J Mol Model (2005) 11: 116–123
DOI 10.1007/s00894-004-0228-3

ORIGINAL PAPER

**Humberto González-Díaz · Luis A. Torres-Gómez
Yaima Guevara · Manuel S. Almeida · Reinaldo Molina
Nilo Castañedo · Lourdes Santana · Eugenio Uriarte**

Markovian chemicals "in silico" design (MARCH-INSIDE), a promising approach for computer-aided molecular design III: 2.5D indices for the discovery of antibacterials

Received: 3 June 2004 / Accepted: 23 November 2004 / Published online: 19 February 2005

Abstract The present work continues our series on the use of MARCH-INSIDE molecular descriptors (parts I and II: J Mol Mod 8:237–245, [2002] and 9:395–407, [2003]). These descriptors encode information pertaining to the distribution of electrons in the molecule based on a simple stochastic approach to the idea of electronegativity equalization (Sanderson's principle). Here, 3D-MARCH-INSIDE molecular descriptors for 667 organic compounds are used as input for a linear discriminant analysis. This 2.5D-QSAR model discriminates between antibacterial compounds and non-antibacterial ones with 92.9% accuracy in training sets. On the other hand, the model classifies 94.0% of the compounds in test set correctly. Additionally, the present QSAR performs similar-to-better than other methods reported elsewhere. Finally, the discovery of a novel compound illustrates the use of the method. This compound, 2-bromo-3-(furan-2-yl)-3-oxo-propionamide has MIC_{50} of 6.25 and 12.50 μg/mL against *Pseudomonas aeruginosa* ATCC 27853 and *Escherichia coli* ATCC 27853, respectively while ampicillim, amoxicillim, clindamycin, and metronidazole have, for instance, MIC_{50} values higher than 250 μg/mL against *E. coli*. Consequently, the present method may becomes a useful tool for the *in silico* discovery of antibacterials.

Keywords Antibacterials · 2.5D-QSAR · Electronegativity equalization · Markov chains · Discriminant analysis

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00894-004-0228-3

H. González-Díaz · L. Santana · E. Uriarte
Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, 15782, Spain

H. González-Díaz (✉) · L. A. Torres-Gómez · Y. Guevara
R. Molina · N. Castañedo
Chemical Bioactives Center,
Central University of Las Villas,
54830, Cuba
E-mail: humbertogd@uclv.edu.cu
Tel.: +53-42-281473
Fax: +53-42-281130

M. S. Almeida
Department of Chemistry, University of Granma, Cuba

M. S. Almeida · R. Molina
Universität Rostock, FB Chemie,
Albert-Einstein-Str 3a, 18059 Rostock, Germany

Introduction

Quantitative-structure-activity relationships (QSAR) methods have emerged because of the interest in finding efficient methods for the discovery of new drug-like compounds, understanding biological mechanism of action, and the search for compounds with the required profile of activity [1]. This method is based on the representation of molecular structure by certain numbers, the so-called molecular descriptors, which are thereafter connected with biological activity by regression techniques [2, 3].

Quantum chemical calculations can be used to obtain a priori descriptors for QSAR studies. Given that some of these quantum properties are not observable, the way to calculate them is not uniquely defined. Consequently, it is likely that there are many different schemes for their calculation, none of which is fundamentally more correct than another. Unfortunately, the computation is often also computationally too demanding for large sets of molecules [4]. In order to tackle this difficulty, Bultinck et al. [5] described an implementation of a com-

putational approach based on the electronegativity equalization principle (EEP) to allow very fast calculation of atomic charges and related molecular descriptors. According to the EEP described by Sanderson, when molecules are formed the electronegativities of the constituent atoms become equal, coinciding with a fixed distribution for probabilities of finding the electrons in the neighborhood of a specific atom in the molecule at the steady state [6–9]. Simpler and faster methods to calculate molecular descriptors based on the idea of EEP are of interest for very large databases of compounds to find drug-like leads.

Consequently, our research group has introduced a Markov chain (MC) approach based on the idea of EEP [10–19]. As a consequence, we were able to derive new molecular descriptors encoding the distribution of electrons in the molecule. The approach termed as Markovian-Chemicals-In silico-Design (MARCH-INSIDE) has been shown to be very useful in drug design, toxicology, proteomics, and bioinformatics [10–19]. MARCH-INSIDE also presents the interesting ability to codify 3D structural features such chirality and Z-E isomerism [12, 13, 19].

This last feature encourages us to study highly 3D-structure-dependent pharmacological activities like antibacterial action [20]. As a result of the widespread use of antibacterials antibacterials-resistant pathogens have become more widespread. This, in turn has fuelled an ever-increasing need for new drugs [21]. 2.5D-MARCH-INSIDE and linear discriminant analysis (LDA) have been used to develop a QSAR in order to classify compounds as anti-bacterial or not, within structurally heterogeneous series. The results are presented in this work. The predictability for test set and comparison with respect to previous models validate this QSAR. In addition, the selection by virtual screening, synthesis, characterization, and preliminary assay of a novel compound illustrates how to use the model in practice. Finally, back-projection analysis of some compounds exemplifies the use of the model for structure-activity-relationship mapping.

