bacterium mentioned above should be an aspect of particular interest in antimicrobial research.

Despite the advances of science and technology, drug discovery remains a slow, expensive, and inefficient process with a low rate of new therapeutic discovery, taking 15−17 years with a cost of approximately US$1778 million.[4] Despite the existence of promising experimental methodologies such as high-throughput screening (HTS), it is not possible to cover the huge molecular space (10^{63} small to medium size molecules).[5] For this reason, while the number of hits have substantially increased with the use of HTS, no corresponding growth in the number of antibacterial (or other) drugs entering the market has been observed.[6] Therefore, it is mandatory to continue the integration of HTS with disciplines involving virtual screening.[7] In this sense, chemoinformatics has served as an essential support for experimental methodologies such as HTS, helping to rationalize the chemical synthesis and contributing to diminish the length of time and cost of the experiments.[8] Regardless, the battle against *E. coli* will depend on the discovery of new and potent antibacterial chemicals, displaying also ADMET (absorption, distribution, metabolism, elimination, and toxicity) profiles that are as desirable as possible. From one side, in the field related to the discovery of antibacterial agents, some relevant works have been reported, dealing with synthesis, evaluation, and *in silico* analysis of anti-*E.*

Received: September 10, 2013
Published: January 2, 2014

Table 1. Cutoff Values for Diverse Measures of Biological Effects

measure of effect (units)	biological profile	concept	cutoff value[a]
AC_{50} (nM)	ADMET (metabolism)	concentration to elicit 50% of the maximal effect	≤6309.58
AUC (μM h), 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 h), oral-H	ADMET (bioavailability, elimination)	area under the curve after oral administration in humans	≥53.45
AUC (μM h), oral-LA	ADMET (bioavailability, elimination)	area under the curve after oral administration in laboratory animals (*M. musculus* and/or *R. norvegicus*)	≥15.00
F (%), oral	ADMET (bioavailability)	oral bioavailability assessed as the fraction of an orally administered drug that reaches systemic circulation	≥60.00
IC_{50} (nM), ip	antibacterial activity	concentration required for 50% inhibition of the activity of a protein or enzyme present in *E. coli*	≤200.00
K_i (nM)	antibacterial activity	inhibition constant associated with a protein or enzyme present in *E. coli*	≤250.00
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	≥1541.75
MIC (nM)	antibacterial activity	minimal inhibitory concentration against *E. coli*	≤20000.00
MIC_{50} (nM)	antibacterial activity	minimal inhibitory concentration at which 50% of isolates belonging to *E. coli* are inhibited	≤9000.30
MIC_{90} (nM)	antibacterial activity	minimal inhibitory concentration at which 90% of isolates belonging to *E. coli* are inhibited	≤17071.41
Papp (nm/s)	ADMET (absorption)	permeability	≥60.00
$T_{1/2}$ (h), iv-LA	ADMET (elimination)	half-life after intravenous administration in laboratory animals (*M. musculus* and/or *R. norvegicus*)	≥1.00
$T_{1/2}$ (h), oral-H	ADMET (elimination)	half-life after oral administration in humans	≥3.00
$T_{1/2}$ (h), oral-LA	ADMET (elimination)	half-life after oral administration in laboratory animals (*M. musculus* and/or *R. norvegicus*)	≥1.52
TD_{50} (μmol/kg), ip	ADMET (toxicity)	toxic dose at 50% after intraperitoneal administration	≥406.54
Vd_{ss} (L/kg), iv	ADMET (distribution)	volume of distribution at steady state after intravenous administration	≥1.00

[a]Condition under which a compound/case is considered to be positive.

coli compounds.[9] However, most of the computer-aided methodologies applied to date have been used against only one target (usually a protein), and the analyses have been realized by employing analogous series of compounds. This fact does not permit a deeper exploration of the chemical space in terms of molecular diversity and complexity. Additionally, advances in systems biology and complexity sciences suggest that complex diseases, including those associated with bacterial pathogens, are multifactorial.[10] On the other hand, when an antibacterial (or any) drug candidate is developed, great concern about the possible undesirable ADMET properties, which constitute a major aspect of disapproval of drugs, is expected.[11] Furthermore, computational approaches applied to the modeling of ADMET profiles exhibit the same disadvantages observed for the design of anti-*E. coli* agents.

Until now, no chemoinformatic model has been able to simultaneously predict anti-*E. coli* activity and ADMET properties. In recent years, several researchers have emphasized the use of methodologies that allow the assessment of multiple pharmacological activities against many biological targets (biomacromolecules, microorganisms, cell lines, etc.),[12] and the integration of different kinds of chemical and biological data.[13] By considering all the ideas mentioned above, we introduce in this work the first multitasking model based on quantitative−-structure biological effect relationships (mtk-QSBER) for simultaneous prediction of anti-*E. coli* activities and ADMET profiles of chemicals.

