Available online at www.sciencedirect.com
SCIENCE DIRECT®
Bioorganic & Medicinal Chemistry 14 (2006) 5973–5980
Bioorganic & Medicinal Chemistry

Unify QSAR approach to antimicrobials. Part 1: Predicting antifungal activity against different species

Humberto González-Díaz,[a,b,*] Francisco J. Prado-Prado,[a] Lourdes Santana[a] and Eugenio Uriarte[a]

[a]*Department of Organic Chemistry, University of Santiago de Compostela 15782, Spain*
[b]*CBQ, Central University of "Las Villas" 54830, Cuba*

Received 23 June 2005; revised 5 May 2006; accepted 15 May 2006
Available online 8 June 2006

Abstract—Most of up-to-date reported molecular descriptors encode only information about the molecular structure. In previous papers, we have extended stochastic descriptors to encode additional information such as target site, partition system, or biological species [*Bioorg. Med. Chem. Lett.* **2005**, *15*, 551; *Bioorg. Med. Chem.* **2005**, *13*, 1119]. This work develops an unify Markov model to describe with a single linear equation the biological activity of 74 drugs tested in the literature against some of the fungi species selected from a list of 87 species (491 cases in total). The data were processed by linear discriminant analysis (LDA) classifying drugs as active or non-active against the different tested fungi species. The model correctly classifies 338 out of 368 active compounds (91.85%) and 89 out of 123 non-active compounds (72.36%). Overall training predictability was 86.97% (427 out of 491 compounds). Validation of the model was carried out by means of leave-species-out (LSO) procedure. After elimination step-by-step of all drugs tested against one specific species, we record the percentage of good classification of leave-out compounds (LSO-predictability). In addition, robustness of the model to the elimination of the compounds (LSO-robustness) was considered. This aspect was considered as the variation of the percentage of good classification of the modified model (Δ) in LSO with respect to the original one. Average LSO-predictability was 86.41 ± 0.95% (average ± SD) and $\Delta = -0.55\%$, being 6 the average number of drugs tested against each fungi species. Results for some of the 87 studied species were *Candida albicans*: 43 tested compounds, 100% of LSO-predictability, $\Delta = -3.49\%$; *Candida parapsilosis* 23, 100%, $\Delta = -0.86\%$; *Aspergillus fumigatus* 21, 95.20%, $\Delta = 0.05\%$; *Microsporum canis* 12, 91.60%, $\Delta = -2.84\%$; *Trichophyton mentagrophytes* 11, 100%, $\Delta = -0.51\%$; *Cryptococcus neoformans* 10, 90%, $\Delta = -0.90\%$. The present one is the first reported unify model that allows one predicting antifungal activity of any organic compound against a very large diversity of fungi pathogens.

1. Introduction

Infections caused by fungi have increased dramatically during the past decades. The reasons for this are manifold. In principle, one has to differentiate between mycoses of the skin or nails and systemic mycoses. Systemic mycoses mainly appear concomitant with other diseases or are caused by treatment with chemotherapeutics. At risk are patients after organ transplantation treated with immunosuppressives or those suffering with a weakened immune system, for example, patients with AIDS. Endocarditic problems caused by fungi have been observed with drug abuse. So far, many of the new compounds could not be introduced into therapy, mainly for two reasons: (i) poor selectivity and (ii) problems in getting into the cytoplasm of fungal cells and/or in gastrointestinal absorption.[1]

Consequently, there is an increasing interest on the development of rational approaches for antifungal drug discovery. In this sense, a very important role may be played by computer-aided drug design techniques based on quantitative-structure–activity-relationships (QSAR) studies. Unfortunately, almost antifungal activity QSAR studies reported up-to-date are based on very limited databases considering only structurally parent compounds acting against one single fungus species.[2] As a result, the researcher interested on predicting the antifungal activity for a given series of compounds needs to use/develop as many QSAR equations as combinations of families of compounds versus fungus species are necessary to be predicted. Therefore, the development of one single unified equation explaining the anti-

Keywords: QSAR; Markov model; Antifungal drugs; Linear discriminant analysis; Antimicrobials.
* Corresponding author. Tel.: +34 981 563100; fax: +34 981 594912; e-mail addresses: gonzalezdiazh@yahoo.es; humbertogd@uclv.edu.cu

0968-0896/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmc.2006.05.018

fungal activity of structurally heterogeneous series of compounds against as many as possible fungus species is of major interest.

