Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Medicinal Chemistry 40 (2005) 1030–1041
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
www.elsevier.com/locate/ejmech

Original article

Unified drug–target interaction thermodynamic Markov model using stochastic entropies to predict multiple drugs side effects

Maykel Cruz-Monteagudo [a,c,*], Humberto González-Díaz [b,c]

[a] *Applied Chemistry Research Center, Central University of "Las Villas", Santa Clara 54830, Cuba*
[b] *Department of Organic Chemistry, Faculty of Pharmacy, University of Santiago de Compostela, Santiago de Compostela 15782, Spain*
[c] *Chemical Bioactives Center, Central University of "Las Villas", Santa Clara 54830, Cuba*

Received 4 February 2005; received in revised form 12 April 2005; accepted 27 April 2005

Available online 13 June 2005

Abstract

Most of present molecular descriptors consider just the molecular structure. In the present article we pretend extending the use of Markov chain (MC) models to define novel molecular descriptors, which consider in addition other parameters like target site or toxic effect. Specifically, this molecular descriptor takes into consideration not only the molecular structure but the specific system the drug affects too. Herein, it is developed a general Markov model that describes 21 different drugs side effects grouped in 10 affected biological systems for 193 drugs, being 311 cases finally. The data were processed by linear discriminant analysis (LDA) classifying drugs according to their specific side effects, forward stepwise was fixed as strategy for variables selection. The average percentage of good classification and number of compounds used in the training/predicting sets were 92.6/91.7% for cardiovascular manifestation (25 out of 27)/(18 out of 20); 89.3/83.9% for dermal manifestations (25 out of 18)/(18 out of 21); 88.9/88.9% for endocrine manifestations (16 out of 18)/(12 out of 14); 88.9/88.2% for psychiatric manifestations (32 out of 36)/(24 out of 27); 88.5/85.6% for systemic phenomena (23 out of 26)/(17 out of 20); 85.7/91.7% for gastrointestinal manifestations (36 out of 42)/(29 out of 32); 83.3/79.2% for metabolic manifestations (15 out of 18)/(11 out of 14); 81.8/78.0% for neurological manifestations (27 out of 33)/(20 out of 25); 75.0/74.0% for hematological manifestations (36 out of 48)/(27 out of 36) and 74.3/72.8% for breathing manifestations (26 out of 35)/(19 out of 26). Finally, application of back-projection analysis (BPA) provides physic interpretation in structural terms through molecular graphics of the toxic effects predicted with these QSTR models. This article develops a mathematical model that encompasses a large number of drugs side effects grouped in specifics systems using stochastic entropies of interaction ($\Theta_k(j)$) by the first time.

Keywords: Markov model; Drugs side effects; Thermodynamic; ADL; QSAR; QSTR

1. Introduction

Traditionally, the search for new drugs has concentrated on the required activity, with considerations of bioavailability and toxicity being left until later in the development process. Numerous examples exist of drugs that have had to be withdrawn, because of unacceptable toxicity, in clinical trials and even after reaching the market-place. Such failures are very expensive, and have given rise to the saying; 'Fail early, fail fast, fail cheap(ly)'. Kennedy [1] reported that in 1997, of new drug entities that failed for reasons other than lack of efficacy, 16% failed in animal toxicity testing and 14% failed because of adverse effects in man. Since a typical drug takes 10–12 years, and costs up to $500 million, to reach the market, it is clearly important to discover potential toxicity as soon as possible [1].

For that reason, in recent years pharmaceutical companies have brought toxicity testing, as well as absorption, distribution, metabolism and excretion (ADME) studies, earlier in the drug development process. The ultimate here would be to use computer-based (in silico) methods to predict toxicity even before a drug candidate was synthesized. There are, however,

* Corresponding author. Tel.: +53 42 28 1473; fax: +53 42 28 1455.
E-mail addresses: maikelcm@gmail.com (M. Cruz-Monteagudo), maikelx77@yahoo.es (M. Cruz-Monteagudo).

0223-5234/$ - see front matter © 2005 Elsevier SAS. All rights reserved.
doi:10.1016/j.ejmech.2005.04.012

difficult problems to be overcome in this regard. Firstly, 'toxicity' covers a wide range of adverse effects; secondly, there is a paucity of data concerning, in particular, chronic toxicities, especially in humans; thirdly, the in silico methods currently available are class-specific and/or are of insufficient accuracy [2].

During last year, the pharmaceutical industries have reoriented their research strategies in order to give more attention to those methods that permit the "rational" selection or design of novel compounds whit the desired biological properties and drug safety profiles [3,4]. In these sense, quantitative structure–toxicity relationships (QSTR) are used as a predictive tools for a preliminary evaluation of the hazard of chemical compounds by using a computer aided models [5–7]. These models represent an alternative to the "real" world of assaying chemical compounds for determining their toxicological properties on live organisms in the laboratory avoiding the expensive, time-consuming and in many cases animal aggressive bioassays, which are now made only after preliminary predictions with computational models [8,9].

In general, González et al., have recently discussed that QSTR molecular models can be applied to congeneric and non-congeneric data sets of compounds. The first permits the understanding of specific mechanism of toxic action for molecules structurally related as well as to identify the different toxicological power of groups or substituents in such chemicals. On the other hand, the use of QSTR models for non-congeneric data sets permits the generalization of such mechanism to structurally diverse compounds as well as the identification of possible toxicophores of different structural nature [10–13].

On the other hand, Markov models are well-known mathematical tools for characterizing biomolecules structure. Markov models have been used for analyzing biological sequence data and they have been used to find new genes from the open reading frames [14,15]. Another use of these models is data-based searching and multiple sequence alignment of protein families and protein domains. Protein turn types and sub-cellular locations have been successfully predicted [16–19]. Hubbard and Park (Hubbard and Park, 1995) used amino acid sequence-based hidden Markov models to predict secondary structures. In this sense, Krogh et al. [16] have also proposed a hidden Markov model architecture. In addition, Markov's stochastic process has been used for protein folding recognition [20]. This approach can also be used for the prediction of protein signal sequences [21,22]. Another seminar works can be found related to the application of MC theory to Proteomic and Bioinformatics. Chou applied Markov models to predict beta turns and their types, and the prediction of protein cleavage sites by HIV protease [23]. Anyhow, have not been very used Markov models to develop QSTR studies and predict drugs side effect.

In this connection, our group has introduced elsewhere a physically meaningful mathematical approach based on Markov models (MARkovian Chemicals In Silico Design: MARCH-INSIDE) encoding molecular backbones information, with several applications in medicinal chemistry like drug design, QSAR/QSTR, molecular modeling and drug-receptor interactions. It allowed us introducing matrix invariants such as stochastic entropies for the study of molecular properties. Specifically, the entropy like molecular descriptors has demonstrated flexibility in many medicinal chemistry problems such as: estimation of anticoccidial activity, modeling the interaction between drugs and HIV-packaging-region RNA, and predicting proteins and virus activity [24–29], but until today multiple drugs side effects have not been modeled using entropy as molecular descriptor encoding information relative to multiple parameters like target site or toxic effect additionally to molecular structure, having a direct physicochemical interpretation in thermodynamic terms. In addition, the stochastic molecular descriptors have demonstrated flexibility in many molecular modeling problems such as: estimation of anticoccidial activity, modeling the interaction between drugs and HIV-packaging-region RNA, and predicting proteins and virus activity [24–29]. Applications to macromolecules have been restricted to the field of RNA without applications to proteins [30–36]. In the field of QSTR our group has reported the first model to predict chemically-induced agranulocytocis by small-to-medium sized drug-like molecules [37].