EXPERIMENTAL PROCEDURES

Database, Molecular Descriptors, and Development of the mtk-QSBER Model. We retrieved a large data set from the highly accurate public source CHEMBL,[14] which is available at http://www.ebi.ac.uk/chembldb/. Our data set contained 37834 cases/endpoints, which are the result of the evaluation of 23705 drugs/chemicals (N_c) under more than one experimental condition c_j. The term "experimental condition" can be considered as an ontology,[13b,d] which has the form $c_j \rightarrow (m_e, b_t, a_i, t_m, l_c)$. In this sense, m_e defines the measure of a specific biological effect (anti-*E. coli* activity, toxicity, half-life time, bioavailability, etc.). The element b_t describes the different biological targets such as strains of *E. coli* and their biomacromolecules (proteins), mice, rats, *Homo sapiens*, etc. On the other hand, and considering the test conditions, a_i focuses on the assay information, i.e., if the test was realized by measuring binding (B), functional (F), or ADMET (A) effects. Also, t_m refers to target mapping, which is the degree of certainty about whether the assay is really intended to measure the effect on a given type of biological target. Finally, the element l_c considers the level of curation/verification of the experimental information of a particular assay. Each combination of the elements m_e, b_t, a_i, t_m, and l_c defines a unique experimental condition c_j. Taking into account all these ideas, we can say that the data set of 37834 cases/endpoints was obtained from 23705 (N_c) different compounds, for which at least one of 18 (Nm_e) measures of biological effects was used, at least one of 148 (Nb_t) biological targets was involved, with at least one of 3 (Na_i) types of assay information, considering at least one of 7 (Nt_m) degrees of target mapping, and with at least one of 3 (Nl_c) levels of curation of experimental information. For the case of the element m_e, several measures of biological effects were found to appear in different units. For this reason, all the values of anti-*E. coli* activities were transformed to nanomoles per liter (nanomolar); the areas under the curves (AUC) were expressed in micromolar per hour, and all toxicity values were converted to micromoles per kilogram.

dx.doi.org/10.1021/co400115s | ACS Comb. Sci. 2014, 16, 78−84

Table 2. Final Variables That Were Used in the mtk-QSBER Model

descriptor	concept
$\Delta q_0^{PSA}(m_e)$	deviation of the quadratic index of order 0 weighted by the polar surface area, depending on the molecular structure and a measure of the biological effect
$\Delta q_2^{Hyd}(a_i)$	deviation of the quadratic index of order 2 weighted by the hydrophobicity, depending on the molecular structure and the assay information
$\Delta q_2^{PSA}(b_t)$	deviation of the quadratic index of order 2 weighted by the polar surface area, depending on the molecular structure and the biological target
$\Delta q_1^{AW}(t_m)$	deviation of the quadratic index of order 1 weighted by the atomic weight, depending on the molecular structure and the target mapping
$\Delta q_5^{R}(l_c)$	deviation of the quadratic index of order 5 weighted by the refractivity, depending on the molecular structure and the level of curation of the experimental information

Each of the 37834 cases was assigned to one of two possible groups related to the biological effect of a defined compound i under a specific condition c_j $[BE_i(c_j)]$. Thus, a compound/case was selected as positive $[BE_i(c_j) = 1]$ when it displayed high anti-*E. coli* activity or any favorable ADMET profile; otherwise, the compound was considered as negative $[BE_i(c_j) = -1]$. All assignments were realized by taking into account certain arbitrary cutoff values of biological effects that are listed in Table 1. For the complete data set, we used a *txt file containing the SMILES codes of the compounds/cases. The extension of this file was changed to *.smi, which was then transformed to *.sdf by using OpenBabel version 2.3.0.[15] For the calculation of molecular descriptors from the *.sdf file, we employed TOMOCOMD-CARDD.[16] Descriptors used here are based on the application of discrete mathematics and linear algebra theory to chemistry,[17] which can be used to characterize the molecular structure. In this study, we selected the descriptors called atom-based quadratic indices (q_k^{PP}), which were calculated from order 0 to 5, considering the hydrophobicity, the polar surface area, the atomic weight, and the refractivity as physicochemical properties (PP). The mathematical background of these descriptors has been explained in several interesting works.[17c,18] Therefore, we will give just the essential ideas. Given a graph-theoretical square matrix $\mathbf{M}$ of $n \times n$ vertices (atom), $q_k(x)$ can be calculated from the kth power of matrix $\mathbf{M}$ by using the following general equation:

$$q_k(\bar{\mathbf{x}}) = [\mathbf{X}^T][\mathbf{M}]^k[\mathbf{X}] = \sum_{i=1}^{n}\sum_{j=1}^{n} {}^k m_{ij} x_i x_j \quad (1)$$

where $\bar{\mathbf{x}}$ is a molecular vector and $[\mathbf{X}^T]$ is the transpose of $[\mathbf{X}]$, which is a column vector ($n \times 1$ matrix) whose components are $x_1, ..., x_n$. In addition, the term ${}^k m_{ij}$ represents the elements of the kth power of $\mathbf{M}$. The components x of the column vector $[\mathbf{X}]$ are used as PP.

