Available online at www.sciencedirect.com
ScienceDirect
Bioorganic & Medicinal Chemistry 15 (2007) 897–902
Bioorganic & Medicinal Chemistry

Unified QSAR approach to antimicrobials. Part 2: Predicting activity against more than 90 different species in order to halt antibacterial resistance

Francisco J. Prado-Prado, Humberto González-Díaz,* Lourdes Santana and Eugenio Uriarte

Department of Organic Chemistry and Institute of Industrial Pharmacy, Faculty of Pharmacy, University of Santiago de Compostela, 15782 Santiago, Spain

Received 27 July 2006; revised 11 October 2006; accepted 19 October 2006
Available online 21 October 2006

Abstract—There are many different kinds of pathogenic bacteria species with very different susceptibility profiles to different antibacterial drugs. One limitation of QSAR models is that they consider the biological activity of drugs against only one species of bacteria. In a previous paper, we developed a unified Markov model to describe the biological activity of different drugs tested in the literature against some antimicrobial species. Consequently, predicting the probability with which a drug is active against different species of bacteria with a single unified model is a goal of major importance. The work described here develops the unified Markov model to describe the biological activity of more than 70 drugs from the literature tested against 96 species of bacteria. We applied linear discriminant analysis (LDA) to classify drugs as active or inactive against the different tested bacterial species. The model correctly classified 199 out of 237 active compounds (83.9%) and 168 out of 200 inactive compounds (84%). Overall training predictability was 84% (367 out of 437 cases). Validation of the model was carried out using an external predicting series, with the model classifying 202 out of 243 (i.e., 83.13%) of the compounds. In order to show how the model functions in practice, a virtual screening was carried out and the model recognized as active 84.5% (480 out of 568) antibacterial compounds not used in the training or predicting series. The current study is an attempt to calculate within a unified framework the probabilities of antibacterial action of drugs against many different species.

1. Introduction

With the increase in resistance of bacteria to antibiotic treatment, attention has focussed on developing novel approaches to antimicrobial therapy. One possibility is to exploit natural mechanisms used by mammals, including humans, to combat microbial invaders. Modern rational drug design widely relies on building extensive QSAR (quantitative structure–activity relationship) models and this represents a substantial part of the current *'in silico'* research. QSAR can then be utilized to optimize both the activity profile of the molecule and its chemical synthesis.[1] Unfortunately, QSAR studies are generally based on databases that consider only parent compounds acting against one single microbial species. As a consequence, to predict the antimicrobial activity for a given series of compounds one has to use/seek as many QSAR models as the number of microbial species for which one wishes to predict the drug's susceptibility.[2] In a previous paper, we developed a single unified Markov model to describe the biological activity of different drugs tested in the literature against different antimicrobial species. In this sense, the development of one single unified equation to calculate the probability of activity of a given drug against different antimicrobial species is very important.

Bacterial infections have increased dramatically in recent years. Bacteria have been the cause of some of the most deadly diseases and widespread epidemics in human civilization. Bacterial diseases such as tuberculosis, typhus, plague, diphtheria, typhoid fever, cholera, dysentery, and pneumonia have taken a high toll on humanity. Water purification, immunization (vaccination), and modern antibiotic treatments continue to

Keywords: QSAR; Antibacterial drugs; *Staphylococcus*; *Streptococcus*; Markov model; Molecular descriptors; Stochastic matrix.
* Corresponding author. Tel.: +34 981 563100; fax: +34 981 594912; e-mail: gonzalezdiazh@yahoo.es

0968-0896/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2006.10.039

reduce the morbidity and the mortality of bacterial disease in the 21st century—at least in the developed world, where these are acceptable cultural practices. However, many new bacterial pathogens have been recognized in the past 25 years and many others, such as *Staphylococcus aureus* and *Streptococcus pneumoniae*, have emerged with new forms of virulence and new patterns of resistance to antimicrobial agents.[3]