However, in spite of several QSTR studies reported there have not been seriously studied almost drug side effects. Unfortunately, the more than 1500 mol descriptors reported have not only been applied to study drug side effects but have very disperse theoretical definition and some times not very well established physical definition. Consequently, becomes a forefront problem applying molecular descriptors to drugs side effect study but at the same time represent them in unified mathematical framework giving better opportunities for physicochemical interpretation [38]. In the current paper we attempt to develop a more serious physicochemical interpretation of the MARCH-INSIDE descriptors with thermodynamic basis, which allow us to contrast the relationship among these descriptors and topologic, flexibility, and quadratic molecular descriptors [39]. These new interpretation allows us built up a molecular descriptor in thermodynamic terms [40] for predicting how likely a given drugs cause a specific side effect with respect to others side effects. This approach is able to take into consideration not only the molecular structure of the drug but the specific system the drug affects too. In particular will be possible correlate more than one property at time, in our case, drugs side effects, making it superior weigh against most of molecular descriptors which simply permit to correlate no more than one property at time, this advantage may be appropriately used in preliminary biological, pharmacological or toxicological studies, specially for comparative studies in early drugs develop stages. This study model a non-congeneric data set of 193 drugs of diverse molecular structure involved in 21 different side effects grouped in 10 affected biological systems, being 311 cases finally.

2. Materials and methods

2.1. Markov thermodynamics for drug–target step-by-step interaction

Let be, a hypothetical situation in which a drug molecule is free in the space at an arbitrary initial 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 in the time on the induction of a side effect. For the sake of simplicity, we are going to consider from now on a general structureless receptor. Understanding as structure-less molecular receptor a model of receptor which chemical structure it is not taken into consideration. The initial free energy of the drug-receptor interaction (^0g_j) is a state function so a reversible process of interaction may be separated on several elemental interactions between the j-th atom and the receptor [40]. Afterwards, interaction continues and we have to define the free energy of interaction between the j-th atom and the receptor given that i-th atom has been interacted at previous time t_k ($^kg_{ij}$). In particular, immediately after of the first interaction ($t_0 = 0$) takes place an interaction $^1g_{ij}$ at time $t_1 = 1$ and so on. So, one can suppose that, atoms begin its interaction whit the structureless molecular receptor binding to this receptor in discrete intervals of time t_k. However, there are several alternative ways in which such step-by-step binding process may occur. Fig. 1 illustrates this idea.

The free energy ^0g_j will be considered here as a function of the absolute temperature (T) of the system and the equilibrium local constant of interaction between the j-th atom and the receptor (^0k_j) [40]. Additionally, the energy $^1g_{ij}$ can be defined by analogy as $^1k_{ij}$:

$$^0g_j = -R \cdot T \cdot \log\ ^0k_j \tag{1}$$

$$^1g_{ij} = -R \cdot T \cdot \log\ ^1k_{ij} \tag{2}$$

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 ^0k_j as the rate of occurrence (n_j) of the j-th atom on molecules inducing the effect under study by molecule–receptor interaction with respect to the number of atoms in the molecule (n). With respect to $^1k_{ij}$ we must taking into consideration that once the j-th atom have interacted the preferred candidates for the next interaction are such i-th atoms bound to j by a chemical bond. Both constants can be then written down as:

$$^0k_j = \frac{n_j}{n} = e^{-\frac{^0g_j}{R \cdot T}} \tag{3}$$

$$^1k_{ij} = \alpha_{ij} \cdot \frac{n_j}{n} = e^{-\frac{^1g_{ij}}{R \cdot T}} \tag{4}$$

where α_{ij} is the elements of the atom adjacency matrix, n_j, n, ^0g_j, $^1g_{ij}$ have been defined in the paragraph above, and R is the gases universal constant.

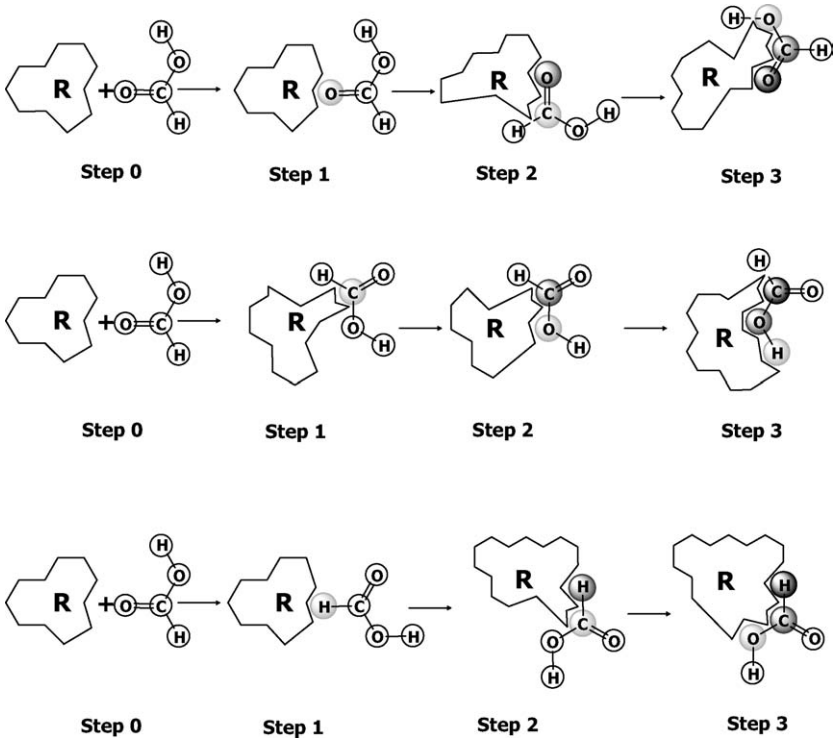


Fig. 1. Stochastic drug–target step-by-step interaction.

The second problem relates to the description of the interaction process at higher times $t_k > t_1$. Therefore, a MC model [30–36] 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 side effects. As depicted in Fig. 1, this model deals with the calculation of the probabilities (${}^{k}\pi_{ij}$) with which any arbitrary molecular atom *j*-th bind to the structureless molecular receptor given that other atom *i*-th has been bound before; along discrete time periods t_k ($k = 1, 2, 3, ...$); ($k = 1$ in light gray), ($k = 2$ in dark gray) and ($k = 3$ in black) throughout the chemical bonding system.

The procedure described here considers as states of the MC the atoms of the molecule. We can built up the corresponding absolute initial probability vector ${}^{A}\boldsymbol{\pi}_0$ and the stochastic matrix ${}^{1}\mathbf{\Pi}$, which has the elements ${}^{A}\pi_0(j)$ and ${}^{1}\boldsymbol{\pi}_{ij}$, respectively. The elements ${}^{A}\pi_0(j)$ of the above-mentioned vector ${}^{A}\boldsymbol{\pi}_0$ constitutes the absolute probabilities with which the *j*-th atom interact with the receptor at the initial time with respect to any atom in the molecule:

$$ {}^{A}\pi_0(j) = \frac{g_0(j)}{\sum_{a=1}^{m} g_0(a)} = \frac{\frac{n_j}{n}}{\sum_{a=1}^{m} \frac{n_a}{n}} = \frac{\frac{1}{n}\cdot n_j}{\frac{1}{n}\cdot\sum_{a=1}^{m} n_a} = \frac{n_j}{\sum_{a=1}^{m} n_a} \tag{5} $$

where *a* 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 ${}^{1}\mathbf{\Pi}$ is built too as a squared Table of order *n*, where *n* represents the number of atoms in the molecule. The elements (${}^{1}\pi_{ij}$) of the 1-step drug–target interaction stochastic matrix are the binding probabilities with which a *j*-th atom bind to a structureless molecular receptor given that other *i*-th atoms have been interacted before at time $t_1 = 1$ (considering $t_0 = 0$):