There are more than 1600 molecular descriptors that may be in principle generalized and used to attack the above-mentioned problem.[3–6] In addition, other QSAR approaches have been introduced recently with demonstrated utility in medicinal chemistry.[7–9] In any case, no one of these indices has been extended yet to encode additional information to chemical structure. Our group has introduced elsewhere one Markov model (MM) encoding molecular backbone information, with several applications in bioorganic medicinal chemistry. The method was named the MARCH-INSIDE approach, MARkovian CHemicals IN SIlico Design. It allowed us introducing matrix invariants such as stochastic entropies and spectral moments for the study of molecular properties.[10–14]

In recent studies, the MARCH-INSIDE method has been extended to encompass interesting molecular environment information in addition to molecular structure.[15,16] This approach is able to take into consideration, for instance, not only the molecular structure of the drug but its average propensity to act against the specific microbial organism the drug has to eliminate too. The present study develops a single linear equation based on these previous ideas to predict the antifungal activity of drugs against different species.

2. Results and discussion

One of the main advantages of the present stochastic approach is the possibility of deriving average structural parameters depending on the probability of the states of the MM.[15–17] In specific, this work introduces for the first time a single linear QSAR equation model to predict the antifungal activity of drugs against different species. The best model found was:

$$\begin{aligned} \text{Actv} = & -2.88\,{}^{0}C_s(X) + 1.26 \cdot {}^{5}C_s(X) \\ & - 1.01 \cdot {}^{0}C_s(T) - 0.78 \cdot {}^{0}C_s(C_{\text{unsatd}}) \\ & + 0.94 \cdot {}^{3}C_s(X) - 0.76 \cdot {}^{4}C_s(T) - 1.17 \\ & \lambda = 0.53,\ F(6,484) = 71.93,\ p < 0.001. \end{aligned} \quad (1)$$

In the model, the coefficient λ is the Wilk's statistics, statistic for the overall discrimination, F is the Fisher ratio, and p is the error level. In this equation, ${}^{k}C_s$ 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 as, for instance, are: halogens (X) or unsaturated carbon atoms (C_{unsatd}). The model correctly classifies 338 out of 368 active compounds (91.85%) and 89 out of 123 non-active compounds (72.36%). Overall training predictability was 86.97% (427 out of 491 compounds).

One of the more serious limitations of the present model is that for some species not many drugs have been biologically tested. Conversely, a huge number of antifungal activity assays have been reported for other species. In addition, some drugs are small sized with not many atoms but other molecules have many of them. In consequence, ${}^{0}c_j(s)$ have different degrees of variability (CV) being lower than 10% for some species and higher than 100% for others (see Table 1 for details). Naturally, the confidence of the probability of action may be affected in this sense. In any case, this possible limitation is an aspect that the researchers have to take into consideration when using the model. Models often misclassify some compounds and this is also a fact we have to take into consideration when using it. One solution, classically accepted in these cases, is checking the Mahalanobis distance (D) of the molecule with respect to the centroids of both groups (active and non-active compounds). In the case of abnormal values for D (very large values) this prediction must not to be used. In any case, we carried out a jack-knife study eliminating some species in order to give an idea of the stability of the model under data variation.[18]

Validation of the model was carried out by means of leave-species-out (LSO) procedure. After elimination step-by-step of all drugs tested against one specific species we record the percentage of good classification of leave-out compounds (LSO-predictability). In addition robustness of the model to the elimination of the compounds (LSO-robustness) was considered. This aspect was measured as the variation of the percentage of good classification of the modified model in LSO (Δ) with respect to the original one. Average LSO-predictability was 86.41 ± 0.95% (average ± SD) and $\Delta = -0.55\%$, being 6 the average number of drugs tested against each fungi species. Results for some of the 87 studied species were *Candida albicans*: 43 tested compounds, 100% of LSO-predictability, $\Delta = -3.49\%$; *Candida parapsilosis* 23, 100%, $\Delta = -0.86\%$; *Aspergillus fumigatus* 21, 95.20%, $\Delta = 0.05\%$; *Microsporum canis* 12, 91.60%, $\Delta = -2.84\%$; *Trichophyton mentagrophytes* 11, 100%, $\Delta = -0.51\%$; *Cryptococcus neoformans* 10, 90%, $\Delta = -0.90\%$. Table 2 depicts the LSO validation results for the model.

The more interesting characteristic of the present model is that the molecular descriptors depend both on the molecular structure of the drug and on the fungus species the drug have to act against. The codification of the molecular structure is in first place due to the use of the adjacency factor α_{ij} to encode atom–atom bond-

Table 1. Variability of selected atoms in all the molecules tested against some fungi species

Fungi	CV[a]			
	C	H	N	O
Coprinus sp.	16.21	37.56	71.32	90.07
Absidia coryambifera	16.42	33.06	77.67	63.59
Wangiella dermatitidis	36.09	36.83	62.45	21.65
Chaetomium nigrocolor	39.43	64.46	53.35	123.26
Aspergillus terreus	51.60	82.31	70.31	129.86
Aspergillus versicolor	51.60	82.31	70.31	129.86
Unidentified basidiomycetes	110.38	74.48	115.70	40.07
Fusarium oxysporum	126.49	87.58	102.17	60.54
Trichophyton erinacei	202.89	131.57	149.03	82.04
Microsporum audouinii	209.35	128.13	183.36	69.48

[a] CV = (standard deviation/mean) × 100.