There are more than 1600 molecular descriptors that can, in principle, be generalized and used to solve the problem outlined above.[4–7] In addition, other QSAR approaches have been introduced recently and their utility has been demonstrated in medicinal chemistry.[8–11] In any case, none of these indices have been extended to encode additional information concerning chemical structure. Our group has previously introduced one Markov Model (MM) to encode molecular backbone information, with several applications in bioorganic medicinal chemistry. The method was named the MARCH-INSIDE approach, MARkovian CHemicals IN SIlico Design. This approach allowed us to introduce matrix invariants such as stochastic entropies and spectral moments for the study of molecular properties. Specifically, the stochastic spectral moments introduced by our group have been largely used for small molecule QSAR problems, including the design of fluckicidal, anticancer, and antihypertensive drugs. Applications to macromolecules have been restricted to the field of RNA without application to proteins.[12–15] The entropy-like molecular descriptors have demonstrated flexibility in many bioorganic and medicinal chemistry problems and these include the estimation of anticoccidial activity, modeling the interaction between drugs and HIV-packaging-region RNA, and predicting protein and virus activity.[16–22]

In recent studies, our group extended the MARCH-INSIDE method to encompass interesting molecular environment information in addition to molecular structure. This new interpretation allows the calculation of the molecular thermodynamic free energy for many physicochemical and biological processes.[23,24] This approach is able to take into consideration not only the molecular structure of the drug but also the free energy of its interaction with the specific microbial organism that the drug has to eliminate. In the present study, we developed a single linear equation to predict the antibacterial activity of drugs against different species based on these previous ideas.

2. Results and discussion

The advantage of the present stochastic approach is the possibility of deriving average contributions to the biological activity depending on the probability of the states of the MM. The generalized parameters fit in a more clearly physicochemical sense in comparison to our previous ones.[23–25] More specifically, this work is the first study in which a single linear QSAR equation model is introduced to predict the antibacterial activity of drugs against different species. A summary of the forward-stepwise analysis shows the variables that enter first in the model (Table 1). The best model found was:

$$\begin{aligned} \text{Actv} = & -1.12 \cdot {}^{1}C_s(T) + 1.34 \cdot {}^{3}C_s(T) + 1.84 \\ & \cdot {}^{0}C_s(\text{C}_{\text{sat}}) - 0.90 \cdot {}^{0}C_s(\text{C}_{\text{uns}}) + 0.88 \\ & \cdot {}^{5}C_s(\text{X}) - 1.27 \cdot {}^{0}C_s(\text{H-Het}) - 0.90 \\ & \cdot {}^{2}C_s(\text{H-Het}) + 0.698 \\ & \lambda = 0.49 \quad Rc = 0.715 \quad p < 0.001 \end{aligned} \tag{1}$$

where λ is the Wilk's statistic for the overall discrimination, F is the Fisher ratio, and p the error level. In this equation, ${}^{k}C_s$ values were calculated for the totality (T) of the atoms in the molecule or for specific collections of atoms. These collections are atoms with a common characteristic, such as halogens (X), unsaturated Carbon atoms (C) or heteroatom-bound hydrogen atoms (H-Het). The model correctly classified 798 out of 848 active compounds (94%) and 312 out of 400 inactive compounds (78%). Overall training predictability was 84.05% (1049 out of 1248 compounds). We validated the model using an external predicting series, with the model classifying 202 out 243 (83.13%) of the compounds (see Table 2). In addition, we used an ROC