$$ {}^{1}\pi_{ij} = \frac{{}^{1}g_{ij}}{\sum_{a=1}^{\delta+1} g_{ia}} = \frac{\alpha_{ij}\cdot\frac{n_j}{n}}{\sum_{a=1}^{\delta+1} \alpha_{ia}\cdot\frac{n_a}{n}} = \frac{\frac{1}{n}\cdot\alpha_{ij}\cdot n_j}{\frac{1}{n}\cdot\sum_{a=1}^{\delta+1} \alpha_{ia}\cdot n_a} = \frac{\alpha_{ij}\cdot n_j}{\sum_{a=1}^{\delta+1} \alpha_{ia}\cdot n_a} \tag{6} $$

where δ is the valence of the *j-th* atom. The method arranges all the ${}^{A}\pi_0(j)$ values in a vector (${}^{A}\boldsymbol{\pi}_0$) and all the ${}^{1}\pi_{ij}$ constants as a squared table (${}^{1}\mathbf{\Pi}$) of $n \times n$ dimension. The calculation of ${}^{A}\boldsymbol{\pi}_0$ and ${}^{1}\mathbf{\Pi}$ is illustrated in Fig. 2. The more important approximation in the present work was considering that by using, both ${}^{A}\boldsymbol{\pi}_0$ and ${}^{1}\mathbf{\Pi}$ and Chapman–Kolgomorov equations [34] one can describe the further evolution of the system, determining the average constant of interaction between the *j*-th atom and the receptor at higher times. In mathematical terms, these probabilities are the elements of the matrices ${}^{k}\mathbf{\Pi}$, which could so be calculated as depicted in Eq. (7) [24–27,29–31,33–35,37,41]:

$$ {}^{k}\mathbf{\Pi} = ({}^{1}\mathbf{\Pi})^{k} \tag{7} $$

As the elements of the matrices ${}^{k}\mathbf{\Pi}$ depend on the adjacency relationships between the atoms on the molecule and the constants of interaction for each atom we can derive the stochastic entropy of interaction ($\Theta_k(j)$) between the drug and the receptor at a specific time:

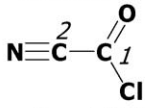
$$ \begin{aligned} \Theta_k(j) &= \mathbf{Sh}\cdot{}^{k}\mathbf{\Pi}\cdot{}^{\mathbf{A}}\pi_0 = \mathbf{Sh}\cdot({}^{1}\mathbf{\Pi})^{k}\cdot{}^{\mathbf{A}}\pi_0 \\ &= \mathbf{S}^{\mathbf{A}}\pi_k = -k_B\sum_{S_x}^{n} {}^{A}\pi_k(j)\cdot log^{A}\pi_k(j) \end{aligned} \tag{8} $$

where:

- ${}^{A}\boldsymbol{\pi}_0$ is the vector listing the absolute initial probabilities with which an atom interact with the receptor at the initial time with respect to any atom in the molecule.
- ${}^{A}\boldsymbol{\pi}_k$ is time dependent vectors, whose components are the absolute probabilities Apk(*j*) with which an atom interact with the receptor at the initial time with respect to any atom in the molecule.
- ${}^{1}\mathbf{\Pi}$ is the Markov or stochastic matrix, the ${}^{1}\mathbf{\Pi}$ matrix built up as a squared matrix $n \times n$ (*n* number of atoms in the molecule). The elements of this matrix of the 1-step drug–target interaction stochastic matrix are the binding probabilities with which a *j*-th atom bind to a structureless molecular receptor given that other *i*-th atoms have been interacted before at time $t_1 = 1$ (considering $t_0 = 0$).
- **Sh** is the log ${}^{A}\boldsymbol{\pi}_{k.}$.
- k_B is the Boltzmann's constant.
- S_x represents a specific group of atoms in the molecule. When S_x contains all the atoms in the molecule, $\Theta_k(j)$ becomes a global molecular index then we write $\Theta_k(T)$. We can calculate different families of molecular descriptor by selecting different S_x conditions, for this purpose atoms were groped in sets or classes (s_x): s_0 = CSat = Saturated carbon atom; s_1 = CInst = Unsaturated carbon atom; s_2 = Hal = Halogens; s_3 = Het = Heteroatoms; or s_4 = HX = Hydrogen bonded to heteroatom in order to describe local aspects of molecular structure.

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 selection of a Markov chain process [42,43] is not arbitrary. Due to atoms interactions are not dependent of previous atoms interactions we can affirm that a MC-based model of a stochastic drug–target step-by-step interaction obeys perfectly to the main characteristics of MC (a memory-less property). This implies that the probability of the occurrence of an event (atom union) does not depend on the history of the process. In other words, such a model will not depend of atoms unions at previous times.

Structure:



Cyanoformyl chloride

Π^1 =

	O	Cl	C_1	C_2	N
O	5.775	0	27.302	0	0
Cl	0	25.566	14.58	0	0
C_1	15.488	14.769	14.932	14.932	0
C_2	0	0	13.467	13.467	10.55
N	0	0	0	8.692	5.775

Π^1 =

	O	Cl	C_1	C_2	N
O	25.566	0	27.302	0	0
Cl	0	25.566	14.58	0	0
C_1	63.712	76.434	65.448	65.448	0
C_2	0	0	47.412	47.412	50.132
N	0	0	0	22.846	25.566

Π^1_{PhD} =

	O	Cl	C_1	C_2	N
O	4.563	0	5.219	0	0
Cl	0	4.563	7.173	0	0
C_1	14.299	12.345	14.955	14.955	0
C_2	0	0	8.255	8.255	8.126
N	0	0	0	4.692	4.563

Definition:

Π^1 =

	O	Cl	C_1	C_2	N
O	$\frac{O}{O+C_1}$	0	$\frac{C_1}{O+C_1}$	0	0
Cl	0	$\frac{Cl}{C_1+Cl}$	$\frac{C_1}{C_1+Cl}$	0	0
C_1	$\frac{O}{O+Cl+C_1+C_2}$	$\frac{Cl}{O+Cl+C_1+C_2}$	$\frac{C_1}{O+Cl+C_1+C_2}$	$\frac{C_2}{O+Cl+C_1+C_2}$	0
C_2	0	0	$\frac{C_1}{C_1+C_2+N}$	$\frac{C_2}{C_1+C_2+N}$	$\frac{N}{C_1+C_2+N}$
N	0	0	0	$\frac{C_2}{C_2+N}$	$\frac{N}{C_2+N}$

Fig. 2. Definition and calculation of Π^1 matrix for a specific compound in three particular cases of side effects. The element symbol is used to denote the value of the rate of recurrence [i.e. Cl represents the rate of recurrence (n_{Cl}) of chlorine atom for the specific side effect]. Thr: thromboembolism, Pat: pancreatitis, PhD: photodermatitis.

2.2. 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:

$$SE_x = b + \sum_{S_x,k} b(S_x, k) \times \Theta_k(S_x) \quad (9)$$

Here, Θ_k (S_x) can act as the local or global molecular descriptors as was defined above, all the Θ_k (S_x) variables where standardized using the Standardize module, implemented in STATISTICA 6.0. We selected linear discriminant analysis (LDA) [44,45] to fit the classification functions. The model deals with the classification of a set of compounds with diverse side effects. A dummy variable (SE_x) codifies the side effect studied. This variable indicates either the presence ($SE_x = 1$) or absence ($SE_x = -1$) of side effect studied. In Eq. (9), b(S_X, 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 [46]. Forward stepwise was fixed as the strategy for variable selection [44,45].

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 [34].