Table 2. Number of drugs tested by species (#), LSO predictability (% LSO), and LSO robustness (Δ)

Species	No.	% LSO	Δ
Candida albicans	43	100	−3.49
Candida parapsilosis	23	100	−0.86
Aspergillus fumigatus	21	95.2	0.05
Candida guilliermondii	20	90	−5.02
Candida tropicalis	17	94.11	−0.68
Candida glabrata	15	93.3	−3.15
Candida krusei	15	93.3	−1.68
Candida dubliniensis	14	100	−0.39
Aspergillus flavus	12	100	−0.57
Microsporum canis	12	91.6	−2.84
Trichophyton mentagrophytes	11	100	−0.51
Cryptococcus neoformans	10	90	−0.90
Saccharomyces cerevisiae	8	87.5	−0.63
Trichophyton violaceum	8	100	−0.43
Aspergillus nidulans	7	100	−2.87
Aspergillus terreus	7	85.7	−0.20
Microsporum cookei	7	85.7	−1.43
Aspergillus niger	6	83.3	−0.79
Candida kefyr	6	100	−0.78
Candida lusitaniae	6	100	−0.38
Malassezia furfur	6	100	−0.38
Microsporum audouinii	6	100	−0.38
Microsporum racemosum	6	100	−0.57
Scedosporium prolificans	6	100	0.63
Trichophyton erinacei	6	100	−0.58
Trichophyton rubrum	6	100	−0.58
Trichophyton verrucosum	6	100	0.04
Aspergillus spp.	5	100	−0.97
Aspergillus sydowii	5	100	0.48
Microsporum ferrugineur	5	100	−0.56
Microsporum fulvum	5	100	−0.55
Microsporum gypseum	5	100	−0.14
Mucor spp.	5	100	−1.38
Rhodotorula glutinis	5	100	−1.37
Scedosporium apiospermum	5	100	−0.76
Trichophyton interdigitale	5	100	1.09
Trichophyton simii	5	100	0.89
Achaetomium strumarium	4	100	−0.11
Apophysomyces elegans	4	100	−1.37
Aspergillus versicolor	4	100	−0.53
Bjerkandera adusta	4	100	−0.73
Cryptococcus laurentii	4	100	−0.73
Microsporum praecox	4	75	−1.35
Paecilomyces lilacinus	4	100	−1.14
Paecilomyces variotii	4	100	−1.14
Sporobolomyces salmonicolor	4	100	0.09
Trichophyton balcaneum	4	100	0.91
Trichophyton concentricum	4	100	0.09
Trichophyton phaseoliforme	4	100	0.91
Trichosporon asahii	4	100	−0.52
Unidentified basidiomycetes	4	100	0.91
Bipolaris spp.	3	100	−0.50
Blastomyces dermatitidis	3	100	−0.70
Coprinus species	3	100	−0.50
Cunninghamella spp.	3	66.6	−0.29
Fusarium spp.	3	100	−0.50
Malassezia sympodialis	3	100	−0.50
Saksenaea vasiformis	3	100	0.73
Trichophyton ajelloi	3	100	−0.70
Trichophyton tonsurans	3	100	−0.29
Trichosporon spp.	3	100	−0.49
Absidia coryambifera	2	50.0	−0.67
Acremonium spp.	2	100	−0.06
Aspergillus flavus	2	100	−0.07
Aspergillus ustus	2	100	0.14
Candida famata	2	100	0.55
Candida neoformans	2	100	−0.65
Chrysosporium spp.	2	100	−0.68
Coccidioides immitis	2	100	−0.47
Microsporum gallinae	2	100	−0.47
Paecilomyces spp.	2	50	−0.87
Rhizopus spp.	2	100	0.76
Aspergillus candidus	1	100	−0.07
Aspergillus ochraceus	1	100	−0.62
Chaetomium atrobrunneum	1	100	−0.62
Chaetomium globosum	1	100	−0.62
Chaetomium nigrocolor	1	100	−0.62
Cokeromyces recurvatus	1	100	−0.42
Fusarium oxysporum	1	100	−0.41
Malassezia pachydermatis	1	100	−0.42
Malassezia slooffiae	1	100	−0.42
Microsporum nanum	1	100	−0.21
Rhizopus oryzae	1	100	−1.02
Schizophyllum commune	1	100	0.79
Trichophyton schoenleinii	1	100	−0.24
Wangiella dermatitidis	1	100	−0.65

ing, molecular connectivity. The other aspect that allows encoding molecular structural changes is that the probabilities calculated are atom class specific. Consequently, one change in the molecular structure of, for example, F by O or O sp^2 by O sp^3 necessarily implies a change in the propensity of interaction. One can note certain homology with the present contributions and free energy[20] but we refrain from using the term contribution here to avoid further misunderstandings. In any case, the more interesting fact is that ${}^{k}\Delta C_s$ are the first molecular descriptors reported for antimicrobial QSAR studies with the skill of discerning among a large number of species. This property is related to the definition of the ${}^{0}c_j(s)$. The values of these species specific atomic contributions reported herein by the first time are given in Table 3 for some atoms and more than 80 species.