Materials and methods

The basis of the 3D-MARCH-INSIDE approach

The basis of the present approach has been explained in detail elsewhere [11]. Briefly, the model constitutes a stochastic approach to EEP. Consider a hypothetical situation in which a series of atoms interact to form a molecule at an arbitrary initial time (t_0) [10–13, 18, 19]. Assume that after this initial situation, electrons start to distribute around atom cores in discrete intervals of time t_k. As depicted in Fig. 1, this model describes the probabilities ($^{k}p_{ij}$) with which electrons move from any arbitrary atom a_i at time t_0 (in black) to other a_j atoms (in white) throughout discrete time periods t_k ($k = 1, 2, 3,...$) and throughout the chemical bonds.

The present procedure considers the external electron layers of any atom core in the molecule (valence shell) as states of the MC. The method uses the matrix $^{1}\Pi$, which has the elements $^{1}p_{ij}$. This matrix is called the 1-step electron-transition stochastic matrix. $^{1}\Pi$ is built as a square table of order n, where n represents the number of atoms in the molecule. The elements ($^{1}p_{ij}$) of the 1-step electron-transition stochastic matrix are the transition probabilities with which electrons move from atom i to j in the interval $t_1 = 1$ (considering $t_0 = 0$) [10–13, 18, 19]:

$$^{1}p_{ij} = \frac{\chi_j . e^{\omega_j}}{\sum_{k=1}^{\delta+1} \chi_k . e^{\omega_k}} \quad (1)$$

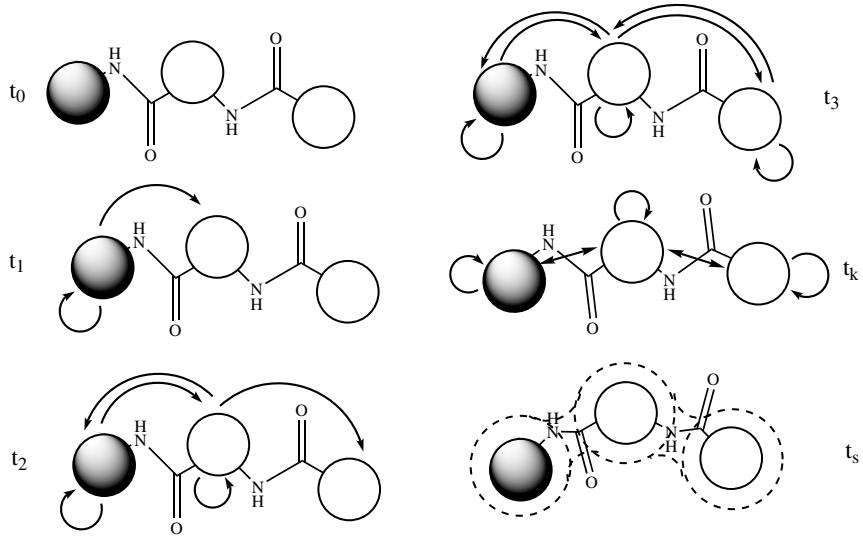


Fig. 1 Diagrammatic representation a Markov model for electrons distribution. The symbol $t_{stationary}$ represent the stationary time, electronegativity equalization

Where, χ_j is Pauling's electronegativity [22] for the atom a_j, which is bound to the atom a_i. In this equation, δ is the number of atoms that compete with a_i by its own electrons (atoms bound to a_i) the number 1 accounts for the atom a_i *per se*. We will only use $^1\Pi$ in the following. The spatial configuration of every atom has been codified through the dummy variable ω_j. This variable (ω_j) takes the value $\omega_j = 1$ if the atom a_j is R, E or *axial* and the values $\omega_j = 0$, -1 if the atom a_j does not have a specific spatial configuration or is present in the S, Z or *equatorial* configuration. The symbols R, S refer to the chirality of the atom. Alternatively, Z-E refers to the 3D characteristic for atoms involved in double bonds [12, 13, 19, 20]. The first step on calculating $^1\Pi$ is to derive an electronegativity matrix χ, whose elements are isomer-indicator exponential functions coinciding with the numerators of expressions like Eq. 1 [12, 13, 19]. In the short-term scale of time ($t_1 = 1$) the movement of electrons is described here by $^1\Pi$, whilst the probabilities of long-term movements are the elements of $^k\Pi$, ($t_k = k > 0$) described herein by the Chapman–Kolgomorov equations [10–23]:

$$^k\Pi = \left(^1\Pi\right)^k \tag{2}$$

The method uses the sum of the self-return probabilities of the natural power of this matrix ($^{SR}\pi_k$) as molecular descriptors. In classical Markov theory, these numbers are the probabilities with which the system returns to the initial state. In the present context, they are the probabilities with which electrons return to the original atoms at different times. That is to say, these numbers encode the distribution of electrons after the formation of the molecule as governed by EEP [10–13]. The calculation of $^{SR}\pi_k$ for any organic or inorganic molecule was carried out using the MARCH-INSIDE software, Tr being the trace operator [24]:

$$^{SR}\pi_k(\omega) = \sum_{i=1}^{g} {}^k p_{ii} = \mathrm{Tr}\left(^1\Pi^k\right) \tag{3}$$

Statistical analysis

In order to discriminate the antibacterial activity of drugs we will use a simple linear QSAR using 3D-MARCH-INSIDE with the general form:

$$A = b + b_0\, x\, {}^{SR}\pi_0(\omega) + b_1\, x\, {}^{SR}\pi_1(\omega) + b_2\, x\, {}^{SR}\pi_2(\omega) + \ldots + b_k\, x\, {}^{SR}\pi_k(\omega) \tag{4}$$

Here, b_k are the discriminant function coefficients fitted by linear discriminant analysis. The model deals with the discrimination of antibacterial chemicals from non-active ones.