Table 1. Summary for the forward-stepwise analysis

	F	p	Effect
${}^{0}C_s$ (Het)	68.2	0.001	In
${}^{0}C_s$ (C_{sat})	151.3	0.001	In
${}^{3}C_s$ (X)	50	0.001	In
${}^{1}C_s$ (T)	59.9	0.001	In
${}^{0}C_s$ (C_{inst})	50.4	0.001	In
${}^{5}C_s$ (C_{sat})	47.6	0.001	In
${}^{2}C_s$ (X)	24.3	0.001	Entered
${}^{3}C_s$ (T)	12.6	0	Out
${}^{1}C_s$ (C_{sat})	0.7	0.398	Out
${}^{2}C_s$ (C_{sat})	0.9	0.334	Out
${}^{3}C_s$ (C_{sat})	0.2	0.623	Out
${}^{4}C_s$ (C_{sat})	0.2	0.641	Out
${}^{5}C_s$ (T)	2.4	0.123	Out
${}^{4}C_s$ (T)	13.1	0	Out
${}^{1}C_s$ (C_{inst})	0.8	0.38	Out
${}^{2}C_s$ (C_{inst})	2.3	0.127	Out
${}^{3}C_s$ (C_{inst})	2.1	0.145	Out
${}^{4}C_s$ (C_{inst})	0.1	0.724	Out
${}^{5}C_s$ (C_{inst})	1.9	0.171	Out
${}^{0}C_s$ (X)	2.9	0.088	Out
${}^{1}C_s$ (X)	0.4	0.511	Out
${}^{2}C_s$ (X)	5.6	0.018	Out
${}^{4}C_s$ (X)	0	0.902	Out
${}^{5}C_s$ (X)	21.9	0	Out
${}^{0}C_s$ (T)	0	0.905	Out
${}^{1}C_s$ (Het)	0.1	0.818	Out
${}^{2}C_s$ (Het)	0.3	0.573	Out
${}^{3}C_s$ (Het)	0	0.943	Out
${}^{4}C_s$ (Het)	0	0.867	Out
${}^{5}C_s$ (Het)	0.3	0.599	Out
${}^{0}C_s$ (H-Het)	0.4	0.533	Out
${}^{1}C_s$ (H-Het)	2.8	0.094	Out
${}^{2}C_s$ (H-Het)	1.9	0.168	Out
${}^{3}C_s$ (H-Het)	2.4	0.124	Out
${}^{4}C_s$ (H-Het)	2.7	0.102	Out
${}^{5}C_s$ (H-Het)	1.6	0.207	Out