As a consequence of the above mentioned flexible definition of the present approach was possible to model by the first time a very heterogeneous a diverse data. The posterior validation probabilities (LSO) predicted for every drug–species pair are depicted in Table 1SM of supplementary material file. The present is the first reported unify model that allows one predicting antifungal activity of any organic compound against a very large diversity of fungi pathogens. As a sort of concluding remark and future research directions one may note that the present QSAR methodology may be able to predict biological activity of drugs in more general situations and to more specific tools than traditional QSAR models may be.

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 the space at an arbitrary initial

Table 3. Some atomic contributions values for atom-receptor interactions

Fungi	C	H	N	O	Cl	F
Absidia coryambifera	0.26	0.27	0.21	0.28	0.30	0.11
Achaetomium strumarium	0.55	0.55	0.57	0.44	0.85	0.54
Acremonium spp.	0.94	0.95	0.96	0.94	1.10	0.87
Apophysomyces elegans	0.22	0.23	0.18	0.26	0.30	0.15
Aspergillus candidus	0.66	0.65	0.69	0.63	0.60	0.73
Aspergillus flavipes	0.72	0.69	0.77	0.65	0.60	0.82
Aspergillus flavus	0.21	0.19	0.19	0.20	0.29	0.23
Aspergillus fumigatus	0.00	0.00	0.00	0.00	0.00	0.00
Aspergillus glaucus	0.60	0.60	0.60	0.60	0.60	0.60
Aspergillus nidulans	0.02	0.01	0.02	0.00	0.00	0.05
Aspergillus niger	0.03	0.03	0.03	0.03	0.03	0.04
Aspergillus ochraceus	0.42	0.40	0.44	0.48	0.18	0.52
Aspergillus spp.	0.11	0.11	0.10	0.12	0.12	0.09
Aspergillus sydowii	1.75	1.73	1.81	1.64	1.94	1.74
Aspergillus terreus	0.64	0.61	0.66	0.58	0.75	0.78
Aspergillus ustus	0.24	0.21	0.33	0.13	0.30	0.22
Aspergillus versicolor	0.93	0.95	0.99	0.95	0.88	0.89
Bipolaris spp.	0.99	1.03	0.94	1.05	1.00	0.93
Bjerkandera adusta	0.60	0.59	0.57	0.59	0.60	0.69
Blastomyces dermatitidis	1.34	1.34	1.38	1.33	1.23	1.40
Candida albicans	0.03	0.03	0.02	0.05	0.04	0.01
Candida dubliniensis	0.01	0.01	0.01	0.01	0.01	0.00
Candida famata	0.33	0.31	0.27	0.33	0.30	0.44
Candida glabrata	0.25	0.25	0.20	0.27	0.26	0.19
Candida guilliermondii	0.32	0.31	0.33	0.34	0.37	0.30
Candida kefyr	0.08	0.07	0.06	0.01	0.12	0.03
Candida krusei	0.02	0.02	0.02	0.02	0.01	0.03
Candida lusitaniae	0.03	0.02	0.03	0.02	0.02	0.07
Candida neoformans	0.29	0.27	0.28	0.26	0.18	0.65
Candida parapsilosis	0.01	0.00	0.01	0.00	0.00	0.02
Candida tropicalis	0.02	0.02	0.02	0.02	0.03	0.02
Candida utilis	0.22	0.22	0.24	0.41	0.30	0.30
Chaetomium atrobrunneum	0.26	0.27	0.27	0.20	0.48	0.23
Chaetomium globosum	0.22	0.24	0.19	0.19	0.30	0.09
Chaetomium nigrocolor	0.27	0.26	0.29	0.20	0.60	0.30
Chrysosporium spp.	1.23	1.24	1.37	1.12	1.23	1.33
Coccidioides immitis	0.24	0.23	0.19	0.22	0.30	0.15
Cokeromyces recurvatus	0.49	0.47	0.42	0.36	0.81	0.36
Coprinus species	0.12	0.15	0.15	0.25	0.00	0.00
Cryptococcus laurentii	0.06	0.05	0.03	0.08	0.12	0.09
Cryptococcus neoformans	0.02	0.02	0.01	0.02	0.03	0.01
Cunninghamella spp.	1.32	1.37	1.11	1.46	1.57	0.15
Epidermophyton floccosum	0.21	0.20	0.25	0.13	0.21	0.26
Fusarium oxysporum	0.52	0.42	0.44	0.17	0.90	0.00
Fusarium solani	0.13	0.10	0.18	0.06	0.22	0.40
Fusarium spp.	0.23	0.20	0.31	0.16	0.54	0.30
Madurella mycetomatis	0.05	0.03	0.10	0.02	0.08	0.90
Malassezia furfur	0.26	0.25	0.22	0.33	0.39	0.23
Malassezia pachydermatis	0.50	0.51	0.45	0.60	0.41	0.61
Malassezia slooffiae	0.29	0.29	0.28	0.28	0.30	0.27
Malassezia sympodialis	0.22	0.23	0.14	0.27	0.22	0.11
Microsporum audouinii	0.25	0.26	0.22	0.25	0.30	0.19
Microsporum canis	0.16	0.16	0.12	0.20	0.28	0.06
Microsporum cookei	0.35	0.36	0.32	0.29	0.32	0.30
Microsporum ferrugineur	0.30	0.30	0.29	0.28	0.30	0.30
Microsporum fulvum	0.16	0.18	0.15	0.22	0.08	0.15
Microsporum gallinae	0.28	0.29	0.26	0.27	0.30	0.23
Microsporum gypseum	0.18	0.19	0.22	0.26	0.16	0.25
Microsporum nanum	0.23	0.25	0.25	0.26	0.08	0.30
Microsporum praecox	0.37	0.36	0.41	0.29	0.15	0.63
Microsporum racemosum	0.12	0.14	0.07	0.19	0.00	0.11
Mucor spp.	0.73	0.74	0.61	0.77	0.93	0.54
Paecilomyces lilacinus	0.07	0.08	0.07	0.09	0.08	0.04
Paecilomyces spp.	0.30	0.30	0.30	0.30	0.30	0.30
Paecilomyces variotii	0.06	0.06	0.06	0.06	0.03	0.08