Clustering of compounds was carried out after previous perform of a canonical analysis using the algorithms implemented in the advanced options for LDA in the STATISTICA 6.0. This analysis offers as outputs the scores of every case for successive canonical roots which are orthogonal centered equations explaining decreased amounts of variance.

Consequently we can plot the scores for each compound in a Cartesian system of coordinates and using a symbol code visually exploring the possibility of clusters formations [34].

2.3. Data set methodology

The data set was conformed by series of the more frequently used drugs which produce side effects in different human organs systems, being these ones extensively tested in clinic and the side effects reported obtained by pharmacovigilance studies. The use of marketed drugs in data set confers a high confidence about the side effect reported. The set of drugs where extracted from a report of drugs side effects listed in literature [47]. The data set was conformed by 21 different drugs side effects grouped in 10 affected biological systems for 193 drugs, being 311 cases finally, taking into consideration that all side effects groups were statistically represented having each one at least seven drugs in order to perform a balanced training series.

2.4. Back-projection analysis and MARCH-INSIDE

In order to calculate the total atom contribution to a specific side effect in the current approach, we make use of the decomposition of total molecular descriptors into local descriptors. More specifically, we decompose the total molecular descriptors into atomic descriptors of the atom in the molecule. For example, the molecular descriptors of chloroform may be decomposed as follows: ${}^{SR}\pi_k(HCCl_3) = {}^{SR}\pi_k(H) + {}^{SR}\pi_k(C) + 3\,{}^{SR}\pi_k(Cl)$. Afterwards, the values of the atomic descriptor for each atom are substituted in the QSTR equation, obtaining the contribution of the atom to the specific side effect where zones shown in light gray (shown in dark gray) are those that have a low (high) contribution to the specific side effect. Only the zones that contribute to the specific side effect were quantified. Estrada and González have explained this procedure in detail for bond spectral moments [48]. The method, called back-projection analysis (BPA), is general for any molecular descriptor, defined a priori as a sum of local descriptors, at least for linear QSTR/QSARs. [49]. The main importance of BPA is that it offers a clear and direct interpretation of results in structural terms. Here we adapt a BPA approach to MARCH-INSIDE and LDA methodology. The present study is aimed on the selection of novel drug candidates for synthesis. Then, we select the different structural synthetic blocks of the molecules as molecular regions for the BPA. As LDA predicts the probability of occurrence of the side effect, we preferred to standardize all of the contribution in order to express them as the percentage of activity that each group accounts for.

3. Results and discussion

3.1. Mathematical model

The main advantage of our stochastic forms is the possibility of deriving average thermodynamic parameters depending on the probability of the states of the MC, which fit on more clearly physicochemical sense with respect to classic vector–matrix–vector forms. In specific, this work introduces by the first time a Markov form to calculate thermodynamic parameters of the drug–target interaction process considering in a unified scheme: time, chemical structure, and system including drug side effects.

Another advantage of the present stochastic vector–matrix–vector forms constitute the fact of it was not necessary considering different rates of occurrence for atoms of the same element but having different configuration e.g.: sp3, sp2, and sp carbons all were considered with the same atoms weights (rate of occurrence herein) for a specific side effect, the rate of carbon atoms. It was possible due to the ${}^{1}\pi_{ij}$ values clearly distinguished among these atoms because of the different connectivity (see Fig. 2). It is clear from Fig. 2 that atoms with different connectivity or configuration will have a different probability of union to the structureless molecular receptor in spite of having the same rate of occurrence. All these are the reasons for the selection in this work of our stochastic forms instead of others.

Aimed on finding some similarity whit others descriptors we could contrast our stochastic vector–matrix–vector forms $\Theta_k(S_x)$ whit Toropov optimization of correlation weights of local graph invariants (OCWLI) named flexible descriptors [50–53] (do not confound with flexibility descriptors). In this sense, both descriptors take into consideration more than one parameter. In flexible descriptors case, it is taken into consideration the abstract parameters (weights) which can be optimized in function of the pursued objectives. On the other hand, the parameters ours molecular descriptors take into consideration cannot be optimized, but have a direct physicochemical interpretation, such aspect has been analyzed in previous paragraphs of Section 2.

Once we perform a representative and balanced training series selection it could be used to fit the classification functions. The models where subjected to the principle of parsimony. Then, we chose a function with high statistical significance but having few terms $b(S_x, k) \times \Theta_k(S_x)$ as possible to each of 21 studied side effects. In order to derive a classification function that permits the classification of drugs as positive (presence of side effect) or negative (absence of side effect) we use the LDA in which stochastic entropies of interactions $\Theta_k(S_x)$ are used as independent variables. The classification models obtained to each studied side effect are given below in Table 1 together whit the statistical parameters of the LDA, validations of the current model by re-substitution of cases in four predicting series results and percents of good classification to each model. All the equations were statistically significant (Fisher ratio that determined a p-level < 0.05), this means that accepting these models as valid presupposes an error lower than 5%. Good average percentages of good classification in training/predicting series were obtained (74.3/72.8% to 92.6/91.7%).

In the models the coefficient U is the Wilk's statistics and F is the Fisher ratio. The Wilk's U-statistic is the standard

Table 1
Overall train accuracy, CV predictability, and models for different drugs side effects