Examination of the Fisher ratio (F) and the p-level (p) determines the quality of the model. We also inspect the percentage of good classification and the proportion between the cases and variables in the equation or variables to be explored in order to avoid over-fitting or chance correlation. Finally, predictability in an external prediction set validates the model. These compounds were never used to develop the classification function [25–27]. The ROC curves, and the area under these curves, were used additionally to validate the model [28].

Each compound was scored in terms of posterior probability by means of the posteriror probability with which the compounds is classified as antibacterial (P). This value constitutes a direct output of the model. This is a rigorous statistical index, which permits us to quote for the error. It possible to classify as antibacterial those compounds with $P\% > 5$, as a consequence that the model p-level threshold limit is 0.05. Conversely, those chemicals with $P\% < -5$ must be classified as inactive ones. Whereas, chemicals in the range $5 > P\% > 0$ must be considered as unclassified by the model at this p-level [10, 11].

Biological activity data used to seek the QSAR

Here we considered a general data set composed by structurally diverse organic chemicals. This original set was split at random in order to design two different sets of antibacterial chemicals and two additional sets of non-antibacterial ones. Both the antibacterial activity and the chemical structure of each compound were verified in different references [29, 30]. For training and predicting sets, those compounds recognized as antibacterials in the referred databases were considered as active without taking into consideration the strain or the concentration of drug required. Conversely, compounds with no effect against any strain were considered as non-active, as is usual practice in the QSARs reviewed below in Table 1.

Synthesis and characterization

Reagents were used as purchased without further purification. Solvents ($CHCl_3$) were dried and freshly distilled before use according to literature procedures. Chromatographic TLC was performed on pre-coated silica gel polyester plates (0.25 mm thickness) with fluorescent indicator UV 254 (Polychrom SI F254). Melting points were determined on a Buchi 510 apparatus and are uncorrected. The IR spectrum was recorded on a Perkin-Elmer 1640FT spectrometer (KBr disk, υ in cm^{-1}). The ^{1}H-NMR and ^{13}C-NMR spectra were recorded on a Bruker WP 200-SY spectrometer at 200 MHz or on a Bruker SY spectrometer (400 MHz), the chemical shifts (σ) are given in ppm downfield from tetramethylsilane. For EIMS analysis, a VG-TS250 apparatus (70 eV) was used. Elemental analysis was performed on a Perkin-Elmer 240B microanalyzer and the values were within $\pm$ 0.4% of calculated ones in all cases.

Table 1 Comparison with other approaches

Models' features to be compared[a]	Antibacterial activity classification models										
	1	2	3	4	5	6	7	8	9	10	11
N total	667	661	661	661	352	111	111	–	972	458	433
N antibacterials	363	249	249	249	174	60	60	–	241	229	217
Technique[b]	LDA	LDA	BLR	ANN	LDA	LDA	ANN	LDA	LDA	LDA	LDA
U-statistics (Wilk's λ)	0.38	–	–	–	0.45	0.28	–	0.57	0.58	0.56	–
F	139.3	–	–	–	48.2	20.9	–	116.6	98	98	–
D^2	5.33	–	–	–	4.9	–	–	–	–	–	–
p-level	0.00	–	–	–	0.00	0.00	–	0.00	0.00	0.00	0.00
3D-topologic indices[c]	yes	no	no	no	no	no	no	no	no	no	no
Explored variables	10	167	167	62	10	16	16	–	–	–	62
Vars. in model	7	6	6	62	7	7	7	8	8	2	6
Back-projection[d]	yes	no	no	no	yes	no	no	no	no	no	no
Training series											
N total	492	661	661	661	289	64	64	294	698	355	433
N antibacterials	274	249	249	249	174	34	34	–	169	161	217
Accuracy (%)	92.85	92.6	94.7	–	91	94.0	89.0	> 90	86.8	~ 85	~ 85
Families of drugs[e]	11	8	8	8	9	3	3	–	> 5	–	> 8
Validation											
Validation method[f]	i	ii	ii	iii	i	i	i	i	i	i	i
N total	175	–	–	63	63	47	47	70	274	103	128
N antibacterials	89	–	–	45	45	26	26	–	72	68	64
Predictability (%)	94	93.6	94.3	~ 94	89	92	97.9	> 90	86.9	~ 85	~ 85
Families of drugs[b]	11	–	–	8	9	3	3	–	> 5	–	> 8

[a]Model 1 is reported in this work, models 2 and 3 were reported by Cronin et al. in reference [35], model 4 appears in reference [36] after Tomás-Vert et al., model 5 was very recently reported by Molina et al. [37] models 6 and 7 were published in 1998 (reference [38]), model 8 was developed by Mut-Ronda et al., [39] two LDA models were recently introduced by Murcia-Soler et al.; model 9 in [39] and 10 in reference [41]. The last model here depicted was published by Gregorio-Alapont et al. [42]

[b]LDA refers to Linear discriminant analysis, ANN to artificial neural network, and BLR to binary logistic regression.