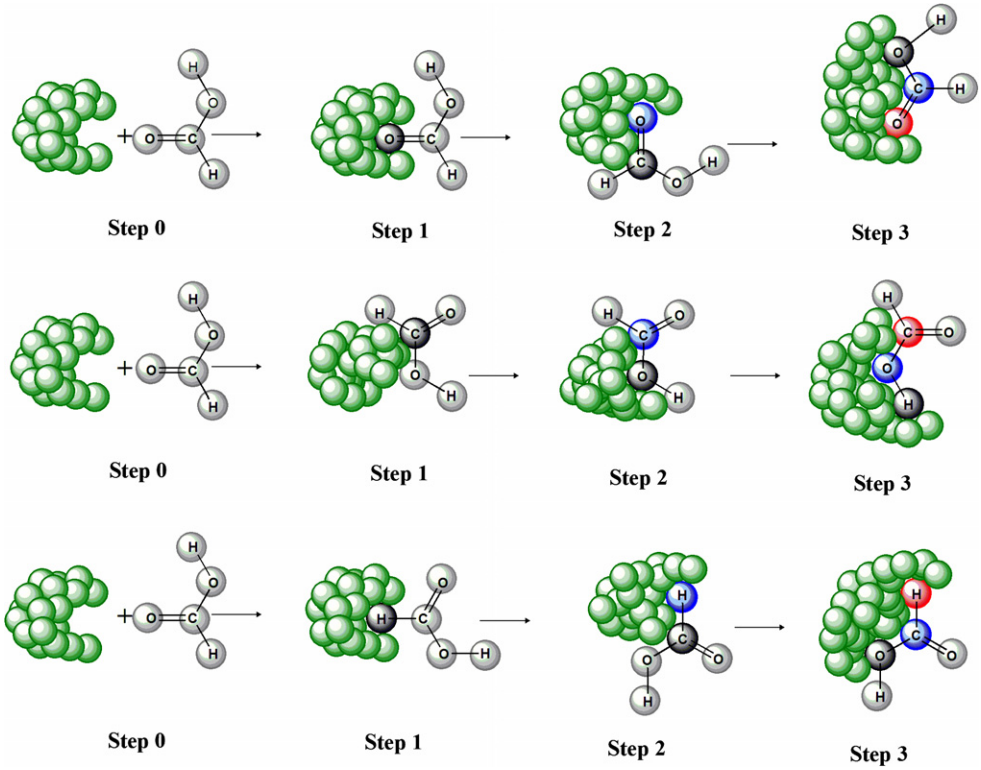

Figure 1. Stochastic drug–target step-by-step interaction.

Table 2. Results of the model, analysis, validation, and virtual screening

	Percent	Antibacterials	Inactive
Analysis			
Antibacterials	84.0	**199**	38
Inactive	84.0	32	**168**
Total	84.0		
Validation			
Antibacterials	83.8	**119**	23
Inactive	82.2	18	**83**
Total	83.0		
Virtual screening			
Antibacterials	84.5	**480**	88

curve (see Fig. 2) to investigate the reliability of the model, with the areas under the curve equal to 0.86 for the predicting series and 0.82 for the training one. These values indicate that the present model gives results that are statistically significant and clearly different from those obtained with a random classifier (area = 0.5). In order to demonstrate how to use the model in practice, we carried out a virtual screening that recognized 480 out of 568 antibacterial compounds (84.5%). These compounds were never used in the training or predicting series.

The most interesting characteristic of the present model is that the $^{k}C_{s}$ used as molecular descriptors depend on both the molecular structure of the drug and the bacterial species against which the drug has to act. The codification of the molecular structure is principally due to the use of the adjacency factor α_{ij} to encode atom–atom bonding, that is, molecular connectivity. The other aspect that allows the molecular structural changes to be encoded is that the atomic contributions $^{0}c_{j}$(s) are atom-class specific. Consequently, one change in the molecular structure, for example, F by O, necessarily involves a change in the interaction. In any case, $^{k}C_{s}$ are the first molecular descriptors reported for antimicrobial QSAR studies that have the ability to discern between a large number of bacterial species. This property is related to the definition of $^{0}c_{j}$(s). The values of these atomic contributions (reported here for the first time for antibacterial action) are given in Table 3 for some atoms and 36 selected species (see Table 2SM in the supplementary material for a detailed list with more than 90 species).

Atomic contributions for antibacterial properties allow not only to distinguish different species (see Table 3) but also to calculate the atomic contributions from different strains of the same species. One advantage of our model is to mark resistant strains or strains that are susceptible to a different drug. For instance, the results in Table 3 show the atomic contributions to antimicrobial action against susceptible and resistant strains of *S. aureus* and *Staphylococcus epidermidis*. For the first of these two species, the regression coefficient between atomic contributions for resistant and susceptible strains is 0.51. Conversely, the regression coefficient is 0.82 for *S. epidermidis*. This notable difference between the two regression coefficients possibly reflects the magnitude of the difference between the respective resistant and susceptible strains. In general, the atomic contributions of