Side Effects	Train	CV	Model
			Breathing manifestations
Infiltrated lung (**IL**)	70.0	61.5	**IL** = –14.21 + 3.66$^{1}\Theta$**(Cinst)** – 3.61$^{5}\Theta$**(Cinst)** – 1.76$^{5}\Theta$**(Csat)** + 16.03$^{1}\Theta$**(T)**
Bronchospasm (**Brch**)	77.3	79.5	**Brch** = –16.43 + 10.8166$^{1}\Theta$**(Cinst)** – 9.58$^{5}\Theta$**(Cinst)** – 0.12$^{5}\Theta$**(Csat)** + 13.18$^{1}\Theta$**(T)**
Total	**74.3**	**72.8**	N = 35 U = 0.723 F = 2.88 P = 0.0396
			Cardiovascular manifestations
Hipertensión (**HyperT**)	84.6	82.7	**HyperT =** –36.85 – 1.65$^{2}\Theta$**(Cinst)** + 229.40$^{5}\Theta$**(Het)** + 45.66$^{0}\Theta$**(T)** – 225.57$^{4}\Theta$**(Het)**
Thromboembolism (**Thr**)	100.0	100.0	**Thr =** –52.17 – 4.18$^{2}\Theta$**(Cinst)** + 344.85$^{5}\Theta$**(Het)** + 56.58$^{0}\Theta$**(T)** – 337.91$^{4}\Theta$**(Het)**
Total	**92.6**	**91.7**	N = 27 U = 0.317 F = 11.87 P = 0.0000
			Hematological manifestations
Agranulocytosis (**Agr**)	75.0	76.3	**Agr =** –58.04 + 8.94$^{0}\Theta$**(HX)** – 22.44$^{5}\Theta$**(Cinst)** + 48.18$^{5}\Theta$**(T)** + 33.10$^{0}\Theta$**(Cinst)**
Hemolytic anemia (**HA**)	75.0	72.3	**HA =** –69.03 + 11.07$^{0}\Theta$**(HX)** – 23.17$^{5}\Theta$**(Cinst)** + 51.56$^{5}\Theta$**(T)** + 34.87$^{0}\Theta$**(Cinst)**
Total	**75.0**	**74.0**	N = 48 U = 0.670 F = 5.29 P = 0.0015
			Gastrointestinal manifestations
Constipation or ileo (**CoI**)	90.9	95.5	**CoI =** –6.81 + 2.34$^{2}\Theta$**(Cinst)** + 1.31$^{5}\Theta$**(Het)** – 1.00$^{5}\Theta$**(HX)**
Pancreatitis (**Pat**)	93.3	93.3	**Pat =** –8.08 + 1.28 $^{2}\Theta$**(Cinst)** + 3.22$^{5}\Theta$**(Het)** + 1.86$^{5}\Theta$**(HX)**
Peptic or hemorrhagic Ulceration (**PoHU**)	75.0	87.5	**PoHU =** –13.78 + 3.05 $^{2}\Theta$**(Cinst)** + 3.23$^{5}\Theta$**(Het)** –0.67$^{5}\Theta$**(HX)**
Total	**85.7**	**91.7**	N = 42 U = 0.252 F = 12.26 P = 0.0000
			Dermal manifestations
Alopecia (**Alp**)	87.5	82.8	**Alp =** –23.05 + 7.63$^{0}\Theta$**(Csat)** + 6.44$^{1}\Theta$**(Cinst)** + 30.43$^{0}\Theta$**(Hal)**
Photodermatitis (**PhD**)	91.7	85.4	**PhD =** –28.09 + 6.71$^{0}\Theta$**(Csat)** + 8.04$^{1}\Theta$**(Cinst)** + 38.29$^{0}\Theta$**(Hal)**
Total	**89.3**	**83.9**	N = 28 U = 0.484 F = 8.54 P = 0.0005
			Systemic phenomena
Anaphylaxis (**Anph**)	85.7	83.9	**Anph =** –95.40 + 103.44$^{0}\Theta$**(T)** + 5.19$^{0}\Theta$**(Het)** + 0.93$^{5}\Theta$**(HX)**
Lupus Erythematosus (**LE**)	91.7	87.5	**LE =** –73.78 + 92.28$^{0}\Theta$**(T)** + 2.9$^{0}\Theta$**(Het)** + 1.97$^{5}\Theta$**(HX)**
Total	**88.5**	**85.6**	N = 26 U = 0.437 F = 9.45 P = 0.0003
			Endocrine manifestations
Galactorrhea (amenorrhea) (**Gal**)	88.9	88.9	**Gal =** –3.46 + 18.78$^{5}\Theta$**(Csat)** – 17.18$^{4}\Theta$**(Csat)**
Thyroid function test disorders (**TFTD**)	88.9	88.9	**TFTD =** –1.94 – 76.55$^{5}\Theta$**(Csat)** + 77.27$^{4}\Theta$**(Csat)**
Total	**88.9**	**88.9**	N = 18 U = 0.529 F = 6.68 P = 0.0084
			Metabolic manifestations
Hyperglycemia (**HyperG**)	77.8	88.9	**HyperG =** –12.00 + 55.68$^{5}\Theta$**(HX)** – 55.57$^{3}\Theta$**(HX)** + 13.02$^{1}\Theta$**(T)**
Hypopotassemia (**HypoP**)	88.9	69.4	**HypoP =** –16.93 + 126.13$^{5}\Theta$**(HX)** – 124.54$^{3}\Theta$**(HX)** + 15.04$^{1}\Theta$**(T)**
Total	**83.3**	**79.2**	N = 18 U = 0.586 F = 3.29 P = 0.0521
			Neurological manifestations
Convulsions (**Cvs**)	90.0	82.5	**Cvs** = –3.68 + 0.21$^{5}\Theta$**(Cinst)** + 1.86$^{0}\Theta$**(Cinst)** + 21.78$^{5}\Theta$**(Hal)** + 3.95$^{0}\Theta$**(Hal)** – 25.24$^{2}\Theta$**(Hal)**
Extrapyramidals effects (**EE**)	70.0	71.1	**EE** = –5.98 + 2.09$^{5}\Theta$**(Cinst)** + 0.09$^{0}\Theta$**(Cinst)** + 6.98$^{5}\Theta$**(Hal)** – 430.00$^{0}\Theta$**(Hal)** + 342.48$^{2}\Theta$**(Hal)**
Total	**81.8**	**78.0**	N = 33 U = 0.614 F = 3.40 P = 0.0164
			Psychiatric manifestations
Dysfunctions of the dream (**DD**)	85.7	85.7	**DD =** –4.88 + 2.58$^{5}\Theta$**(Cinst)**
Somnolence (**Snl**)	90.9	89.8	**Snl =** –13.03 + 4.43$^{5}\Theta$**(Cinst)**
Total	**88.9**	**88.2**	N = 36 U = 0.480 F = 36.81 P = 0.0000

$^{0}\Theta_k$ **(T)** represents a global molecular index. $^{0}\Theta_k$ **(CSat),** $^{1}\Theta_k$ **(Cinst),** $^{2}\Theta_k$ **(Hal),** $^{3}\Theta_k$ **(Het) and** $^{4}\Theta_k$ **(HX)** represents local molecular indices describing saturated carbon atom, unsaturated carbon atom, halogens, heteroatoms and hydrogen bonded to heteroatoms, respectively.

statistic that is used to denote the statistical significance of the discriminatory power of the current model [30,54]. Results displayed in Table 2 prove the robustness and predictability of the mathematical models obtained. The average percentage of good classification reached for cross-validation (CV) in training/predicting (robustness/predictability) series were (72.8/77.1% to 91.7/92.6%).

Using this approach is possible to take into consideration not only the molecular structure of the drug but the specific system the drug affects too. These new interpretation allows us built up a molecular descriptor in thermodynamic terms [40] for predicting how likely a given drugs cause a specific side effect with respect to others side effects. Applying these molecular descriptors to drugs side effect study is possible represent them at the same time in one unified mathematical framework giving better opportunities for a more serious physicochemical interpretation with thermodynamic basis [38]. In particular will be possible correlate more than one property at time, in our case, drugs side effects.

3.2. Canonical analysis

In order to simplify the equations for the purposes of interpretation and the possibility of graphical representation, we