The purpose here is to predict the biological effect of any compound depending on its molecular structure and the experimental conditions c_j under which the compound was tested. Then, if we use the original quadratic indices weighted by any physicochemical property (q_k^{PP}), we will not be able to discriminate the biological effect of a compound by varying the different elements of the condition/ontology c_j. For this reason, inspired by the application of the moving average approach (MAA),[19] we introduced new sets of molecular descriptors like q_k^{PP}, according to the following simple equation:

$$\Delta q_k^{PP}(c_j) = q_k^{PP} - \text{avg}q_k^{PP}(c_j) \quad (2)$$

where q_k^{PP} is the quadratic index of order k, weighted by a defined PP. The descriptor $\text{avg}q_k^{PP}(c_j)$ characterizes each set n_j of compounds assayed under the same experimental condition c_j, being calculated as the average of all the q_k^{PP} values for compounds in a subset of n_j, which were considered as positive cases $[BE_i(c_j) = 1]$ in the same element of the ontology c_j. Thus, in the case of the element b_t, the descriptor $\text{avg}q_k^{PP}(c_j)$ for a set n_j of compounds tested against a specific biological target b_t (protein, cell line, etc.) was calculated as the average of the q_k^{PP} values by considering only the subset of n_j, i.e., those compounds that were considered positive $[BE_i(c_j) = 1]$ against that biological target. A similar operation was also realized for elements m_e, a_i, t_m, and l_c. Regardless, in eq 2, the most important element is the descriptor $\Delta q_k^{PP}(c_j)$, which describes both the molecular structure and the experimental condition c_j. Then, only $\Delta q_k^{PP}(c_j)$ descriptors (120 in total) were used to develop the mtk-QSBER model.

The data set containing 37834 cases was randomly split into two series: training and prediction sets. The training set was used to construct the mtk-QSBER model, containing 28383 cases, with 13203 of them considered as positive and 15180 as negative. The prediction (validation) set was employed to demonstrate the predictive power of the model. This set comprised 9451 cases, 4431 positive and 5020 negative. Linear discriminant analysis (LDA) was used as the pattern classification technique to find the best model, using a forward stepwise procedure as variable selection strategy. To perform this task, STATISTICA version 6.0 was used.[20] The mtk-QSBER model has the following general equation:

$$BE_i(c_j) = a_0 + \sum_{j=1}^{5} b_i \times \Delta MD_i(c_j) \quad (3)$$

where $\Delta MD_i(c_j)$ refers to $\Delta q_k^{PP}(c_j)$ and $BE_i(c_j)$ is the real score that represents the propensity of compound i to have a given biological effect under a specific experimental condition c_j. In eq 3, a_0 is the constant, and b_i values can be considered as the coefficients of the variables. The quality of the mtk-QSBER model was analyzed by examining some statistical indices such as Wilks' lambda (λ), χ^2, and the p level.[21] We also calculated the sensitivity (percentage of correct classification for positive cases), specificity (percentage of correct classification for negative cases), accuracy (overall correct classification), and the Mathews correlation coefficient (MCC) as measure of the quality of binary classifications,[22] and the areas under the receiver operating characteristic (ROC) curves.[23]

RESULTS AND DISCUSSION

mtk-QSBER Model. Together with the variable selection strategy mentioned above, we also applied the principle of parsimony, which means that the model with the highest statistical quality but with the fewest possible descriptors should be selected. Thus, the best mtk-QSBER model that was found (a five-variable equation) is given below:

dx.doi.org/10.1021/co400115s | ACS Comb. Sci. 2014, 16, 78−84

$$BE_i(c_j) = 1.216 \times 10^{-3} \times \Delta q_0^{PSA}(m_e) + 0.018\Delta q_2^{Hyd}(a_i) - 9.372 \times 10^{-5} \times \Delta q_2^{PSA}(b_t) - 0.061 \Delta q_1^{AW}(t_m) + 5.530 \times 10^{-5} \times \Delta q_5^{R}(l_c) + 1.281$$

$$N = 28383 \quad \lambda = 0.412 \quad \chi^2 = 25169.57 \quad p < 10^{-16} \tag{4}$$

The symbols and meanings of the different variables (descriptors) used in the model are listed in Table 2. With regard to the different statistical indices mentioned above, we can say that small values of λ and p level and the large χ^2 are indicative of the high statistical quality of our mtk-QSBER model, which could correctly classify 27235 of 28383 cases, for an accuracy of 95.96% in the training set; in the prediction set, 9047 of 9451 cases were rightly classified, for an accuracy of 95.73%. More details about specific percentages of classification appear in Table 3, while all relevant biological and chemical data associated with each case/compound as well as their respective classifications are listed in the first file of Supporting Information. Additionally, all the average descriptors [used to generate those of the type $\Delta q_k^{PP}(c_j)$] that depend on the elements m_e, b_t, a_i, t_m, and l_c, can be found in the second file of the Supporting Information.