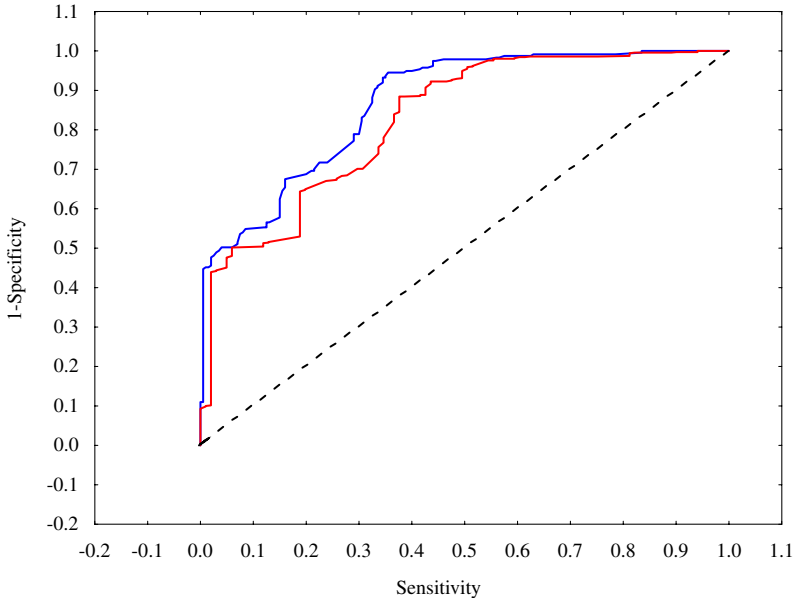

Figure 2. Results for the ROC curve.

Table 3. Some atomic contribution values for atom–receptor interactions including resistant and susceptible species

Bacteria species	C	N	O	H	S	F	Cl
Acinetobacter baumanni	0.22	0.18	0.2	0.23	0.18	0	0
Bacteroides thetaiotaomicron	0.21	0.23	0.2	0.21	0.2	0.3	0
Clostridium perfringens	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Corynebacterium pseudodiphtheriticum	0.15	0.18	0.17	0.12	0.12	0	0.2
Chlamydia trachomatis	0.3	0.3	0.3	0.3	0	0.3	0
Citrobacter freundii	0.18	0.17	0.17	0.18	0.16	0.3	0
Clostridium difficile	0.25	0.25	0.25	0.25	0.17	0.27	0.2
Eikenella corrodens	0.15	0.21	0.12	0.12	0	0.3	0
Enterococcus faecium	0.24	0.22	0.24	0.24	0.19	0.26	0.3
Eubacterium lentum	0.22	0.2	0.21	0.22	0.22	0.3	0.3
Haemophilus influenzae	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Klebsiella oxytoca	0.24	0.23	0.23	0.24	0.22	0.3	0
Legionella pneumophila	0.3	0.3	0.3	0.3	0	0.3	0.3
Listeria monocytogenes	0.3	0.3	0.3	0.3	0.3	0	0.3
Mycobacterium avium	0.1	0.11	0.11	0.05	0	0.14	0
Micoplasma pneumoniae	0.3	0.3	0.3	0.3	0	0.3	0.3
Moraxella catarrhalis	0.3	0.3	0.3	0.3	0.3	0.3	0.3
Morganella morganii	0.25	0.26	0.25	0.26	0.26	0.3	0
Staphylococcus aureus	0.24	0.25	0.24	0.24	0.23	0.31	0.2
MrSa[a]	0.14	0.15	0.14	0.13	0.08	0.22	0.2
MsSa	0.23	0.23	0.22	0.22	0.24	0.22	0
Staphylococcus epidermidis	0.29	0.27	0.28	0.29	0.25	0.3	0.3
MrSe[a]	0.18	0.15	0.19	0.18	0.1	0	0.3
MsSe[a]	0.24	0.24	0.24	0.24	0.24	0.3	0
Staphylococci coagulase-negative	0.15	0.16	0.12	0.13	0.05	0.3	0.1
Staphylococcus haemolyticus	0.24	0.21	0.23	0.24	0.14	0.3	0.3
Streptococcus pneumoniae	0.27	0.28	0.27	0.26	0.29	0.29	0.3
Streptococcus pyogenes	0.26	0.28	0.27	0.27	0.3	0.3	0.3
Streptococci β-hemolytic	0.4	0.28	0.27	0.28	0.3	0.3	0.3

[a] *MrSa*: *Staphylococcus aureus* methicillin resistant, *MrSe*: *Staphylococcus epidermidis* methicillin resistant, *MsSa*: *Staphylococcus aureus* methicillin susceptible, *MsSe*: *Staphylococcus epidermidis* methicillin susceptible.

different atoms to the antibacterial property against all the species studied are connected with one another. The high regression coefficients for some of these contributions are shown in Table 4.