(*continued on next page*)

Table 3 (*continued*)

Fungi	C	H	N	O	Cl	F
Rhizopus oryzae	0.72	0.73	0.58	0.73	0.85	0.78
Rhizopus spp.	0.96	0.99	0.90	1.02	1.10	0.78
Rhodotorula glutinis	0.16	0.15	0.16	0.15	0.22	0.19
Saccharomyces cerevisiae	0.13	0.12	0.17	0.12	0.30	0.25
Saksenaea vasiformis	0.35	0.29	0.42	0.21	0.40	0.57
Scedosporium apiospermum	0.10	0.10	0.15	0.10	0.06	0.15
Scedosporium prolificans	0.09	0.08	0.10	0.07	0.06	0.12
Schizophyllum commune	0.55	0.51	0.47	0.39	0.90	0.38
Sporobolomyces salmonicolor	0.29	0.28	0.27	0.21	0.64	0.25
Trichophyton ajelloi	0.20	0.20	0.19	0.25	0.23	0.21
Trichophyton balcaneum	0.28	0.29	0.26	0.27	0.30	0.23
Trichophyton concentricum	0.28	0.29	0.26	0.27	0.30	0.23
Trichophyton erinacei	1.64	1.73	1.57	1.83	1.13	1.65
Trichophyton interdigitale	0.47	0.47	0.51	0.37	0.32	0.73
Trichophyton mentagrophytes	0.16	0.16	0.12	0.20	0.28	0.06
Trichophyton phaseoliforme	0.47	0.47	0.51	0.37	0.32	0.73
Trichophyton rubrum	0.16	0.17	0.13	0.22	0.28	0.06
Trichophyton schoenleinii	0.28	0.29	0.26	0.27	0.30	0.23
Trichophyton simii	0.33	0.33	0.34	0.31	0.30	0.36
Trichophyton tonsurans	0.30	0.30	0.30	0.30	0.30	0.30
Trichophyton verrucosum	0.34	0.35	0.32	0.29	0.32	0.30
Trichophyton violaceum	0.21	0.23	0.21	0.28	0.10	0.28
Trichosporon asahii	0.08	0.07	0.12	0.05	0.22	0.15
Trichosporon spp.	0.21	0.23	0.18	0.26	0.30	0.15
Unidentified basidiomycetes	0.28	0.28	0.23	0.29	0.30	0.21
Wangiella dermatitidis	0.25	0.23	0.27	0.30	0.00	0.00

time (t_0). It is then interesting to develop a simple stochastic model for a step-by-step interaction between the atoms of a drug molecule and a molecular receptor at the time of beginning of the pharmacological effect. For the sake of simplicity, we are going to consider from now on a model in which the chemical structure of the receptor is unknown or not taken into consideration.