Table 3. Performance of the mtk-QSBER Model

classification[a,b]	training set	prediction set
NC_{total}	28383	9451
$NC_{positive}$	13203	4431
$CCC_{positive}$	12679	4235
sens (%)[c]	96.03	95.58
$NC_{negative}$	15180	5020
$CCC_{negative}$	14556	4812
spec (%)[d]	95.89	95.86
acc (%)[e]	95.96	95.73
MCC[f]	0.919	0.914

[a]NC, number of cases. [b]CCC, correctly classified cases. [c]Sensitivity. [d]Specificity. [e]Accuracy. [f]Mathews correlation coefficient.

The areas under the ROC (receiver operating characteristic) curves were analyzed as final evidence of the quality and predictive power of the mtk-QSBER model. The value was 0.992 for both training and prediction sets (Figure 1), demonstrating that our mtk-QSBER model is not a random classifier because the areas under the ROC curves are very different from those obtained by random classifiers (which have an area of 0.5). We need to emphasize that the cases belonging to the prediction set were never used to construct the model. So, from an analysis of Table 3, the ROC curves, and the first and second files of the Supporting Information, it is intuitive to see that the mtk-QSBER model developed here has an excellent quality and predictive power, which is in agreement with other reports from the literature.[12,13]

An important element regarding the creation of our mtk-QSBER model described by eq 4 is that the descriptors that were employed can have simple physicochemical and/or structural information (Table 2). In essence, the biological effect of any compound can be enhanced by increasing the global polar surface area of the whole molecule, which is described by $\Delta q_0^{PSA}(m_e)$, depending on the molecular structure and the measure of the biological effect. At the same time, the descriptor mentioned above is constrained by $\Delta q_2^{PSA}(b_t)$,

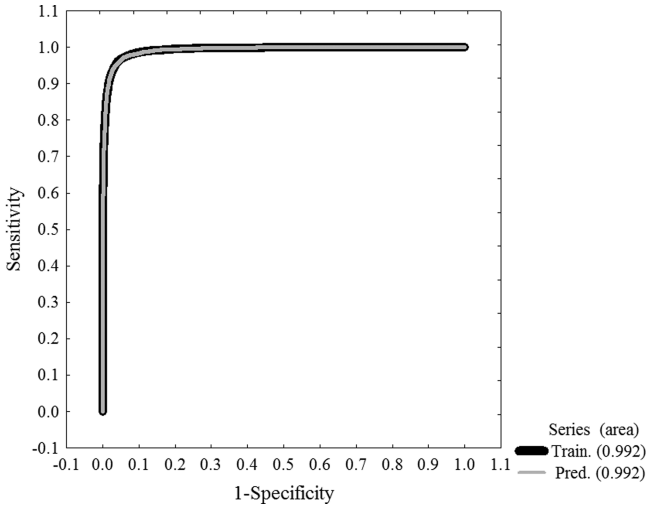


Figure 1. ROC curves obtained from the mtk-QSBER model.

which represents the diminution of the polar surface area in regions where the topological distance (number of bonds) between atoms is two. In this sense, the diminution expressed by $\Delta q_2^{PSA}(b_t)$ will take into consideration both the molecular structure and the biological target. Regardless, the previous ideas about $\Delta q_0^{PSA}(m_e)$ and $\Delta q_2^{PSA}(b_t)$ demonstrate that polar interactions (for example, hydrogen bonds) are important for the increment of a specific biological effect of a molecule under defined experimental conditions. On the other hand, the information provided by the descriptor $\Delta q_2^{Hyd}(a_i)$ is consistent with that described by $\Delta q_2^{PSA}(b_t)$, because the diminution of the polar surface area in the same regions (topological distance equal to two) means an increment in the hydrophobicity. However, in this case, $\Delta q_2^{Hyd}(a_i)$ will depend on the molecular structure and the assay information. Additionally, $\Delta q_1^{AW}(t_m)$ is dependent on the chemical structure and the target mapping (type of biological target). With regard to the appearance and/or enhancement of a biological effect, this descriptor will contain information about the reduction of the global molecular size, considering each atom and its chemical environment formed by the other atoms adjacent to it. Finally, the descriptor $\Delta q_5^{R}(l_c)$ will depend on both the molecular structure and the quality and reliability of the assay, involving a diminution of the polarizability of those molecular regions where the topological distance (number of bonds) between atoms is five.