Finally, we have depicted the names of all the drugs used, the bacterial species tested, and detailed results for training and validation in see Table 1SM. The aforementioned flexible definition of the present approach

Table 4. Correlation values of atomic values

	C	N	O	H	S	F	Cl
C	1.00	**0.95**	**0.96**	**0.98**	0.54	0.40	*0.33*
N		1.00	**0.97**	**0.96**	0.45	0.38	*0.23*
O			1.00	**0.98**	0.44	0.28	*0.33*
H				1.00	0.52	0.36	*0.32*
S					1.00	*0.26*	*0.43*
F						1.00	−0.05
Cl							1.00

makes it possible to model very heterogeneous antibacterial activity data for the first time. In fact, the present study is the first reported unified model that allows the prediction of the antibacterial activity of any organic compound against a very large and diverse range of bacterial pathogens. As a concluding remark, and concerning the future research outlook, one can note that the present QSAR methodology may be able to predict the biological activity of drugs in more general situations than the traditional QSAR models.

3. Methods

3.1. Markov model for drug–target step-by-step interaction

We will consider a hypothetical situation in which a drug molecule is free in space at an arbitrary initial time (t_0). It is then of interest to develop a simple stochastic model for a step-by-step interaction between the atoms of a drug molecule and a molecular receptor from the time when the pharmacological effect begins. For the sake of simplicity, we consider a model in which the chemical structure of the receptor is either unknown or is not taken into consideration.

Let the initial contribution of the jth atom to the drug–receptor interaction be ${}^0c_j(\mathrm{s})$. In this term the symbol c represents contribution, the 0 indicates that we refer to the initial atom–receptor interaction, and s indicates that the contribution depends on the specific microbial species. One must then define the contribution ${}^kc_{ij}(\mathrm{s})$ of the interaction between the jth atom and the receptor given that ith atom has interacted at previous time t_k. With respect to ${}^1c_{ij}(\mathrm{s})$ we must take into consideration that once the jth atom has interacted the preferred candidates for the next interaction are ith atoms bound to j by a chemical bond. In particular, immediately after the first interaction takes place ($t_0 = 0$), an interaction ${}^1c_{ij}(\mathrm{s})$ occurs at time $t_1 = 1$ and so on. Therefore, we defined ${}^1c_{ij}(\mathrm{s}) = \alpha_{ij} \cdot {}^0c_j(\mathrm{s})$, where $\alpha_{ij} = 1$ if the jth atom is adjacent to the ith one and $\alpha_{ij} = 0$ otherwise. One can suppose that atoms bind to their receptor in discrete intervals of time t_k. There are several alternative ways in which such a step-by-step binding process may occur. This concept is illustrated in Figure 1.

The Markov Model allowed us to derive the average contributions ${}^kC_s$ of the atoms in the molecule to the gradual interaction between the drug and the receptor at a specific time k in a given microbial species (s). These ${}^kC_s$ values are derived by summing up all the atomic contributions of interaction ${}^0c_j(\mathrm{s})$ pre-multiplied by the absolute probabilities of drug–target interaction ${}^Ap_k(j,\mathrm{s})$[23–25]:

$$ {}^kC_s = \sum_{j=1}^{n} {}^Ap_k(j,\mathrm{s}) \cdot {}^0c_j(\mathrm{s}) \tag{2} $$

Such a model is stochastic per se (probabilistic step-by-step atom–receptor interaction in time) but also considers molecular connectivity (the step-by-step atom union in space throughout the chemical bonding system). The Markov Model for the drug–target step-by-step interaction method was described in a previous paper.[26]