[c] Considers the capability of the method to encode together chirality, Z-E, and axial-equatorial isomerism.

[d]Considers the possibility of deriving a map with the calculated contribution of any atom in the molecule to the biological activity.

[e]Only largely represented families were considered, e. g, methods 1 and 2 used 3 in training quinolones, sulphonamides, and cephalosporins but add only diaminopyridine (1 compound), cephamicins (2), oxacephems (1) and sulfones (1) to predicting series.

[f]Validation methods are: (i) external predicting series, (ii) leave-30%-out crossvalidation, and (iii) 100-times-averaged re-substitution technique. Furthermore, note that methods ii and iii are cross-validation methods

2-Bromo-3-(furan-2-yl)-3-oxo-propionamide

To a solution of 2-furoylacetonitrile (2.7 g, 0.02 mol) in chloroform (20 mL), was added under strong stirring anhydrous benzoyl peroxide (4.84 g, 0.02 mol). Thereafter, bromine (3.2 g, 0.02 mol) was added and the mixture was stirred for 2 h at r. t. The solution was washed with water and $NaHCO_3$ 5%, dried over Na_2SO_4 and the solvent was evaporated under vacuum. The solid residue was recrystallized from ethanol to give the desired compound (3.6 g, 77% yield).

- Mp (dec.) = 155.8–156.7 °C.
- ^{1}H-NMR (acetone- d_6), δ: furanic protons [7.94 (m, lH, H-5), 7.55 (m, lH, H-3), 6.75 (m, lH, H-4)], 7.0–7.75 (m, 2H exch., NH_2), 5.73 (s, lH, CH).
- ^{13}C-NMR (acetone- d_6), δ: furanic carbons [150.82 (C-2), 149.25 (C-5), 120.92 (C-3), 113.7 (C-4)], 178.83 (C = O), 184.5 (C = O), 47.51 (CH).
- EIMS (70 eV) m/z (%): 231 (M^+), 233 ($[M+2]^+$), 152 (41), 95 (98), 31 (100).
- IR υ : 3,423, 3,301, 1,666, 1,656.
- Anal. $C_7H_6BrNO_3$: C, H, N.

Biological assay of a new compound

This study was carried out with a new compound not contained in the training or predicting series but predicted afterwards. In vitro minimal inhibitory concentration (MIC_{50}) assays were carrying out throughout the Mueller-Hinton serial dilution method according to the recommendations of the National Committee for Clinical Laboratory Standards [31, 32]. The MIC_{50} value was determined for two ATCC reference bacterial strains: *Ps. aeruginosa* ATCC 27853 and *E. coli* ATCC 27853.

Results and discussion

QSAR modeling

Once we split at random the original data in representative training and test set, the training set may be used to seek the discriminant function using LDA. The model selection was subjected to the principle of parsimony or

Occam's Razor [27]. As a result; we chose a function with higher statistical significance but with few parameters (b_k) as possible:

$$A_{stand} = 0.02^{SR}\pi_0(\omega) + 1.57^{SR}\pi_2(\omega) - 3.56^{SR}\pi_3(\omega) + 4.33^{SR}\pi_4(\omega) - 19.5^{SR}\pi_6(\omega) + 24.8^{SR}\pi_7(\omega) - 7.8^{SR}\pi_9(\omega) + 4.221$$
$$F = 139.27 \; p<0.001 \quad (5)$$

Where, A_{stand} is a dummy variable (1 for antibacterial compounds and −1 for non-active ones). Prior to fitting, all molecular descriptors were mean-centered and normalized to avoid certain descriptors dominating the model. The Fisher test allows us to test the hypothesis of separation of groups with a probability of error (p-level) $p < 0.05$. All the parameters have the same values for Eq. 5 [10, 11, 27, 33]

The model correctly classifies 92.9% of the compounds in the training set. Specifically, the model correctly classifies 252 out of 274 (91.9%) antibacterial compounds and 205 out of 218 (93.9%) inactive compounds in the training set. Although the model is not strictly 3D, the use of ω allowed us to take compounds with specific 3D structure into consideration. The names of all compounds used to derive the QSAR as well as their predicted activity appear in Table S1 (see supplementary material). Compounds were ordered according to different intervals of predicted activity.

On the other hand, the prediction series shows a 94.0% global predictability. In this study the discriminant function gave rise for a good classification of 85 out of 92 (95.9%) and 79 out of 83 (92.2%) non-active drugs respectively. The names of all compounds used to validate the QSAR as well as their predicted activities appear in Table S2, see supplementary information. Both, training and predicting sets percentages of good classification validates the model for the use in virtual screening taking into consideration that 85.0% is considered as an acceptance threshold limit for this kind of analysis [34].