Avarofloxacin. Virtual Prediction of Diverse Biological Effects. To demonstrate the practical utility of our mtk-QSBER model, we performed a simultaneous virtual prediction of anti-*E. coli* and ADMET profiles for avarofloxacin (AVX) (Figure 2), a potent and versatile antibacterial agent that has undergone phase II clinical trials with excellent results. AVX was discovered by Janssen Pharmaceutica N.V., a unit of the Johnson & Johnson company, and Furiex Pharmaceuticals has

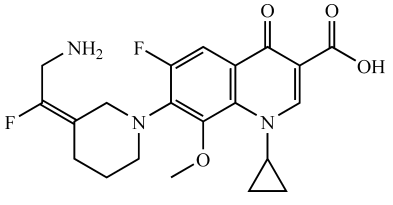


Figure 2. Chemical structure of avarofloxacin (JNJ-Q2).

dx.doi.org/10.1021/co400115s | ACS Comb. Sci. 2014, 16, 78−84

obtained a license to globally develop and commercialize the product. Detailed information regarding the status of this investigational antibacterial drug is available on the Web site of Furiex Pharmaceuticals at http://www.furiex.com/pipeline/discoverydevelopment-pipeline/fluoroquinolone/. AVX, also known as JNJ-Q2, has displayed very high antibacterial activity against both Gram-positive and Gram-negative bacteria, including strains of *E. coli* with different degrees of sensitivity to antibiotics.[24] In this sense, AVX has been reported to have a MIC_{50} of 0.06 μg/mL (143.05 nM) and a MIC_{90} of 0.25 μg/mL (596.06 nM) against susceptible and intermediate strains of *E. coli*, with the highest value of MIC being 0.50 μg/mL (1192.12 nM).[24] In the same report, and for the case of resistant strains of the same bacterium, AVX had a MIC_{50} of 4 μg/mL (9536.98 nM) and a MIC_{90} of 16 μg/mL (38147.92 nM). Thus, in terms of biological activity against *E. coli*, and considering the cutoff values listed in Table 1, AVX can be classified as positive. We did not find preclinical studies of this investigational drug available in the literature. However, clinical assays have been recently reported,[25] showing that AVX had an absolute bioavailability F of 0.65 (65%) and a volume of distribution at steady state of 146.7 L after intravenous administration, which divided by the maximum weight (100 kg) for a healthy human participating in clinical assays yields a values for Vd_{ss} (iv) of 1.467 L/kg. Also, other desirable ADMET parameters were as follows: AUC = 35.8 μg h mL^{-1} after oral administration, which is equivalent to AUC (oral-H) = 85.36 μM h, and $T_{1/2}$ (oral-H) = 14.4 h. According to all these experimental ADMET parameters, and by applying the same criteria of cutoff values listed in Table 1, we should be able to classify AVX as being positive.

We used our mtk-QSBER model to perform virtual prediction of many biological effects of AVX under 260 diverse experimental conditions, which represent combinations of the five elements of the ontology c_j explained in the previous section. Results of these virtual predictions are listed in the third file of the Supporting Information, indicating that AVX is very active against drug susceptible and intermediate strains of *E. coli*. Predictions also suggest that this investigational antibacterial drug may be active against fluoroquinolone-resistant and MDR strains. At the molecular level, and considering the cutoff values listed in Table 1 [IC_{50} (ip) $\leq$ 200.00 nM, and $K_i \leq$ 250.00 nM], AVX was predicted to be positive, i.e., an inhibitor of several proteins, including topoisomerases.

On the other hand, and with regard to ADMET profiles, virtual predictions confirm the drug mentioned above is a safe antibacterial agent. In this sense, AVX was predicted to be positive, displaying good bioavailability, and a volume of distribution at the steady state. Also, AVX was classified as desirable with respect to its AUC (μM h), oral-H, and elimination profiles [$T_{1/2}$ (h), oral-H]. As explained above, preclinical assays are not reported. However, the predictions associated with toxicity profiles such as LD_{50} and TD_{50}, which act as complements of clinical assays, permit us to recognize AVX as safe regardless of the different laboratory animal strains on which the tests were conducted. Similar results of safety were predicted for pharmacokinetic parameters such as AUC and $T_{1/2}$, AVX being predicted to be positive by considering intravenous and oral routes of administration. In terms of metabolism, our calculations suggest that, as expected for fluoroquinolones, several cytochromes (CYP) can be inhibited by AVX. If we take into account all the virtual predictions realized for AVX, we can say that our mtk-QSBER model helped to demonstrate why this investigational drug is ready to be used in phase III clinical trials. At the same time, we are confirming the ability of our mtk-QSBER model to be employed for virtual screening of safer anti-*E. coli* agents, serving as excellent ally to HTS in the field of antibacterial research to rationally filter huge libraries of chemicals and drugs. Consequently, this chemoinformatic approach that was applied to develop the mtk-QSBER model could be extended to diverse areas, permitting integration of diverse pharmacological activities with ADMET properties, in an attempt to search for more effective drugs with desirable profiles.