3.2. Statistical analysis

As a continuation of the previous sections, we can attempt to develop a simple linear QSAR using the MARCH-INSIDE methodology, as defined previously, with the general formula:

$$ \mathrm{Actv} = b_0 \cdot {}^0C_s + b_1 \cdot {}^1C_s + b_2 \cdot {}^2C_s + b_3 \cdot {}^3C_s + \cdots + b_k \cdot {}^kC_s + b \tag{3} $$

Here, ${}^kC_s$ act as the microbial species-specific molecule/target interaction descriptors. Linear discriminant analysis (LDA)[18] was selected to fit the classification functions. The model deals with the classification of a set of compounds as active or inactive against different microbial species. A dummy variable (Actv) was used to codify the antimicrobial activity. This variable indicates either the presence (Actv = 1) or absence (Actv = −1) of antimicrobial activity for the drug against the specific species. In Eq. 3, b_k represents the coefficients of the classification function, determined by the least squares method as implemented in the LDA module of the STATISTICA 6.0 software package.[27] Forward stepwise was fixed as the strategy for variable selection.[19,20]

The quality of LDA models was determined by examining Wilk's U statistic, the Fisher ratio (F), and the p-level (p). We also inspected the percentage of good classification and the ratios between the cases and variables in the equation and variables to be explored in order to avoid over-fitting or chance correlation. Validation of the model was corroborated by resubstitution of cases in four predicting series.[26,27]

3.3. Data set

The data set consisted of a set of marketed and/or very recently reported antibacterial drugs with $\mathrm{MIC}_{50} \leqslant 10\ \mu\mathrm{M}$ against different bacteria. The three data sets used were as follows: training series, 199 active compounds plus 168 inactive compounds (367 in total); predicting series, 137 + 106 = 243 in total; virtual screening, 568 active compounds. The literature reports contain experimental test data for each drug against some, but not all, species from a list of 137. Consequently, we were able to collect 1248 cases (drug/species pairs). For the sake of brevity, the names or codes for all compounds, as well as the references consulted, are depicted in Table 1SM of the supplementary material.

Acknowledgments

Authors thank projects PXIB20304PR and BTF20302PR from Xunta de Galiza as well as PI061457 from Ministerio de Sanidad y Consumo for partial financial support. González-Díaz H acknowledges two months contract as guest professor from the Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, Spain (period August–September 2006).

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2006.10.039.