A more serious validation was carried out by calculating the areas under the receiver operating characteristic (ROC) curve to show how well the model classifies. Figure 2 depicts the ROC curves obtained for compounds on training and predicting sets with areas under the curves of 0.98 and 0.97, respectively. A ROC area of 1.0 indicates perfect classification. A clear difference between the two ROC curves and the line in the main diagonal that represents a random classifier area under curve equal to 0.5 can be seen [28].

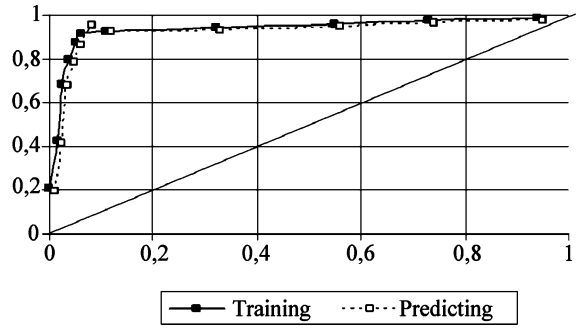


Fig. 2 Operating receive characteristic curve (ROC-curve) for training and predicting series of antibacterial and non-active compounds

Comparison with other models

The present QSAR performs better-to-similar with respect to the other 11 models based on large heterogeneous series of antibacterial/non-antibacterial compounds, see Table 1 of the supplementary material [35–41]. In addition to the ten models shown in the Table 1, another interesting study of antibacterial activity was reported by Mishra et al. [42] However, the largest data set studied by Mishra et al. [42] incorporated only 463 compounds (242 antibacterials) with about 84% overall predictability and without 3D or 2.5D indices. Briefly, the present model has some interesting characteristics (see Table 1, salient points in boldface):

1. It uses the largest up-to-date reported data set of experimentally corroborated antibacterial compounds (363) for a QSAR study.
2. A broad range of applicability for this model can be stressed considering the larger number of families of different organic compounds used.
3. It must also be noted that on seeking the present model we explored only ten molecular descriptors.
4. This model makes use of chiral topologic indices [43] for the search of antibacterial compounds for the first time.
5. The quality of the predictions of this model have been assessed using a more rigorous test set method, resubstitution. Some of the other models assessed predictability using cross-validation methods (e.g. leave-one-out or jacknife).

The previous topological models do not consider the 3D structure. Although the present model does not consider geometric 3D information, it may discriminate drugs with different 3D structure due to the use of ω, including optical and Z-E isomers. Consequently, predictions using the earlier models are expected to fail in such very commonly occurring cases in which stereochemistry determines the biological activity. [12, 18–20, 44] The discrimination of different kinds of stereoisomers has been fully exemplified before for 3D-MARCH-INSIDE [12, 13, 19]. Figure 3 illustrates this aspect for a simple example. It is straightforward to realize that, the more different the matrices are for a pair of isomers the larger the differences for their calculated molecular descriptors. We would like to highlight that the present

Fig. 3 Depicts the different electronegativity matrix χ (which elements are equal to $\chi_j \cdot \exp(\omega_j)$) if atom a_i bounds to atom a_j or equal to 0 otherwise, the normalized stochastic or Markov matrix $^1\Pi$ derived from χ and its first power $^2\Pi$ for a pair of mirror isomers

[R]-CHClBrF

χ_j	H	F	C	Br	Cl
H	2.1	0	6.8	0	0
F	0	4	6.8	0	0
C	2.1	4	6.8	2.8	3
Br	0	0	6.8	2.8	0
Cl	0	0	6.8	0	3

$^1\Pi$[R]

	H	F	C	Br	Cl
H	0.24	0	0.76	0	0
F	0	0.4	0.63	0	0
C	0.11	0.2	0.36	0.15	0.16
Br	0	0	0.71	0.29	0
Cl	0	0	0.69	0	0.31

$^2\Pi$[R]

	H	F	C	Br	Cl
H	0.14	0.2	0.46	0.11	0.12
F	0.09	0.4	0.25	0.12	0.13
C	0.1	0.2	0.35	0.15	0.16
Br	0.1	0.2	0.26	0.29	0.15
Cl	0.1	0.2	0.26	0.13	0.31

[S]-CHClBrF

χ_j	H	F	C	Br	Cl
H	2.1	0	0.92	0	0
F	0	4	0.92	0	0
C	2.1	4	0.92	2.8	3
Br	0	0	0.92	2.8	0
Cl	0	0	0.92	0	3

$^1\Pi$[R]

	H	F	C	Br	Cl
H	0.7	0	0.3	0	0
F	0	0.8	0.19	0	0
C	0.16	0.3	0.07	0.22	0.23
Br	0	0	0.25	0.75	0
Cl	0	0	0.23	0	0.77

$^2\Pi$[R]

	H	F	C	Br	Cl
H	0.53	0.1	0.23	0.07	0.07
F	0.03	0.6	0.35	0.04	0.04
C	0.09	0.2	0.43	0.13	0.14
Br	0.04	0.1	0.4	0.45	0.05
Cl	0.03	0.1	0.39	0.05	0.47

comparison does not involve the abilities of different indices as molecular descriptors. In fact, recent results report the generalization of classic connectivity indices for chirality codification [43–46]. The present comparison refers only to the range of applicability of the reported QSARs with respect to the present one.