Table 2
CV predictability and robustness for different drugs side effects

Side effects/CV	1 [a]	2 [a]	3 [a]	4 [a]	**Mean** [a]	1 [a]	2 [a]	3 [a]	4 [a]	**Mean** [a]
	Breathing manifestations									
	Predictability					Robustness				
Infiltrated lung	69.2	46.2	53.8	76.9	**61.5**	60.0	60.0	70.0	88.9	**69.7**
Bronchospasm	72.7	86.4	77.3	81.8	**79.5**	70.6	82.4	87.5	87.5	**82.0**
Total	**71.4**	**71.4**	**68.6**	**80.0**	**72.8**	**66.7**	**74.1**	**80.8**	**88.0**	**77.4**
	Cardiovascular manifestations									
	Predictability					Robustness				
Hypertension	84.6	76.9	84.6	84.6	**82.7**	80.0	90.0	80.0	88.9	**84.7**
Thromboembolism	100.0	100.0	100.0	100.0	**100.0**	100.0	100.0	100.0	100.0	**100.0**
Total	**92.6**	**88.9**	**92.6**	**92.6**	**91.7**	**90.5**	**95.2**	**90.0**	**94.7**	**92.6**
	Hematological manifestations									
	Predictability					Robustness				
Agranulocytosis	85.0	70.0	75.0	75.0	**76.3**	86.7	73.3	73.3	73.3	**76.7**
Hemolytic anemia	60.7	75.0	78.6	75.0	**72.3**	76.2	76.2	85.7	71.4	**77.4**
Total	**70.8**	**72.9**	**77.1**	**75.0**	**74.0**	**80.6**	**75.0**	**80.6**	**72.2**	**77.1**
	Gastrointestinal manifestations									
	Predictability					Robustness				
Constipation or ileo	100.0	90.9	90.9	100.0	**95.5**	88.9	87.5	75.0	87.5	**84.7**
Pancreatitis	93.3	93.3	93.3	93.3	**93.3**	91.7	90.9	100.0	90.9	**93.4**
Peptic or hemorrhagic ulceration	87.5	87.5	87.5	87.5	**87.5**	66.7	75.0	75.0	83.3	**75.0**
Total	**92.9**	**90.5**	**90.5**	**92.9**	**91.7**	**81.8**	**83.9**	83.9	**87.1**	**84.2**
	Dermal manifestations									
	Predictability					Robustness				
Alopecia	87.5	81.3	81.3	81.3	**82.8**	91.7	83.3	83.3	75.0	**83.3**
Photodermatitis	66.7	91.7	91.7	91.7	**85.4**	88.9	88.9	88.9	88.9	**88.9**
Total	**78.6**	**85.7**	**85.7**	**85.7**	**83.9**	**90.5**	**85.7**	**85.7**	**81.0**	**85.7**
Side Effects/CV	1*	2*	3*	4*	**Mean***	1*	2*	3*	4*	**Mean***
	Systemic phenomena									
	Predictability					Robustness				
Anaphylaxis	85.7	85.7	85.7	78.6	**83.9**	90.9	81.8	80.0	80.00	**83.2**
Lupus erythematosus	91.7	91.7	75.0	91.7	**87.5**	100.0	88.9	77.8	88.89	**88.9**
Total	**88.5**	**88.5**	**80.8**	**84.6**	**85.6**	**95.0**	**85.0**	**78.9**	**84.21**	**85.8**
	Endocrine manifestations									
	Predictability					Robustness				
Galactorrhea (amenorrhea)	88.9	100.0	100.0	66.7	**88.9**	85.7	100.0	100.0	83.3	**92.3**
Thyroid function test disorders	88.9	88.9	88.9	88.9	**88.9**	85.7	85.7	100.0	83.3	**88.7**
Total	**88.9**	**94.4**	**94.4**	**77.8**	**88.9**	**85.7**	**92.9**	**100.0**	**83.3**	**90.5**
	Metabolic manifestations									
	Predictability					Robustness				
Hyperglycemia	88.9	77.8	100.0	88.9	**88.9**	85.7	71.4	100.0	66.7	**81.0**
Hypopotassemia	66.7	77.8	66.7	66.7	**69.4**	85.7	100.0	100.0	83.3	**92.3**
Total	**77.8**	**77.8**	**83.3**	**77.8**	**79.2**	**85.7**	**85.7**	**100.0**	**75.0**	**86.6**
	Neurological manifestations									
	Predictability					Robustness				
Convulsions	80.0	90.0	90.0	70.0	**82.5**	73.3	86.7	93.3	86.7	**85.0**
Extrapyramidals effects	76.9	61.5	61.5	84.6	**71.1**	70.0	70.0	70.0	77.8	**71.9**
Total	**78.8**	**78.8**	**78.8**	**75.8**	**78.0**	**72.0**	**80.0**	**84.0**	**83.3**	**79.8**
	Psychiatric manifestations									
	Predictability					Robustness				
Dysfunctions of the dream	85.7	85.7	85.7	85.7	**85.7**	81.8	90.9	80.0	90.0	**85.7**
Somnolence	90.9	86.4	90.9	90.9	**89.8**	100.0	88.2	87.5	87.5	**90.8**
Total	**88.9**	**86.1**	**88.9**	**88.9**	**88.2**	**92.9**	**89.3**	**84.6**	**88.5**	**88.8**

[a] % of good classification based on posterior probabilities for four different training and predicting sets; predictability refers to compounds within predicting sets and robustness to compounds within training ones.

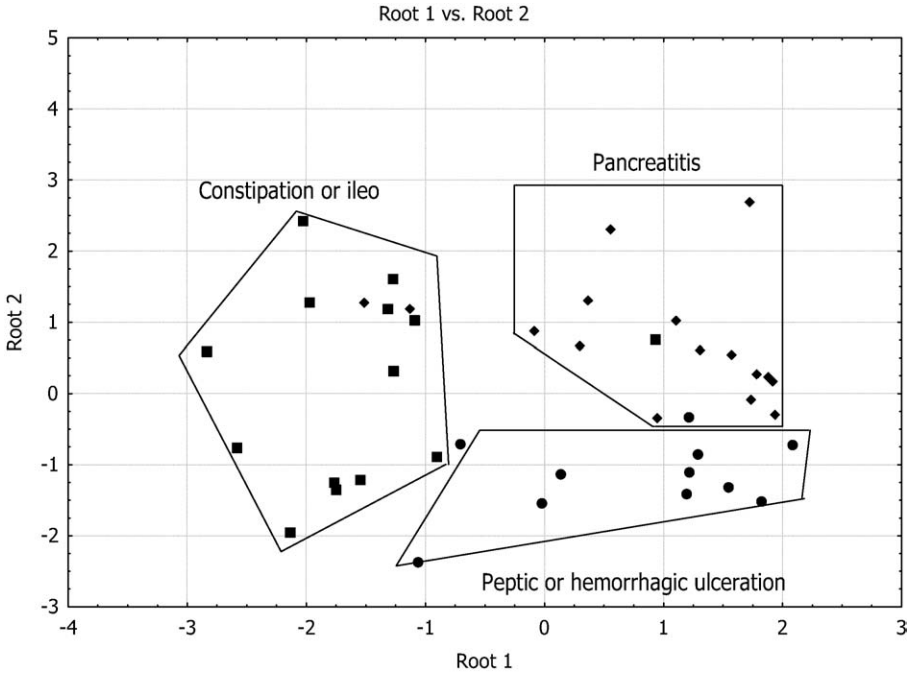


Fig. 3. Canonical roots analysis for gastrointestinal manifestations.

performed a canonical analysis [44] for gastrointestinal manifestations side effects group with the only purpose of illustrate the capability of the equations obtained to condense more than two side effects groups in only one simple equation (Root function) and its ability to discriminate between several side effects groups. The main root obtained (Root 1) proved to be a simple equation centered to 0:

$$Root1 = 0.87\Theta_2(CInst) - 0.25\Theta_5(Het) - 0.67\Theta_5(HX) \quad (10)$$
$$Root2 = 0.45\Theta_2(CInst) + 0.80\Theta_5(Het) + 0.13\Theta_5(HX)$$

This canonical root presented an eigen-value of 1.41 and an acceptable regression coefficient of 0.76, which it is statistically significant (P-level < 0.05), together whit a Chi-squared statistic of 52.22 and Wilk's lambda of 0.25 Fig. 3 show three side effects groups clearly distinguished and discriminated in a canonical space graph.

3.3. Back-projection analysis

Finally, we applied a BPA in order to carry out a physical interpretation in structural terms of the models obtained. BPA graphics for four somnolence inducer drugs (Pinazepam, Lorazepam, Diazepam and Oxazepam) where developed. As was explained in Section 2.3, zones shown in light gray (shown in dark gray) are those that have a low (high) contribution to the specific side effect.

Somnolence is the main side effect of ansiolitic benzodiazepines [55], an extension of its pharmacological action [56] due to its depressive action over central nervous system (CNS). All the important benzodiazepines depressives of the CNS contains a 5-aril or 5-ciclohexenil substituent, so the term benzodiazepines has ended up meaning the 5-aril-1,4-benzodiazepines. Besides the apparent requirement of the 5-Aril group, the structure–activity relationships are not strict [57,58].

In the graphic obtained by means of the BPA it is shown that 5-aril group (only requirement for the depressive action over CNS) shows a high contribution to the induction of somnolence (38.63%, 43.94%, 43.92% and 44.02%) for Pinazepam, Lorazepam, Diazepam and Oxazepam, respectively. All the above-mentioned confirm in structural terms the classification of these four benzodiazepines with a probability of 0.98, 0.97, 0.94 and 0.94, respectively, of producing somnolence. Fig. 4 show the BPA results.