CONCLUSIONS

Over the centuries, humankind has been the victim of bacterial diseases. Today, novel chemoinformatic approaches are required to rationalize drug discovery in antimicrobial research, to eliminate the alarming resistance of bacteria to current antibiotics. Our unified mtk-QSBER model represents an attempt to face this problem. In this sense, the mtk-QSBER model that was developed by using a large and heterogeneous data set of compounds allowed us to perform simultaneous virtual prediction of anti-*E. coli* activities and ADMET properties. The model, displaying very good accuracy for classification of compounds as positive or negative, was employed with one single but important objective, to help powerful methodologies such as HTS to prioritize and/or filter molecules from the huge chemical space, looking for the structural patterns associated with or related to the biological profiles discussed in our study. This fact was confirmed by performing multiple virtual predictions of different properties of AVX under rigorous conditions, experimental and theoretical results being strongly related. The development of this mtk-QSBER model can be viewed as a new horizon for the design and *in silico* selection of desirable anti-*E. coli* agents.

ASSOCIATED CONTENT

Supporting Information

Additional data

. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

*E-mail: alejspivanovich@gmail.com. Fax: +351 220402659.

*E-mail: ncordeir@fc.up.pt;. Fax: +351 220402659.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

A.S.-P. acknowledges the Portuguese Fundação para a Ciência e a Tecnologia (FCT) and the European Social Found for financial support (Grant SFRH/BD/77690/2011).

REFERENCES

(1) Brachman, P. S.; Abrutyn, E. *Bacterial Infections of Humans: Epidemiology and Control*, 4th ed.; Springer Science+Business Media, LLC: New York, 2009.

(2) Ryan, K. J. Enterobacteriaceae. In *Sherris Medical Microbiology. An Introduction to Infectious Diseases*, 4th ed.; Ryan, K. J., Ray, C. G., Eds.; McGraw-Hill Companies, Inc.: New York, 2004; pp 343−371.

(3) Poirel, L.; Lagrutta, E.; Taylor, P.; Pham, J.; Nordmann, P. Emergence of metallo-β-lactamase NDM-1-producing multidrug-