References and notes

1. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley-VCH: Weinheim, Germany, 2000.
2. Fratev, F.; Benfenati, E. *J. Chem. Inf. Model.* **2005**, *45*, 634.
3. Todar's Online Textbook of Bacteriology: http://www.textbookofbacteriology.net/.
4. Kubinyi, H.; Taylor, J.; Ramdsen, C. Quantitative drug design. In *Comprehensive Medicinal Chemistry*; Hansch, C., Ed.; Pergamon, 1990, Vol. 4, p 589.
5. González, M. P.; Morales, A. H.; Molina, R. *Polymer* **2004**, *45*, 2773.
6. Cabrera, M. A.; Bermejo, S. *Bioorg. Med. Chem.* **2004**, *22*, 5833.
7. Marrero-Ponce, Y.; Medina-Marrero, R.; Torrens, F.; Martinez, Y.; Romero-Zaldivar, V.; Castro, E. A. *Bioorg. Med. Chem.* **2005**, *13*, 2881.
8. Marrero-Ponce, Y.; Castillo-Garit, J. A.; Olazabal, E.; Serrano, H. S.; Morales, A.; Castañedo, N.; Ibarra-Velarde, F.; Huesca-Guillen, A.; Sánchez, A. M.; Torrens, F.; Castro, E. A. *Bioorg. Med. Chem.* **2005**, *13*, 1005.
9. Marrero-Ponce, Y.; Montero-Torres, A.; Romero-Zaldivar, C.; Iyarreta-Veitía, M.; Mayón-Peréz, M.; García-Sánchez, R. N. *Bioorg. Med. Chem.* **2005**, *13*, 1293.
10. González-Díaz, H.; Olazábal, E.; Castañedo, N.; Hernádez, S. I.; Morales, A.; Serrano, H. S.; González, J.; Ramos de, A. R. *J. Mol. Mod.* **2002**, *8*, 237.
11. González-Díaz, H.; Gia, O.; Uriarte, E.; Hernádez, I.; Ramos, R.; Chaviano, M.; Seijo, S.; Castillo, J. A.; Morales, L.; Santana, L.; Akpaloo, D.; Molina, E.; Cruz, M.; Torres, L. A.; Cabrera, M. A. *J. Mol. Mod.* **2003**, *9*, 395.
12. González-Díaz, H.; Hernández, S. I.; Uriarte, E.; Santana, L. *Comput. Biol. Chem.* **2003**, *27*, 217.
13. González-Díaz, H.; Ramos de, A. R.; Molina, R. R. *Bull. Math. Biol.* **2003**, *65*, 991.
14. González-Díaz, H.; Uriarte, E.; Ramos de, A. R. *Bioorg. Med. Chem.* **2004**, *13*, 323.
15. González-Díaz, H.; Molina, R. R.; Uriarte, E. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4691–4695.
16. González-Díaz, H.; Ramos de, A. R.; Molina, R. R. *Bioinformatics* **2003**, *19*, 2079.
17. González-Díaz, H.; Molina, R. R.; Uriarte, E. *Polymer* **2004**, *45*, 3845.
18. Ramos de, A. R.; González Díaz, H.; Molina, R.; González, M. P.; Uriarte, E. *Bioorg. Med. Chem.* **2004**, *12*, 4815.
19. Ramos de, A. R.; González-Díaz, H.; Molina, R. R.; Uriarte, E. *Proteins, Struct. Func. and Bioinf.* **2004**, *56*, 715.
20. González-Díaz, H.; Bastida, I.; Castañedo, N.; Nasco, O.; Olazabal, E.; Morales, A.; Serrano, H. S.; Ramos de, A. R. *Bull. Math. Biol.* **2004**, *66*, 1285.
21. González-Díaz, H.; Marrero, Y.; Hernández, I.; Bastida, I.; Tenorio, I.; Nasco, O.; Uriarte, E.; Castañedo, N. C.; Cabrera-Pérez, M. A.; Aguila, E.; Marrero, O.; Morales, A.; González, M. P. *Chem. Res. Tox.* **2003**, *16*, 1318.
22. González-Díaz, H.; Agüero, G.; Cabrera, M. A.; Molina, R.; Santana, L.; Uriarte, E.; Delogu, G.; Castañedo, N. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 551.
23. González-Díaz, H.; Cruz-Monteagudo, M.; Molina, R.; Tenorio, E.; Uriarte, E. *Bioorg. Med. Chem.* **2005**, *13*, 1119.
24. Cruz-Monteagudo, M.; González-Díaz, H. *Eur. J. Med. Chem.* **2005**, doi:10.1016/j.ejmech.2005.04.012.
25. Van Waterbeemd, H. Discriminant Analysis for Activity Prediction. In *Method and Principles in Medicinal Chemistry*; Manhnhold, R., Krogsgaard-Larsen, H., Timmerman, H., Eds.; Chemometric Methods in Molecular Design; Van Waterbeemd, H., Ed.; VCH: Weinhiem, 2005; Vol. 2, pp 265–282.
26. Gonzalez-Diaz, H.; Prado-Prado, F. J.; Santana, L.; Uriarte, E. *Bioorg. Med. Chem.* **2006**, *14*, 5973.
27. STATISTICA for Windows release 6.0. Statsoft Inc., 2001.