Let the initial contribution of the *j*th atom to the drug–receptor interaction is ${}^0c_j(s)$. In this symbol, the *c* points to contribution, the 0 indicates that we refer to the initial atom–receptor interaction, and the *s* indicates that the contribution depends on the specific microbial species. Afterwards, we have to define the contribution ${}^kc_{ij}(s)$ of interaction between the *j*th atom and the receptor given that *i*th atom has been interacted at previous time t_k. With respect to ${}^1c_{ij}(s)$ we must take into consideration that once the *j*th atom has interacted the preferred candidates for the next interaction are such *i*th atoms bound to *j* by a chemical bond. In particular, immediately after the first interaction ($t_0 = 0$) takes place an other interaction ${}^1c_{ij}(s)$ at time $t_1 = 1$, and so on. In consonance, we defined ${}^1c_{ij}(s) = \alpha_{ij} \cdot {}^0c_j(s)$, being $\alpha_{ij} = 1$ if the *j*th atom is adjacent to the *i*th one and $\alpha_{ij} = 0$ otherwise. So, one can suppose that, atom binds to its receptor in discrete intervals of time t_k. There several alternative ways in which such a step-by-step binding process may occur. Figure 1 illustrates this idea.[15–17]

The contribution ${}^0c_j(s)$ will be considered here as a function of the absolute temperature of the system and the equilibrium local frequency of interaction between the *j*th atom and the receptor ${}^0\Gamma_j(s)$ for a given microbial species. Additionally, ${}^1c_{ij}(s) = \alpha_{ij} \cdot {}^0c_j(s)$ can be defined by analogy from $\Gamma_{ij}(s)$:[15–17]

$$ {}^0c_j(s) = -R \cdot T \cdot \log {}^0\Gamma_j(s) \tag{2} $$

$$ {}^1c_{ij}(s) = -R \cdot T \cdot \log {}^1\Gamma_{ij}(s) \tag{3} $$

The present approach to drug–receptor interaction has two main drawbacks. The first is the difficulty on the definition of the constants. In this work, we solve the first question estimating ${}^0\Gamma_j(s)$ as the rate of occurrence $n_j(s)$ of the *j*th atom on active molecules against a given specie with respect to the number of atoms of the *j*th class in the molecules tested against the same specie $n_T(s)$. Both constants can be then written down as:[15–17]

$$ {}^0\Gamma_j(s) = \left(\frac{n_j(s)}{n_T(s)} + 1\right) = e^{\frac{{}^0c_j(s)}{R \cdot T}} \tag{4} $$

$$ {}^1\Gamma_{ij}(s) = \left(\alpha_{ij} \cdot \frac{n_j(s)}{n_T(s)} + 1\right) = e^{\frac{{}^1c_{ij}(s)}{R \cdot T}} \tag{5} $$

where, as explained α_{ij} coincide with the elements of the atom adjacency matrix, and $n_j(s)$, $n_T(s)$, ${}^0c_j(s)$, and ${}^1c_{ij}(s) = \alpha_{ij} \cdot {}^0c_j(s)$ have been defined in the paragraph above, R is the universal gas constant, and T is the absolute temperature. The number 1 is added to avoid scale and logarithmic function's definition problems. The second problem relates to the description of the interaction process at higher times $t_k > t_1$. Therefore, MM theory enables a simple calculation of the probabilities with which the drug–receptor interaction takes place in the time until the studied effect is achieved. In this work, we are going to focus on drugs–microbial structure less target interaction. As depicted in Figure 1, this model deals with the calculation of the probabilities (${}^kp_{ij}$) with which any

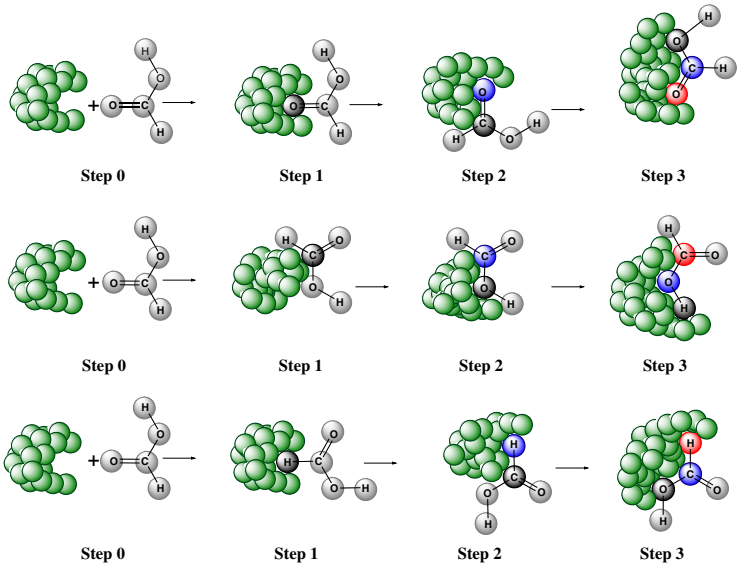

Figure 1. Stochastic drug–target step-by-step interaction.