Virtual screening

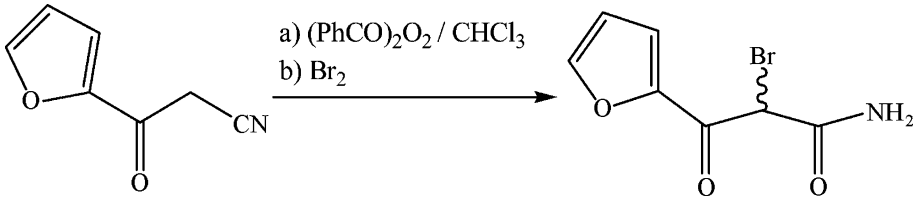
Finally, the discovery of 2-bromo-3-(furan-2-yl)-3-oxo-propionamide as a novel antibacterial compound illustrates the use of the model in practice. First, the $^{SR}\pi_k(\omega)$ values were calculated for a large dataset of organic compounds. Unfortunately, the data it is not available at the moment due to ongoing patenting process. However, this data it is not necessary to reproduce the present work, taking into consideration that only training and validation sets must be used in doing so. Secondly, the posterior probabilities of antibacterial activity were predicted with the 2.5D-QSAR. Lastly, the compound with the highest probabilities was synthesized starting from 2-furoyl-acetonitrile, with a yield of 77.0% (Fig. 4).

After biological assay this amide showed MIC_{50} values of 6.25 and 12.50 μg/mL against *Ps. aeruginosa* ATCC 27853 and *E. coli* ATCC 27853, respectively. These are strains of bacterial species with high clinical incidence. With these MIC_{50}, the compound may be considered useful as a lead compared for instance with the MIC_{50} for ceftriaxone, a comercial drug, which has an $MIC_{50}$of 0.06 against μg/mL *E. coli* but 64.0 μg/mL against *Ps. aeruginosa* [47]. However, *E. coli* has developed resistance against several broad-spectrum antibacterial drugs such as ampicillim, amoxicillim, clindamycin, and metronidazole with MIC_{50} values higher than 250.0 μg/mL in all cases [48]. In the present case, both enantiomers were predicted with similarly high probabilities, so we decided not to separate them in this preliminary study. More rigorous studies aimed at the synthesis, characterization, stability, biological testing, and the mechanism of action (now unknown) of both enantiomers and their derivatives are beyond of the scope of the present paper.

Back-projection analysis

Finally, to gain further insight into the role played by the different molecular features a back-projection approach was used. Specifically, the use of back-projectable approaches enables the variables in the QSAR to be projected back into the molecular space, providing biologically and chemically significant conclusions. The MARCH-INSIDE descriptors introduced by our research group constitute another example of novel back-projectable molecular descriptors. Specifically, the model introduced in the present work may be used to draw visual structure-activity maps for drugs in the training and test sets, as well as for the novel compound reported here by the first time. This analysis makes it possible to calculate the contribution of the different groups of atoms in the molecule to the pharmacological activity. First, the $SR^{\pi}{}_{k}$ values for each atom in the