Equation obtained for this drug side effect (**Snl** = –13.03 + 4.43$\boldsymbol{\Theta_5}$ **(Cinst)**) show a positive contribution of unsaturated carbons. The penetration of the molecule in to the central nervous system (CNS) thru the hematoencefalic barrier could be increased, probably due to the hydrophobic characteristics of the molecule originated by the prevalence of unsaturated carbons and at the same time could confer a high reactivity to the molecule of drug increasing the probability of interaction with the molecular receptor. This tentative interpretation is supported by the results obtained in BPA, proved in structural terms.

4. Conclusions

The fusion of high throughput screening and QSAR/QSTR [24–37,41,59] techniques in attempt to develop novel compounds with favorable drug-like characteristics specifically defined by pharmaceutic, pharmacokinetic, and drug safety profiles and minimize the costs in terms of time, financial, human and animal resources is becoming a viable alternative

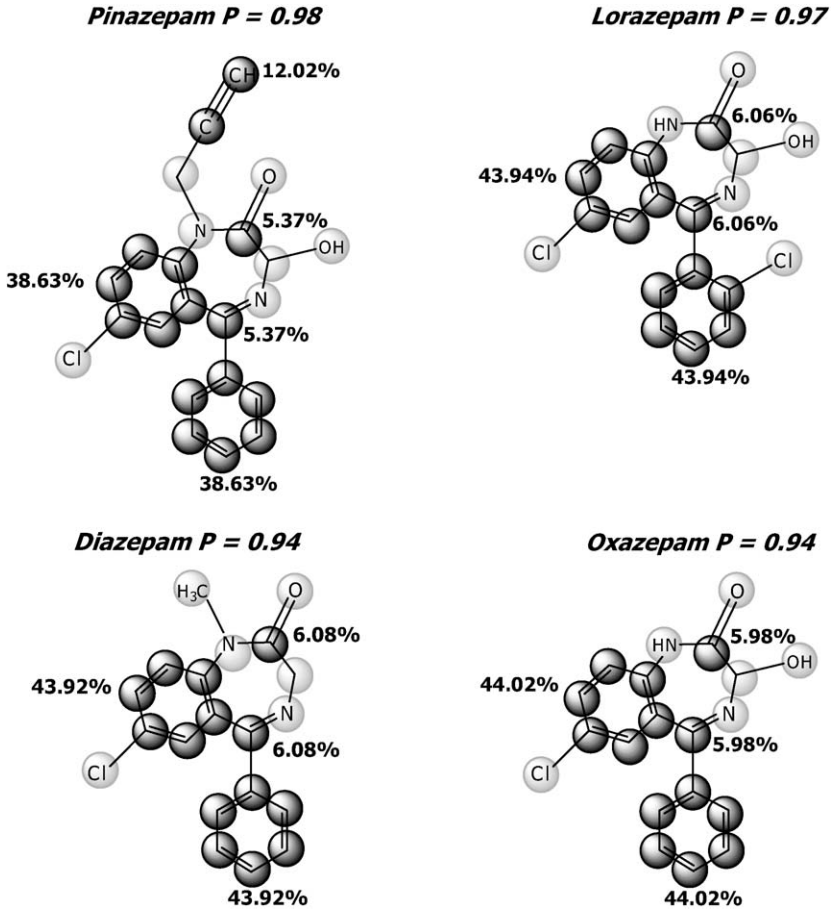


Fig. 4. Back-projection graphic for four drugs classified as able to induce somnolence (Pinazepam, Lorazepam, Diazepam and Oxazepam).

to massive screening. The results described here have demonstrated that MARCH-INSIDE methodology encode molecular backbones information, with several applications in medicinal chemistry like drug design, QSAR/QSTR, molecular modeling and drug-receptor interactions. This work confirms previous results on a better physicochemical interpretation for MARCH-INSIDE than for others [60]. Specifically, stochastic entropy of interaction (Θ_k (j)) is able to provide a physicochemical direct interpretation for drug–target step-by-step interaction taking into consideration not only the molecular structure of the drug but the specific system the drug affects too. In particular, thru this molecular descriptor will be possible correlate more than one property at time (in our case, drugs side effects) having a more serious physicochemical interpretation in thermodynamic terms. This fact make the present descriptors superior weigh against most of molecular descriptors, which correlate no more than one property at time [61–63]. This advantage may be appropriately used in preliminary biological, pharmacological or toxicological studies, especially for comparative studies in drug development early stages.

List of acronyms

MC: Markov chains
LDA: linear discriminant analysis
BPA: back-projection analysis
ADME: absorption, distribution, metabolism and excretion
QSTR: quantitative structure–toxicity relationships
HIV: human immunodeficiency virus
MARCH-INSIDE: Markovian chemicals in silico design
QSAR: quantitative structure–activity relationships
RNA: ribonucleic acid
OCWLI: optimization of correlation weights of local graph invariants
CNS: central nervous system

Acknowledgments

Cruz-Monteagudo, M. and González-Díaz, H. would like to express his more sincere gratitude to Professor Castro, E. and Professor Toropov, A.A. by kindly supply some very interesting materials on the topic of flexible indices which contribute to increase the quality of the present paper. Both authors also would like to acknowledge to Professor Olivier Lafont for his kind attention and to the two unknown referees for the opportune observations made which constitute a valuable help to increase the scientific quality of this article.

References

[1] T. Kennedy, Drug Discov. Today 2 (1997) 436.

[2] M.T.D. Cronin, Curr. Opin. Drug Discov. Dev. 3 (3) (2000) 292–297.

[3] J.M. Briggs, T.J. Marrone, J.A. McCammon, Trends Cardiovasc. Med. 6 (1996) 198–203.

[4] G. Wess, Drug Discov. Today 1 (1996) 529–532.

[5] M.T.D. Cronin, Pharm, Pharmacol. Commun. 4 (1998) 157–163.

[6] D.E.V. Lewis, Computer assisted methods in the evaluation of chemical toxicity, in: K.B. Lipkowitz, D.B. Boyd (Eds.), Reviews in Computational Chemistry, vol. 3, VCH Publishers, New York, 1992, pp. 173–222.

[7] M.T.D. Cronin, J.C. Dearden, Quant. Struct.-, Act. Relat. 4 (1995) 1–7.

[8] J.C. Dearden, M.T.D. Cronin, A.J. Dobbs, Chemosphere 31 (1995) 2521–2528.

[9] D.W. Roberts, An analysis of published data on fish toxicity of nitrobenzenes and aniline derivatives, in: K.L.E. Kaiser (Ed.), QSAR in Environmental Toxicology-II, D. Raidel, Dordrecht, The Netherlands, 1987, pp. 295–308.

[10] M.P. González, A.H. Morales, R. Molina, J.F. García, Polym. 45 (2004) 2773–2779.

[11] M.P. González, A.H. Morales, H. González-Díaz, Polymers 45 (2004) 2073–2079.

[12] M.P. González, H. González-Díaz, M.A. Cabrera-Pérez, R.R. Molina, Bioorg. Med. Chem. 12 (2004) 735–744.

[13] A.H. Morales, M.P. González, J.B. Rieumont, Polymers 45 (2004) 2045–2050.

[14] M. Vorodovsky, E.V. Koonin, K.E. Rudd, Trends Biochem. Sci. 19 (1994) 309.

[15] M. Vorodovsky, J.D. Macininch, E.V. Koonin, K.E. Rudd, C. Médigue, A. Danchin, Nucleic Acids Res. 23 (1995) 3554.

[16] A. Krogh, M. Brown, I.S. Mian, K. Sjeander, D.J. Haussler, Mol. Biol. 235 (1994) 1501.

[17] K.C. Chou, Biopoymers. 42 (1997) 837.