arbitrary molecular atom *j*th binds to the structureless molecular receptors given that other atom *i*th has been bound before; along discrete time periods t_k ($k = 1, 2, 3, \ldots$); ($k = 1$ in grey), ($k = 2$ in blue) and ($k = 3$ in red) throughout the chemical bonding system. The procedure described here considers as states of the MM the atoms of the molecule. The method arranges all the ${}^0c_j(s)$ values in a vector **C(s)** and all the ${}^1c_{ij}(s) = \alpha_{ij} \cdot {}^0c_j(s)$ as a squared table of $n \times n$ dimension. After normalization of both the vector and the matrix we can build up the corresponding absolute initial probability vector $\boldsymbol{\varphi}(\mathbf{s})$ and the stochastic matrix ${}^1\boldsymbol{\Pi}(\mathbf{s})$, which has the elements ${}^Ap_0(j,s)$ and ${}^1p_{ij}(s)$, respectively. The elements ${}^Ap_0(j,s)$ of the above-mentioned vector $\boldsymbol{\varphi}(\mathbf{s})$ constitute the absolute probabilities with which the *j*th atom interacts with the molecular target or receptor in the species *s* at the initial time with respect to any atom in the molecule:[15–17]

$$
\begin{aligned}
{}^Ap_0(j,s) &= \frac{{}^0c_j(s)}{\sum_{a=1}^{m} {}^0c_a(s)} \\
&= \frac{-RT \cdot \log\left(\frac{n_j(s)}{n_T(s)} + 1\right)}{\sum_{a=1}^{m} - RT \cdot \log\left(\frac{n_a}{n_T(s)} + 1\right)} \\
&= \frac{\log\left(\frac{n_j(s)}{n_T(s)} + 1\right)}{\sum_{a=1}^{m} \log\left(\frac{n_a}{n_T(s)} + 1\right)}
\end{aligned}
\tag{6}
$$

where *m* represents all the atoms in the molecule including the *j*th, n_a is the rate of occurrence of any atom a including the *j*th with value n_j. On the other hand, the matrix is called the 1-step drug–target interaction stochastic matrix. ${}^1\boldsymbol{\Pi}(\mathbf{s})$ is built too as a squared table of order *n*, where *n* represents the number of atoms in the molecule. The elements ${}^1p_{ij}(s)$ of the 1-step drug–target interaction stochastic matrix are the binding probabilities with which a *j*th atom binds to a structureless molecular receptor, given that other *i*th atoms have been interacted before at time $t_1 = 1$ (considering $t_0 = 0$):[15–17]

$$
\begin{aligned}
{}^1p_{ij}(s) &= \frac{\alpha_{ij} \cdot {}^0c_j(s)}{\sum_{a=1}^{m} \alpha_{ia} \cdot {}^0c_a(s)} \\
&= \frac{\alpha_{ij} \cdot (-RT) \cdot \log\left(\frac{n_j(s)}{n(s)} + 1\right)}{\sum_{a=1}^{n} \alpha_{ia} \cdot (-RT) \cdot \log\left(\frac{n_a(s)}{n_T(s)} + 1\right)} \\
&= \frac{\alpha_{ij} \cdot \log\left(\frac{n_j(s)}{n_T} + 1\right)}{\sum_{a=1}^{n} \alpha_{ia} \cdot \log\left(\frac{n_j(s)}{n_T(s)} + 1\right)}
\end{aligned}
\tag{7}
$$

By using, $\boldsymbol{\varphi}(\mathbf{s})$, ${}^1\boldsymbol{\Pi}(\mathbf{s})$ and Chapman–Kolgomorov equations one can describe further evolution of the system.[10–17] Summing up all the atomic contributions of interaction ${}^0c_j(s)$ pre-multiplied by the absolute probabilities of drug–target interaction ${}^Ap_k(j,s)$ one can derive the average changes in contribution ${}^kC_s$ 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*):[15]

$$
\begin{aligned}
{}^kC_s &= \boldsymbol{\varphi}(\mathbf{s}) \cdot {}^k\boldsymbol{\Pi}(\mathbf{s}) \cdot {}^0\mathbf{C}(\mathbf{s}) = \boldsymbol{\varphi}(\mathbf{s}) \cdot \left[{}^1\boldsymbol{\Pi}(\mathbf{s})\right]^k \cdot {}^0\mathbf{C}(\mathbf{s}) \\
&= \sum_{\mathrm{j}=1}^{\mathrm{n}} {}^k\mathrm{c_j}(s) = \sum_{\mathrm{j}=1}^{\mathrm{n}} {}^Ap_k(j,s) \cdot {}^0\mathrm{c_j}(s)
\end{aligned}
\tag{8}
$$

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).