Fig. 4 Synthesis of 2-bromo-3-(fur-2-yl)-3-oxo-propionamide



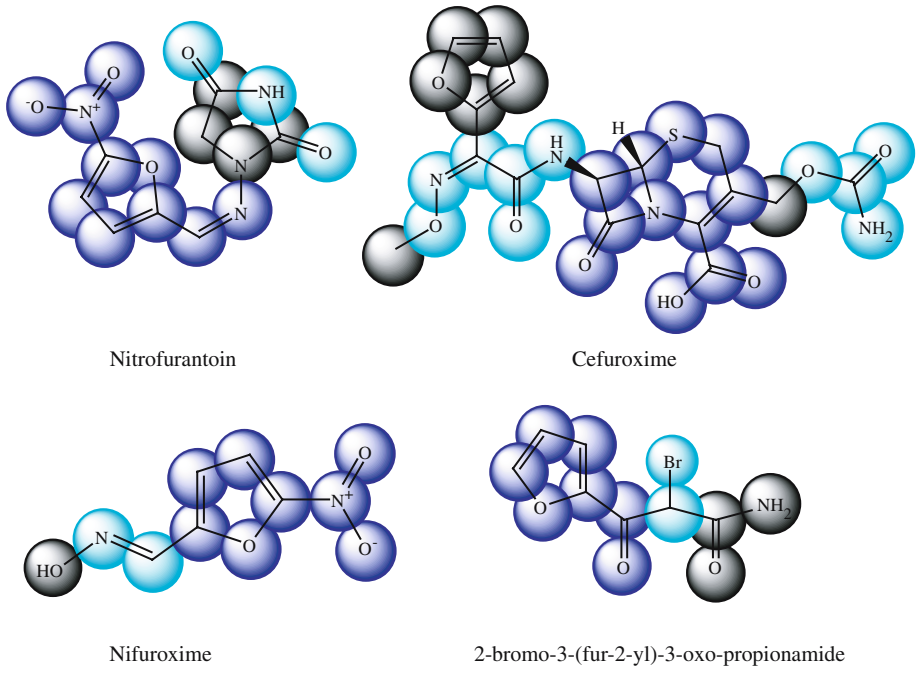


Fig. 5 Colour scaled backprojection analysis of some compounds in training and predicting sets: nitrofurantoin (*left top*), cefuroxime (*right top*), nifuroxime (*left bottom*) and 2-bromo-3-(fur-2-yl)-3-oxo-propionamide (*right bottom*). Colour code is as follows: *blue* structural framework with high (more than 50%) contribution to the property, *light blue* group with significant contribution (20–50%), *grey* group with low (< 10%) or not contribution to the property

molecule are calculated and afterwards they are evaluated in the QSAR equation. All values are normalized within 0–100 scales [49].

Figure 5 depicts these maps for nitrofurantoin, cefuroxime, nifuroxime, and 2-bromo-3-(fur-2-yl)-3-oxo-propionamide. Interestingly, one can note that the furan ring shows high positive contributions to the activity in nitrofurantoin, and nifuroxime, which are 5-nitro-furans with a double bond (C = N) attached to position 2 of the furan ring. These classes of compounds are expected to bind the target by nucleophilic substitution at the double bond activated by the electron-withdrawing nitro group. In this sense, our maps coincide with previous knowledge. Conversely, in cefuroxime the highest contribution was calculated for the β-lactamic framework with no significant contribution of the furan ring. This coincides with the structure-activity relationships for cephalosporin. In the case of 2-bromo-3-(fur-2-yl)-3-oxo-propionamide, a high contribution for the furan ring is also predicted. This indicates that this compound is more like the nitro-furans [21].

The explosion in the use of novel topologic molecular descriptors will continue in the future. The fusion of high throughput screening with QSAR techniques is a promising new field [50]. In conclusion, the above modeling results introduce a timely way for the discovery of antibacterial lead-like compounds taking into consideration 3D structural features.

Acknowledgements We thank the Spanish Ministry of Science and Technology (SAF2003-02222), for partial financial support. Molina RR, Castañedo C, and Almeida SM, acknowledge support from the Universität Rostock, Germany. Both unknown referees are acknowledged by their useful comments.

References

1. Randič M, Sabljič A, Nikolič S, Trinajstič N (1988) Int J Quantum Chem Quantum Biol Symp 15:267–285
2. Kier LB, Hall LH (1999) Topological indices and related descriptors in QSAR and QSPR. Gordon and Breach, Amsterdam, pp 455–489
3. Todeschini R, Consonni V (2000) Handbook of molecular descriptors. Mannhold R, Kubinyi H, Timmermann H (eds) Wiley-VCH, Weinheim
4. Karelson M, Lobanov VS, Katritzky AR (1996) Chem Rev 96:1027–1043
5. Bultinck P, Langenaeker W, Carbó-Dorca R, Tollenaere JP (2003) J Chem Inf Comput Sci 43:422–428
6. Sanderson RT (1951) Science 114:670–672
7. Sanderson RT (1983) Polar covalence. Academic, New York
8. Gasteiger J, Marsili M (1980) Tetrahedron 36:3219–3228
9. Mortier WJ, Gosh SK, Shankar S (1986) J Am Chem Soc 108:4315–4320
10. González-Díaz H, Olazábal E, Castañedo N, Hernádez SI, Morales A, Serrano HS, González J, Ramos de Armas R (2002) J Mol Mod 8:237–245
11. González-Díaz H, Gia O, Uriarte E, Hernádez SI, Ramos de Armas R, Chaviano M, Seijo S, Castillo JA, Morales L, Santana L, Akpaloo D, Molina E, Cruz M, Torres LA, Cabrera MA (2003) J Mol Mod 9:395–407
12. González-Díaz H, Bastida I, Castañedo N, Nasco O, Olazábal E, Morales A, Serrano SS, Ramos de Armas R (2004) Bull Math Biol 66:1285–1311. DOI 10.1016/j.bulm.2003.12.003
13. González-Díaz H, Marrero Y, Hernández SI, Bastida I, Tenorio I, Nasco O, Uriarte E, Castañedo N, Cabrera M, Aguila E, Marrero O, Morales A, Pérez M (2003) Chem Res Toxicol 16:1318–1327
14. González-Díaz H, Ramos de Armas R, Uriarte E (2002) Online J Bioinf 1:83–95
15. González-Díaz H, Ramos de Armas R, Molina RR (2003) Bull Math Biol 65:991–1002
16. González-Díaz H, Ramos de Armas R, Molina RR (2003) Bioinformatics 19:2079–2087