[18] Z. Yuan, FEBS Lett. 451 (1999) 23–26.

[19] S. Hua, Z. Sun, Bioinformatics 17 (2001) 721–728.

[20] V. Di Francesco, P.J. Munson, J. Garnier, Bioinformatics 15 (1999) 131.

[21] K.C. Chou, Peptides 22 (2001) 1973–1979.

[22] K.C. Chou, Curr. Protein Pept. Sci. 3 (2002) 615–622.

[23] K.C. Chou, J. Biol. Chem. 268 (1993) 16938–16948.

[24] H. González-Díaz, R. Ramos De A, R.R. Molina, Bioinformatics. 19 (2003) 2079–2087.

[25] H. González-Díaz, R.R. Molina, E. Uriarte, Polym. 45 (2004) 3845–3853.

[26] H. González-Díaz, R.R. Molina, E. Uriarte, Bioorg. Med. Chem. Lett. 14 (2004) 4691–4695.

[27] H. González-Díaz, I. Bastida, N. Castañedo, O. Nasco, E. Olazábal, A. Morales, H.S. Serrano, R. Ramos De A, Bull. Math. Biol. 66 (2004) 1285–1311.

[28] R. Ramos de A, H. González-Díaz, R.R. Molina, M.P. González, E. Uriarte, Bioorg. Med. Chem. 12 (2004) 4815–4822.

[29] R. Ramos de A, H. González-Díaz, R.R. Molina, E. Uriarte, Proteins, Struct. Funct. Bioinf. 56 (2004) 715–723.

[30] H. González-Díaz, E. Olazábal, N. Castañedo, S.I. Hernádez, A. Morales, H.S. Serrano, J. González, R. Ramos de A, J. Mol. Mod. 8 (2002) 237–245.

[31] H. González-Díaz, E. Uriarte, R. Ramos de Armas, Bioorg. Med. Chem. 13 (2005) 323–331.

[32] H. González-Díaz, R. Ramos de Armas, E. Uriarte, J. Online, Bioinf. 1 (2002) 83–95.

[33] H. González-Díaz, O. Gia, E. Uriarte, I. Hernádez, R. Ramos, M. Chaviano, S. Seijo, J.A. Castillo, L. Morales, L. Santana, D. Akpaloo, E. Molina, M. Cruz-Monteagudo, L.A. Torres, M.A. Cabrera, J. Mol. Mod. 9 (2003) 395–407.

[34] H. González-Díaz, S.I. Hernández, E. Uriarte, L. Santana, Comput. Biol. Chem. 27 (2003) 217–227.

[35] H. González-Díaz, R. Ramos de A, R.R. Molina, Bull. Math. Biol. 65 (2003) 991–1002.

[36] H. González-Díaz, E. Uriarte, R. Ramos de A, Bioorg. Med. Chem. 13 (2004) 323–331.

[37] H. González-Díaz, Y. Marrero, I. Hernández, I. Bastida, E. Tenorio, O. Nasco, E. Uriarte, N.C. Castañedo, M.A. Cabrera-Pérez, E. Aguila, O. Marrero, A. Morales, M.P. González, Chem. Res. Toxicol. 16 (2003) 1318–1327.

[38] R. Todeschini, V. Consonni, Handbook of Molecular Descriptors, Wiley VCH, Weinheim, Germany, 2000.

[39] H. Kubinyi, J. Taylor, C. Ramdsen, Quantitative drug design, in: C. Hansch (Ed.), Comprehensive Medicinal Chemistry, vol. 4, Pergamon, Oxford, 1990, pp. 589.

[40] A. Villa, R. Zangi, G. Pieffet, A.E. Mark, J. Comput, Aid. Mol. Des. 17 (2003) 673–686.

[41] R. Ramos de Armas, H. González-Díaz, R. Molina, E. Uriarte, Biopolymers 77 (2005) 247–256.

[42] B. Gnedenko, in: The Theory of Probability, Ed. Mir Publishers, Moscow, 1978, pp. 107–112.

[43] J.A. Freund, T. Poschel, Stochastic processes in physics, chemistry, and biology, in: Lecture Notes in Physics, Springer-Verlag, Berlin, Germany, 2000.

[44] H. Van Waterbeemd, Discriminant analysis for activity prediction, in: R. Manhnhold, P. Krogsgaard-Larsen, H. Timmerman, Van Waterbeemd, V.C.H. Weinhiem (Eds.), Method and Principles in Medicinal Chemistry, vol. 2. "Chemometric Methods in Molecular Design", 1995, pp. 265–282.

[45] R.B. Kowalski, S. Wold, Pattern recognition in chemistry, in: P.R. Krishnaiah, L.N. Kanal (Eds.), Handbook of Statistics, North Holland Publishing Company, Amsterdam, 1982, pp. 673–697.

[46] STATISTICA for Windows release, version, 6.0., 2001, Statsoft Inc.

[47] A.G. Garcia, J.F. Horga de la Parte, Reacciones adversas a los fármacos, in: INTERCON (Ed.), Índice de Especialidades Farmacéuticas. Prescripción Racional de Fármacos, Editores Médicos S.A. EDIMSA, Madrid, 1994, pp. 155–173.

[48] M.A. Cabrera, D.H. González, C. Teruel, J.M. Pla-Delfina, M. Bermejo del Val, Eur. J. Pharm. Biopharm. 53 (2002) 317–325.

[49] N. Stief, K. Baumann, J. Med. Chem. 46 (2003) 1390–1407.

[50] A.A. Toropov, A.P. Toropova, J. Mol. Struct. Theochem. 538 (2001) 287–293.

[51] A.A. Toropov, A.P. Toropova, J. Mol. Struct. Theochem. 581 (2002) 11–15.

[52] A.A. Toropov, A.P. Toropova, J. Mol. Struct. Theochem. 637 (2003) 1–10.

[53] A.A. Toropov, E.J. Benfenati, J. Mol. Struct. Theochem. 676 (2004) 165–169.

[54] R. Franke, Theoretical Drug Design Methods, Elsevier, Amsterdam, 1984.

[55] T.R. Browne, N. Clonazepam, Engl. J. Med. 299 (1978) 812–816.

[56] P.L. Morselli, Psychotropic drugs, in: P.L. Morselli (Ed.), Drug Disposition During Development, Spectrum Publications, Inc., New York, 1977, pp. 431–474.

[57] L.H. Sternbach, Chemistry of 1,4-benzodiazepines and some aspects of the structure–activity relationship, in: S. Garahini, E. Mussini, L.O. Randall (Eds.), The Benzodiazepines, Raven Press, New York, 1973, pp. 1–26.

[58] D.J. Greenblatt, R.I. Shader, Benzodiazepines in Clinical Practice, Raven Press, New York, 1974.

[59] H. González-Díaz, G. Agüero, M.A. Cabrera, R.R. Molina, L. Santana, E. Uriarte, et al., Bioorg. Med. Chem. Lett. 15 (2005) 551–557.

[60] O. Gia, S.M. Magno, H. González-Díaz, E. Quezada, L. Santana, E. Uriarte, L. DallaVia, Bioorg. Med. Chem. 13 (2005) 809–817.

[61] M.A. Cabrera-Pérez, M. Bermejo, M.P. Gonzalez, R. Ramos de A, J. Pharm. Sci. 7 (2004) 1701–1717.

[62] M.A. Cabrera-Pérez, M. Bermejo-Sanz, Bioorg. Med. Chem. 12 (2004) 5833–5843.

[63] M.A. Cabrera-Pérez, A.R. García, C.F. Teruel, I.G. Álvarez, M. Bermejo-Sanz, Eur. J. Pharm. Biopharm. 56 (2003) 197–206.