Notably, we should expect different predictions for sp^3 oxygen than for sp^2 one. Conversely, the contributions ${}^0c_j(s)$ and ${}^1c_{\mathrm{ij}}(s) = \alpha_{\mathrm{ij}} \cdot {}^0c_{\mathrm{j}}(s)$ of an atom for instance oxygen ${}^0c_{\mathrm{O}}(s)$ get the same values for all different oxygen types. However, is straightforward to realize that probabilities involving a specific atom are different for differ-

ent states of this atom. Each probability ${}^1p_{ij}$ is calculated dividing the conditional contribution of the atom ${}^1c_{ij}(s) = \alpha_{ij} \cdot {}^0c_j(s)$ by the sum of the contributions of the atoms chemically bound to it and itself. Consequently, a probability formula involving oxygen sp^3 in the matrix for the molecule of methyl alcohol is ${}^1p_{CO} = {}^1c_{CO}(s)/({}^1c_{OO}(s) + {}^1c_{CO}(s) + {}^1c_{HO}(s))$, while for the oxygen sp^2 in acetone we have the equation ${}^1p_{CO} = {}^1c_{CO}(s)/({}^1c_{OO}(s) + {}^1c_{CO}(s))$. As a result, the probabilities and therefore the predictions for sp^3 oxygen are different from those for the sp^2 oxygen and the same apply for any other atoms. This aspect has been discussed before for electronegativities, please see, for instance, the page 398 in Ref. 11.

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:

$$\text{Actv} = b_0 \cdot {}^0C_s + b_1 \cdot {}^1C_s + b_2 \cdot {}^2C_s + b_3 \cdot {}^3C_s \cdots + b_k \cdot {}^kC_s + b \quad (9)$$

Here, ${}^kC_s$ act as the microbial species-specific molecule–target interaction descriptors. We selected linear discriminant analysis (LDA)[18] to fit the classification functions. The model deals with the classification of a set of compounds as active or not 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 of the drug against the specific species. In Eq. (9), b_k represents the coefficients of the classification function, determined by the least square method as implemented in the LDA module of the STATISTICA 6.0 software package.[19] Forward stepwise was fixed as the strategy for variable selection.[18,19]

The quality of LDA models was determined by examining Wilk's U statistic, 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 re-substitution of cases in four predicting series.[19]

3.3. Data set

The data set was conformed by a set of marketed and/or very recently reported antifungal drugs which reported low MIC_{50} <10 μM against different fungi. The data set was conformed by 74 different drugs experimentally tested against some species of a list of 87. Not all drugs were tested in the literature against all listed species so we were able to collect 491 cases (drug/species pairs) instead of 74×87 cases. The names or codes and activity for all compounds as well as the references used to collect it are depicted in Table 1SM of supplementary material file.

Acknowledgments

The authors would like to acknowledge the editor and the unknown referees for kind attentions and changes suggested. González-Díaz H acknowledges funds from program 'Axuda para a incorporacion de investigadores tecnologos/visitantes da CONSELLEREIA DE INOVACION, INDUSTRIA E COMERCIO, IN8061 2005/63-0.' This author also expresses his gratitude to the Laboratory of Medicinal Chemistry, Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, Spain, for financial support. Last but not least, the authors sincerely acknowledge kind attention and useful comments from the editor Professor Herbert Waldmann as well as from two unknown referees.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmc.2006.05.018.

References and notes

1. Otzen, T.; Wempe, E. G.; Kunz, B.; Bartels, R.; Lehwark-Yvetot, G.; Hansel, W.; Schaper, K. J.; Seydel, J. K. *J. Med. Chem.* **2004**, *47*, 240.
2. Fratev, F.; Benfenati, E. *J. Chem. Inf. Model.* **2005**, *45*, 634.
3. Todeschini, R.; Consonni, V. *Handbook of Molecular Descriptors*; Wiley VCH: Weinheim, Germany, 2000.
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.; Hernández, 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.; Uriarte, E.; Ramos de, A. R. *Bioorg. Med. Chem.* **2005**, *13*, 323.
13. González-Díaz, H.; Molina, R. R.; Uriarte, E. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 4691.
14. Ramos de, A. R.; González Díaz, H.; Molina, R.; González, M. P.; Uriarte, E. *Bioorg. Med. Chem.* **2004**, *12*, 4815.
15. 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.

16. González-Díaz, H.; Cruz-Monteagudo, M.; Molina, R.; Tenorio, E.; Uriarte, E. *Bioorg. Med. Chem.* **2005**, *13*, 1119.
17. Cruz-Monteagudo, M.; González-Díaz, H. *Eur. J. Med. Chem.* **2005**, *40*, 1030.
18. Van Waterbeemd, H. Discriminant analysis for activity prediction. In *Method and Principles in Medicinal Chemistry*; Manhnhold, R., Krogsgaard-Larsen, H., Timmerman, , Eds.; Chemometric Methods in Molecular Design; Van Waterbeemd, H., Ed.; VCH: Weinhiem, 1995; Vol. 2, pp 265–282.
19. STATISTICA for Windows release 6.0. Statsoft Inc., 2001.
20. Villa, A.; Zangi, R.; Pieffet, G.; Mark, A. E. *J. Comput. Aided Mol. Des.* **2003**, *17*, 673.