resistant *Escherichia coli* in Australia. *Antimicrob. Agents Chemother.* **2010**, *54*, 4914−4916.
(4) Paul, S. M.; Mytelka, D. S.; Dunwiddie, C. T.; Persinger, C. C.; Munos, B. H.; Lindborg, S. R.; Schacht, A. L. How to improve R&D productivity: The pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discovery* **2010**, *9*, 203−214.
(5) Jahnke, W.; Erlanson, D. A. *Fragment-based Approaches in Drug Discovery*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2006.
(6) Bleicher, K. H.; Bohm, H. J.; Muller, K.; Alanine, A. I. Hit and lead generation: Beyond high-throughput screening. *Nat. Rev. Drug Discovery* **2003**, *2*, 369−378.
(7) Bajorath, J. Integration of virtual and high-throughput screening. *Nat. Rev. Drug Discovery* **2002**, *1*, 882−894.
(8) (a) Gasteiger, J. *Handbook of Chemoinformatics*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2003. (b) Oprea, T. *Chemoinformatics in Drug Discovery*; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2005; Vol. 23. (c) Bunin, B. A.; Bajorath, J.; Siesel, B.; Morales, G. *Chemoinformatics: Theory, Practice and Products*; Springer: Dordrecht, The Netherlands, 2007.
(9) (a) Fosso, M. Y.; Chan, K. Y.; Gregory, R.; Chang, C. W. Library synthesis and antibacterial investigation of cationic anthraquinone analogs. *ACS Comb. Sci.* **2012**, *14*, 231−235. (b) Liu, T.; Qian, Z.; Xiao, Q.; Pei, D. High-throughput screening of one-bead-one-compound libraries: Identification of cyclic peptidyl inhibitors against calcineurin/NFAT interaction. *ACS Comb. Sci.* **2011**, *13*, 537−546. (c) Thaker, H. D.; Cankaya, A.; Scott, R. W.; Tew, G. N. Role of Amphiphilicity in the Design of Synthetic Mimics of Antimicrobial Peptides with Gram-negative Activity. *ACS Med. Chem. Lett.* **2013**, *4*, 481−485. (d) Fenner, A. M.; Oppegard, L. M.; Hiasa, H.; Kerns, R. J. Selective inhibition of bacterial and human topoisomerases by arylacyl-sulfonated aminoglycoside derivatives. *ACS Med. Chem. Lett.* **2013**, *4*, 470−474. (e) Mor, S.; Pahal, P.; Narasimhan, B. Synthesis, characterization, biological evaluation and QSAR studies of 11-p-substituted phenyl-12-phenyl-11a,12-dihydro-11H-indeno[2,1-c][1,5]-benzothiazepines as potential antimicrobial agents. *Eur. J. Med. Chem.* **2012**, *57*, 196−210. (f) Deng, L.; Diao, J.; Chen, P.; Pujari, V.; Yao, Y.; Cheng, G.; Crick, D. C.; Prasad, B. V.; Song, Y. Inhibition of 1-deoxy-D-xylulose-5-phosphate reductoisomerase by lipophilic phosphonates: SAR, QSAR, and crystallographic studies. *J. Med. Chem.* **2011**, *54*, 4721−4734. (g) Shakibaei, G. I.; Feiz, A.; Khavasi, H. R.; Soorki, A. A.; Bazgir, A. Simple three-component method for the synthesis of spiroindeno[1,2-b]pyrido[2,3-d]pyrimidine-5,3′-indolines. *ACS Comb. Sci.* **2011**, *13*, 96−99.
(10) Gonzalez-Diaz, H.; Munteanu, C. R. *Topological Indices for Medicinal Chemistry, Biology, Parasitology, Neurological and Social Networks*; Transworld Research Network: Kerala, India, 2010.
(11) (a) Borchardt, R. T.; Kerns, E. H.; Hageman, M. J.; Thakker, D. R.; Stevens, J. L. *Optimizing the "Drug-Like" Properties of Leads in Drug Discovery*; Springer Science+Business Media, LLC: New York, 2006; Vol. *IV*. (b) Croes, S.; Koop, A. H.; van Gils, S. A.; Neef, C. Efficacy, nephrotoxicity and ototoxicity of aminoglycosides, mathematically modelled for modelling-supported therapeutic drug monitoring. *Eur. J. Pharm. Sci.* **2012**, *45*, 90−100. (c) Hau, J.; Schapiro, S. J. *Handbook of Laboratory Animal Science: Essential Principles and Practices*, 3rd ed.; CRC Press, Taylor & Francis Group, LLC: Boca Raton, FL, 2011; Vol. *1*.
(12) (a) Vina, D.; Uriarte, E.; Orallo, F.; Gonzalez-Diaz, H. Alignment-free prediction of a drug-target complex network based on parameters of drug connectivity and protein sequence of receptors. *Mol. Pharmaceutics* **2009**, *6*, 825−835. (b) Garcia, I.; Fall, Y.; Gomez, G.; Gonzalez-Diaz, H. First computational chemistry multi-target model for anti-Alzheimer, anti-parasitic, anti-fungi, and anti-bacterial activity of GSK-3 inhibitors in vitro, in vivo, and in different cellular lines. *Mol. Diversity* **2011**, *15*, 561−567. (c) Prado-Prado, F.; Garcia-Mera, X.; Abeijon, P.; Alonso, N.; Caamano, O.; Yanez, M.; Garate, T.; Mezo, M.; Gonzalez-Warleta, M.; Muino, L.; Ubeira, F. M.; Gonzalez-Diaz, H. Using entropy of drug and protein graphs to predict FDA drug-target network: Theoretic-experimental study of MAO inhibitors and hemoglobin peptides from *Fasciola hepatica*. *Eur. J. Med. Chem.* **2011**, *46*, 1074−1094. (d) Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S. A ligand-based approach for the in silico discovery of multi-target inhibitors for proteins associated with HIV infection. *Mol. BioSyst.* **2012**, *8*, 2188−2196. (e) Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S. Chemoinformatics in anti-cancer chemotherapy: Multi-target QSAR model for the in silico discovery of anti-breast cancer agents. *Eur. J. Pharm. Sci.* **2012**, *47*, 273−279. (f) Speck-Planche, A.; Luan, F.; Cordeiro, M. N. D. S. Abelson tyrosine-protein kinase 1 as principal target for drug discovery against leukemias. Role of the current computer-aided drug design methodologies. *Curr. Top. Med. Chem.* **2012**, *12*, 2745−2762. (g) Gonzalez-Diaz, H.; Prado-Prado, F. J. Unified QSAR and network-based computational chemistry approach to antimicrobials, part 1: Multispecies activity models for antifungals. *J. Comput. Chem.* **2008**, *29*, 656−667. (h) Marzaro, G.; Chilin, A.; Guiotto, A.; Uriarte, E.; Brun, P.; Castagliuolo, I.; Tonus, F.; Gonzalez-Diaz, H. Using the TOPS-MODE approach to fit multi-target QSAR models for tyrosine kinases inhibitors. *Eur. J. Med. Chem.* **2011**, *46*, 2185−2192.
(13) (a) Tenorio-Borroto, E.; Penuelas Rivas, C. G.; Vasquez Chagoyan, J. C.; Castanedo, N.; Prado-Prado, F. J.; Garcia-Mera, X.; Gonzalez-Diaz, H. ANN multiplexing model of drugs effect on macrophages; theoretical and flow cytometry study on the cytotoxicity of the anti-microbial drug G1 in spleen. *Bioorg. Med. Chem.* **2012**, *20*, 6181−6194. (b) Luan, F.; Cordeiro, M. N. D. S.; Alonso, N.; Garcia-Mera, X.; Caamano, O.; Romero-Duran, F. J.; Yanez, M.; Gonzalez-Diaz, H. TOPS-MODE model of multiplexing neuroprotective effects of drugs and experimental-theoretic study of new 1,3-rasagiline derivatives potentially useful in neurodegenerative diseases. *Bioorg. Med. Chem.* **2013**, *21*, 1870−1879. (c) Tenorio-Borroto, E.; Garcia-Mera, X.; Penuelas-Rivas, C. G.; Vasquez-Chagoyan, J. C.; Prado-Prado, F. J.; Castanedo, N.; Gonzalez-Diaz, H. Entropy model for multiplex drug-target interaction endpoints of drug immunotoxicity. *Curr. Top. Med. Chem.* **2013**, *13*, 1636−1649. (d) Speck-Planche, A.; Cordeiro, M. N. D. S. Simultaneous modeling of antimycobacterial activities and ADMET profiles: A chemoinformatic approach to medicinal chemistry. *Curr. Top. Med. Chem.* **2013**, *13*, 1656−1665.
(14) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: A large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **2012**, *40*, D1100−D1107.
(15) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *J. Cheminf.* **2011**, *3*, 33.
(16) Marrero-Ponce, Y.; Valdés-Martini, J. R.; García-Jacas, C. R.; Silveira-Vaz d'Almeida, Y.; Morell, C.; Cubillan, N. *TOMOCOMD-CARDD. Module QUBILs-MAS*, version 1.0; CAMD-BIR Unit: Santa Clara, Villa Clara, 2012.
(17) (a) Marrero-Ponce, Y. Total and Local Quadratic Indices of the Molecular Pseudograph's Atom Adjacency Matrix: Applications to the Prediction of Physical Properties of Organic Compounds. *Molecules* **2003**, *8*, 687−726. (b) Marrero-Ponce, Y. Linear indices of the "molecular pseudograph's atom adjacency matrix": Definition, significance-interpretation, and application to QSAR analysis of flavone derivatives as HIV-1 integrase inhibitors. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2010−2026. (c) Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castanedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Jorge, E.; del Valle, A.; Torrens, F.; Castro, E. A. TOMOCOMD-CARDD, a novel approach for computer-aided 'rational' drug design: I. Theoretical and experimental assessment of a promising method for computational screening and in silico design of new anthelmintic compounds. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 615−634. (d) Marrero-Ponce, Y.; Castillo-Garit, J. A. 3D-chiral atom, atom-type, and total non-stochastic and stochastic molecular linear indices and their applications to central chirality codification. *J. Comput.-Aided Mol. Des.* **2005**, *19*, 369−383. (e) Castillo-Garit, J. A.; Marrero-Ponce, Y.; Torrens, F. Atom-based 3D-chiral quadratic indices. Part 2: Prediction of the corticosteroid-binding globulin binding affinity of the 31 benchmark steroids data set.