17. González-Díaz H, Molina RR, Uriarte E (2004) Polymer 45:3845–3853
18. Ramos de Armas R, González-Díaz H, Molina RR, Uriarte E (2004) Proteins: Struct Funct Bioinf 56:715–723. DOI 10.1002/prot.20159
19. González-Díaz H, Hernández SI, Uriarte E, Santana L (2003) Comput Biol Chem 27:217–227
20. Eliel E, Wilen S, Mander L (1994) Stereo chemistry of organic compounds. Wiley, New York, pp 103–112
21. Hardman GJ, Lee EI (1996) Goodman and Gilman's, the pharmacological basis of therapeutics, 9th edn, McGraw-Hill, New York
22. Pauling L (1939) The nature of chemical bond. Cornell University Press, Ithaca, pp 2–60
23. Freund JA, Poschel T (2000) Stochastic processes in physics, chemistry, and biology. In: Lect Notes Phys. Springer, Berlin Heidelberg New York
24. González-Díaz H, Molina RR, Hernández SI (2003) MARCH-INSIDE 2.0 (Markovian Chemicals "In Silico" Design) Chemical Bioactives Center, Central University of "Las Villas", Cuba. The present constitutes a preliminary experimental version; future professional versions shall be available to the public free of charge. For any information about it, sends and e-mail to the corresponding author: humbertogd@vodafone.es or humbertogd@cbq.uclv.edu.cu.
25. Kowalski RB, Wold S (1982) Pattern recognition in chemistry. In: Krishnaiah PR, Kanal LN (eds) Handbook of statistics. North Holland, Amsterdam, pp 673–697
26. STATISTICA (2002) Statsoft Inc, version. 6.0
27. van Waterbeemd H (1995) Discriminant analysis for activity prediction. In: Method and principles in medicinal chemistry. Chemometric methods in molecular design. VCH, Weinheim, pp 265–282
28. Swets JA (1988) Science 240:1285–1293
29. Kleeman A, Engel J, Kutscher B, Reichert D (2001) Pharmaceutical Substances 4th. George Thieme Verlag, Stuttgart
30. Negwer M (1987) Organic chemical drugs and their synonyms. Akademie, Berlin
31. The ATCC Bacteriology Collection (2002): http://www.atcc.org/SearchCatalogs/Bacteria.cfm
32. National Committee for Clinical Laboratory Standards (2000) Methods for dilution antimicrobial susceptibility tests for bacteria that gorw aerobically. Approved standard M7-A5. Wyne, Pa
33. González MP, González-Díaz H, Molina RR, Cabrera MA, Ramos de Armas R (2003) J Chem Inf Comput Sci 43:1192–1199
34. Gálvez J, García R, Salabert MT, Soler R (1994) J Chem Inf Comput Sci 34:520–525
35. Cronin MTD, Aynur AO, Dearden CJ, Deffy CJ, Netzeva IT, Patel H, Rowe HP, Schultz TW, Worth PA, Voutzolidis K, Schüürmann G (2002) J Chem Inf Comput Sci 42:869–878
36. Tomás-Vert F, Pérez-Giménez F, Salabert-Salvador MT, García-March FJ, Jaén-Oltra J (2000) J Mol Struct Theochem 504:249–259
37. Molina E, González-Díaz H, González MP, Rodríguez E, Uriarte E (2004) J Chem Inf Comput Sci 44:515–521
38. García-Domenech R, Julián-Ortiz JV (1998) J Chem Inf Comput Sci 38:445–449
39. Mut-Ronda S, Salabert-Salvador MT, Duart MJ, Antón-Fos GM (2003) Bioorg Med Chem Lett 13:2699–2702
40. Murcia-Soler M, Pérez-Giménez F, García-March FJ, Salabert-Salvador MT, Díaz-Villanueva W, Medina-Casamayor PJ (2003) Mol Graph Mod 21:375–390
41. Murcia-Soler M, Pérez-Giménez F, García-March FJ, Salabert-Salvador MT, Díaz-Villanueva W, Castro-Bleda MJ, Villanueva-Pareja AJ (2004) Chem Inf Comput Sci 44:1031–1041
42. Mishra RK, Garcia-Domenech R, Galvez J (2001) J Chem Inf Comput Sci 41:383–393
43. Golbraikh A, Bonchev D, Tropsha A (2001) J Chem Inf Comput Sci 41:147–158
44. Julián-Ortiz JV, de Gregorio Alapont C, Ríos-Santamaria I, García-Domenech R, Gálvez J (1988) J Mol Graphics Modell 16:14–18
45. Kellogg GE, Kier LB, Gaillard P, Hall LH (1996) J Comput-Aided Mol Des 10:513–520
46. Golbraik A, Bonchev D, Tropsha A (2001) J Chem Inf Comput Sci 41:147–158
47. Fujimura T, Yamano Y, Yoshida I, Shimada J, Kuwahara S (2003) Antimicrob Agents Chemoter 47:923–931
48. Handal T, Caugant DA, Olsen I (2003) Antimicrob Agents Chemoter 47:1443–1446
49. Stilf N, Baumann K (2003) J Med Chem 46:1390–1407
50. Devillers J, Balaban AT (2000) Topological indices and related descriptors in QSAR and drug design. Amsterdam, pp 3–41