Bioorg. Med. Chem. **2006**, *14*, 2398−2408. (f) Casanola-Martin, G. M.; Marrero-Ponce, Y.; Khan, M. T.; Ather, A.; Sultan, S.; Torrens, F.; Rotondo, R. TOMOCOMD-CARDD descriptors-based virtual screening of tyrosinase inhibitors: Evaluation of different classification model combinations using bond-based linear indices. *Bioorg. Med. Chem.* **2007**, *15*, 1483−1503.

(18) Marrero-Ponce, Y.; Iyarreta-Veitia, M.; Montero-Torres, A.; Romero-Zaldivar, C.; Brandt, C. A.; Avila, P. E.; Kirchgatter, K.; Machado, Y. Ligand-based virtual screening and in silico design of new antimalarial compounds using nonstochastic and stochastic total and atom-type quadratic maps. *J. Chem. Inf. Model.* **2005**, *45*, 1082−1100.

(19) Hill, T.; Lewicki, P. *STATISTICS Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining*; StatSoft: Tulsa, OK, 2006.

(20) *STATISTICA. Data analysis software system*, version 6.0; StatSoft: Tulsa, OK, 2001.

(21) Huberty, C. J.; Olejnik, S. *Applied MANOVA and discriminant analysis*, 2nd ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2006.

(22) Speck-Planche, A.; Kleandrova, V. V.; Luan, F.; Cordeiro, M. N. D. S. Rational drug design for anti-cancer chemotherapy: Multi-target QSAR models for the in silico discovery of anti-colorectal cancer agents. *Bioorg. Med. Chem.* **2012**, *20*, 4848−4855.

(23) Hanczar, B.; Hua, J.; Sima, C.; Weinstein, J.; Bittner, M.; Dougherty, E. R. Small-sample precision of ROC-related estimates. *Bioinformatics* **2010**, *26*, 822−830.

(24) Morrow, B. J.; He, W.; Amsler, K. M.; Foleno, B. D.; Macielag, M. J.; Lynch, A. S.; Bush, K. In vitro antibacterial activities of JNJ-Q2, a new broad-spectrum fluoroquinolone. *Antimicrob. Agents Chemother.* **2010**, *54*, 1955−1964.

(25) Davenport, J. M.; Covington, P.; Gotfried, M.; Medlock, M.; Watanalumlerd, P.; McIntyre, G.; Turner, L.; Almenoff, J. Summary of pharmacokinetics and tissue distribution of a broad-spectrum fluoroquinolone, JNJ-Q2. *Clinical Pharmacology in Drug Development* **2012**, *1*, 121−130.